Stochasticity and Heterogeneity in Growing Bacterial Populations

From Neutral Evolution to Stochastic Toxin Release

Matthias Ferdinand Lechner

München 2017

Stochasticity and Heterogeneity in Growing Bacterial Populations

From Neutral Evolution to Stochastic Toxin Release

Matthias Ferdinand Lechner

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Matthias Ferdinand Lechner
aus München

München, den 13. Dezember 2017

Zusammenfassung

Heterogene Bakterienpopulationen stehen unter dem ständigen Einfluss stochastischer Effekte, welche deren Diversität sowohl erhöhen, als auch zerstören können. In meiner Dissertation präsentiere ich drei verschiedenen Projekte, in denen ich auf unterschiedlichen Ebenen diese Einflüssse untersucht habe.

1. Nichtselektive Evolution wachsender Populationen

mit Karl Wienand, Felix Becker, Heinrich Jung, und Erwin Frey.

In unserem ersten Projekt haben wir die nicht-selektive Evolution zweier Spezies untersucht. In der bisherigen Forschung wurde dabei stets eine konstante Populationsgröße angenommen, wodurch eine der beiden Spezies letztendlich ausstirbt. Mit Hilfe eines Pólya-Urnenmodells untersuchten wir den realistischeren Fall einer exponentiell wachsenden Population. Unsere Ergebnisse zeigen, dass die Zusammensetzung von Populationen unter exponentiellem Wachstum nicht fixiert, sondern gegen einen zufälligen Grenzwert konvergiert. Wir konnten diese Grenzwerte sowie die Zeitentwicklung der ersten drei Momente ermitteln, und unsere Ergebnisse experimentell bestätigen. Dieses Projekt hat zu einer Publikation geführt, die dieser Doktorarbeit angefügt ist.

2. Post-transkriptionelle Regulation der ColicinE2-Expression in Escherichia Coli

mit Mathias Schwarz, Alexandra Götz, Madeleine Opitz, und Erwin Frey.

ColicinE2 ist ein Bakterientoxin, welches nach Stress-Signalen produziert und durch Zell-Lyse freigesetzt wird. In diesem Projekt haben wir das Regulationsnetzwerk für die Produktion und Freisetzung von ColicinE2 mathematisch modelliert, und seine Reaktion auf stochastische Effekte untersucht. Dabei zeigte sich, dass die hierarchische Struktur der regulierenden Komponenten einen bisher unbekannten Mechanismus zur Unterdrückung von Fluktuationen darstellt. Danach koppelten wir unser Modell an stochastische Stress-Signale. Dies hat eine, auch im Experiment beobachtete, breite Verteilung an Lysezeiten zur Folge, die wir quantitativ bestimmt haben. Im zweiten Teil dieses Projektes unterstützten wir mit unserem Modell den erstmaligen experimentellen Nachweis einer regulierende Funktion von *single-stranded DNA* (ssDNA). Die Ergebnisse dieses Projektes haben zu einer Publikation geführt, die dieser Doktorarbeit angefügt ist. Eine weitere Publikation ist in Vorbereitung.

3. Optimale Zeitverteilungen für lysebasierte Toxin-Sekretion mit Erwin Frey.

Ausgehend von unseren Arbeiten über ColicinE2 untersuchten wir die Frage, welche Verteilung der ColicinE2-Lysezeiten ideal ist, um Konkurrenten abzuwehren, und welche Faktoren dies beeinflussen. Dazu entwickelten wir ein einfaches, konzeptionelles Modell, um mittels eines genetischen Algorithmus die optimalen Verteilungen zu ermitteln. Die anschließenden numerischen Simulationen zeigen, dass breite Verteilungen mit positiver Schiefe in allen untersuchten Parameter-Konstellationen die beste Lyse-Strategie darstellen. Diese generelle Form zeigt sich auch in experimentell beobachteten Verteilungen. Eine Publikation der Ergebnisse ist in Vorbereitung.

Summary

My thesis deals with the interplay of heterogeneity and stochasticity in growing bacterial populations. This interplay is an important factor in the evolution of microbial organisms, and proves to be relevant from the population level down to protein synthesis in cells. We worked with two different model systems in three projects, which also structure this thesis into three chapters. Each chapter starts with a short project abstract, which I reprint here to give a short initial summary of this thesis.

1 Non-Selective Evolution of Growing Populations

with Karl Wienand, Felix Becker, Heinrich Jung, and Erwin Frey.

In this joint theoretical and experimental project, we focussed on populations of two species with equal growth rates. This non-selective form of evolution has been previously studied, but only for the special case of fixed population size. It was shown using Wright-Fisher models that, in this case of fixed populations, demographic fluctuations (*genetic drift*, see subsection 1.1.1) eventually lead to the fixation of one species, and thus to the loss of diversity in the population. Together with Karl Wienand and Erwin Frey, I studied the more general case of growing populations using a Pólya urn model approach (subsection 1.1.2). We found that, after a short initial genetic drift, the population composition "freezes" to a non-fixated steady state. Consequently, genetic diversity is maintained. Our theoretical results were experimentally verified by Felix Becker and Heinrich Jung. This joint work has been published in the research paper "Non-Selective Evolution of Growing Populations", which is reprinted in section 1.5 (and is also reprinted in the PhD theses of Karl Wienand and Felix Becker).

2 Post-transcriptional Regulation of ColicinE2 Expression in Escherichia Coli

with Mathias Schwarz, Alexandra Götz, Madeleine Opitz, and Erwin Frey.

The secretion of toxins is a potent mechanism for bacteria to kill other strains in the competition for resources. In this project, I analysed (together with Mathias Schwarz and Erwin Frey) the regulatory network (see 2.1.2) of the toxin ColicinE2. The ColicinE2 system is activated by stochastic SOS responses (see 2.1.3), and regulates the production and release of the toxin using three hierarchically ordered components (see 2.1.4). We developed a rate equation model for this system by introducing effective quantities and reducing the complexity of the regulatory network with a time scale separation approach. A parameter analysis of the resulting system revealed that the hierarchical nature of the network reduces internal fluctuations, and allows for a fine-tuned response to SOS signals. Moreover, we coupled our reduced system to a stochastic SOS response model and reproduced broad lysis time distributions. These distributions are also found in experiments from the group of Madeleine Opitz. Our work on this model and its analysis have been published in the research paper "Hierarchical Post-transcriptional Regulation of Colicin E2 Expression in

Escherichia coli", which is reprinted in section 2.5. In a subsequent experimental study on five different strains containing the ColicinE2 system, Alexandra Götz and Madeleine Opitz found experimental evidence that single stranded DNA (ssDNA) in the cells acts as fourth regulative component. Together with Erwin Frey, I extended the model for ColicinE2 regulation with this fourth component, and could show that the additional regulator is indeed necessary to reproduce the experimental results. Moreover, we complemented the model analysis by investigating the influence of the different plasmid compositions in the investigated strains. The extension of the model with ssDNA, as well as the analysis of the experimental and theoretical results, are detailed in the paper draft "CsrA and its regulators control the time-point of ColicinE2 release in *Escherichia coli*", which is reprinted in section 2.6 (and also in the PhD thesis of Alexandra Götz).

3 Optimal Time Distributions for Lysis-based Toxin Release
with Erwin Frey.

This follow-up project on the ColicinE2 system investigates, which toxin release distribution is optimal to fend off competitors, and how this result is determined by the parameters of the system. In chapter 2, we found that changes in parameters of our model for ColicinE2 regulation alter the resulting lysis time distribution. As the toxins are released to defend their producers against competitors, this raises the question as to which distribution is best in killing other bacteria. To identify the factors affecting the distribution, I created a conceptual model for the self-destructive toxin release of bacteria (with Erwin Frey). Using this model in combination with a genetic algorithm (see 3.1.2), we determined optimal lysis time distributions for given parameter sets, and analysed our results in context of phenotypic heterogeneity (see 3.1.1). The optimal distributions were then successfully put to the test on a stochastic lattice-gas model. The detailed development of our model and the full presentation of the results are given in the paper draft "Optimal Time-Distributions for Lysis-based Toxin Release", which is reprinted in section 3.5.

Contents

Contents

Introduction

When we think of bacterial growth, we most likely imagine a large population of genetically identical bacteria growing isolated on or in a medium. It is easy to understand why this idea was conceived: Both experimental and theoretical studies are much easier to set up, analyse, interpret and replicate if only one type of organism is involved. In nature, however, such homogeneous conditions are rare, and *heterogeneous populations* are the rule.

In the human large intestine, for instance, the microbiome of a single human comprises more than 500 different bacterial strains [1], which all compete for the same space and resources. The ubiquity of heterogeneity also means that many bacteria co-evolved with each other, and thus developed a variety of mechanisms for interaction. This comprises means to gain advantages over competing strains, like the production of toxins, but also cooperative phenomena such as public goods or division of labour.

Another important aspect for researching growing populations is stochasticity, which is also directly connected to heterogeneity: random mutations in the genome [2] or, for instance, noisy expression of proteins [3], are major causes for diversity in bacterial colonies. Random genetic drift, however, has been shown to produce homogeneous populations [4]. This complex interplay between the different components, in concert with the stochastic interactions, makes heterogeneous populations particularly attractive for physicists.

In the following, I present my work on three projects with two bacterial systems that exhibit different forms of heterogeneous populations. While the projects deal with effects on very different levels (bacterial populations, post-transcriptional regulation within a cell) , they all present different viewpoints of the overarching question of this thesis: **What mechanisms maintain diverse populations in the interplay between heterogeneity and stochasticity?**

In this thesis, I address each of the three projects in a separate chapter. The main content of my research is given in form of papers or paper drafts, which are reprinted in full at the end of each chapter. Beforehand, I give an abstract of the project, along with a short introduction to biological and mathematical background topics, which might be less familiar to a general physicist audience. This is followed by a summary of the results, and a more detailed conclusion, which discusses our results as well as their relevance to current and future studies.

Contents

Part I

Neutral Growth of Bacteria

Non-Selective Evolution of Growing Populations

1

Abstract and Contributions   In this joint theoretical and experimental project, we focussed on populations of two species with equal growth rates. This non-selective form of evolution has been previously studied, but only for the special case of fixed population size. It was shown using WRIGHT-FISHER models that, in this case of fixed populations, demographic fluctuations (*genetic drift*, see subsection 1.1.1) eventually lead to the fixation of one species, and thus to the loss of diversity in the population. Together with KARL WIENAND and ERWIN FREY, I studied the more general case of growing populations using a Pólya urn model approach (subsection 1.1.2). We found that, after a short initial genetic drift, the population composition "freezes" to a non-fixated steady state. Consequently, genetic diversity is maintained. Our theoretical results were experimentally verified by FELIX BECKER and HEINRICH JUNG. This joint work has been published in the research paper "Non-Selective Evolution of Growing Populations", which is reprinted in section 1.5 (and is also reprinted in the PhD theses of KARL WIENAND and FELIX BECKER).

1.1 Background

1.1.1 Selection and Neutral Evolution

natural
selection

For many years after the release of CHARLES DARWIN's "On the origin of species" [5], evolution has been primarily considered in the form of *natural selection*. This theory comprises three essential principles: Mutations create a (1) *variation* of traits within a population, which are (2) *inherited* to each trait's offspring, but show (3) *different survival rates* (*fitnesses*). This last principle is most characteristic for natural selection, and interestingly contrasts with the first: Mutations are stochastic, and thus create new traits at random, while selection eventually removes less fit traits from the population. The interplay of these forces shapes the composition of populations, and "directs" them towards adaptation to the environment.

neutral
evolution

However, this interplay relies on the fact that the different traits indeed have a different fitness. The color of a prey animal, for instance, is a very important selective trait if its predators have good vision, but becomes completely irrelevant if all predators in its habitat are blind (e.g. underground). For such neutral settings, natural selection theory is not able to predict the outcome of evolution. This gap was closed in the 1950s and 1960s by MOTOO KIMURA's theory of neutral evolution [6]. His work highlighted the influence of stochastic effects to evolution, and started an ongoing debate on whether selective or neutral forces are dominant (see, for instance, Ref. [7]).

genetic
drift

Particularly, he studied *genetic drift*[1] in finite populations [4] due to random sampling (which corresponds to demographic fluctuations in population dynamics). This neutral theory relies on three main assumptions: Only two traits/species, A and B, exist in the population (that is, no mutations), none of the traits has a selective advantage, and the population size $N = N_A + N_B$ is fixed. The evolution of this population occurs via joint birth/death events, in

Wright-Fisher
model

particular, as a WRIGHT-FISHER model [2, 8, 9]: A random individual is selected and produces an offspring of its own trait, which then replaces a random individual in the population. This is often analysed using an urn model analogy, mapping individuals to coloured marbles and the population to an urn. The state of the population can be represented by $x = N_A/N$, the fraction of A individuals in the population. Then, the evolution of the population is equivalent to a random walk between the boundaries 0 and 1. If the population fraction reaches one of

fixation

the boundaries, the population *fixates*: All individuals are either A (if $x = 1$) or B (if $x = 0$). As also all offspring will keep this trait, the boundaries are absorbing.

For a more detailed study on the stochastics of genetic drift, KIMURA considered distributions of population compositions, $P(x, t)$, and derived a FOKKER-PLANCK equation for their

diffusion
equation

evolution, which reads

$$\frac{\partial}{\partial t} P(x, t) = \frac{1}{4N} \frac{\partial^2}{\partial x^2} \left[x(1 - x) P(x, t) \right]. \tag{1.1}$$

As perfectly neutral conditions are assumed, it is a diffusion equation with no additional terms. The analysis of this equations shows that any initial coexistence state of A and B decays exponentially to fixation. More specifically, during this evolution, any initial probability

---

[1]From a physicists perspective, the naming convention "drift" is unfortunate, as the Fokker-Planck equation describing this phenomenon, eq. (1.1), only contains a diffusion, but no drift term.

distribution $P(x, t_0)$ quickly flattens out, and turns to a U-shape. Like all phenomena based on demographic fluctuations, these effects become smaller when the population size N is increased, because then single birth/death events affect x less.

1.1.2 Pólya Urn Model

In our work, we make use of a well-studied variant of the WRIGHT-FISHER urn model: the PÓLYA urn model [10, 11]. It describes the same setup as the WRIGHT-FISHER model, but differs in one crucial step: When the "offspring marble" is added to the population, it does not replace an existing marble. Each birth event thus increases the population size N by 1.

Pólya urns

Due to the lack of death events, PÓLYA urns have interesting properties: First, and most simply, no trait can fixate. Second, they constitute self-reinforcing processes, as every marble extraction increases the share of the selected trait in the population, thereby increasing the chance of another extraction. However, the impact of noise from the stochastic birth events constantly decreases, as the population size grows. Third, while the population size N grows, the urn composition x converges to a limit value x^*, which is a random variable. The limit distribution has been shown to be a beta-binomial distribution [12].

1.2 Motivation and Research Question

Previous research in our group dealt with the co-evolution of two bacterial strains in the context of the "dilemma of cooperation" [13–15]. These studies considered very small sub-populations in liquid medium, all consisting of different, randomly pipetted compositions of the two strains. Indeed, situations like this occur frequently in nature (often referred to as *founder effect* [2, 16]), for instance, when random pathogens from a host reservoir spread to a new host. Despite the focus on selective evolution in these studies, the experimental testing of the models required an initial phase of neutral growth to increase . However, it was unclear how the composition of these sub-populations evolves: The growth conditions are neutral, and thus, the theory of genetic drift (see above) should apply; yet, the fact that the populations are growing violates a fundamental assumption of KIMURA's theory (fixed population size). There are concepts for effective population sizes [17] which solve this problem approximately, but they break down when the population grows over several orders of magnitude for a prolonged time. This gave rise to the research question of this project:

How does exponential growth affect genetic drift?

1.3 Summary of Results

Stochastically initialised Pólya urn model for experimental setting. To find a model for non-selectively growing bacteria in our experimental setting, we first focussed on the initial condition. The specific pipetting procedure in the experiment is resembled by drawing the initial abundances of A and B from POISSON distributions, with means $\overline{N}_0 \overline{x}_0$ and $\overline{N}_0(1 - \overline{x}_0)$, respectively. We apply this initialisation method to a large ensemble of populations, which then grow according to a PÓLYA urn process. The waiting times between the birth events are

assumed to follow a exponential distribution (Poissonisation). With this assumption, the probability of finding N_A individuals of type A in a population at time t follows the master equation

$$\frac{d}{dt}P(N_A, t) = (N_A - 1)P(N_A - 1, t) - N_A P(N_A, t).$$ (1.2)

For B, an analogous equation holds.

Growth freezes genetic drift.   We implemented our model in numerical simulations, which ran on large ensembles of stochastically initialised populations. The results show that the distribution of x initially broadens (like in Kimura's genetic drift model), but then quickly freezes to a steady state. In particular, the populations do not fixate (except for populations that already started fixated), contrary to their behaviour for genetic drift.

Approximative solution for time evolution of moments.   To see how the distribution of x evolves over time for arbitrary $\overline{N}_0$ and $\overline{x}_0$, we determined an approximate solution for the first three moments of x. These can be computed using the first three moments for N_A and N_B, which we obtained from (1.2) and its B counterpart, employing a generating function approach. The first moment, $\langle x \rangle$, the mean of x, remains constant. If this were not the case, the process would be selective, and not neutral. The second moment reads

$$\mathrm{Var}(x) = \frac{2 - e^{-t}}{\overline{N}}\overline{x}(1 - \overline{x}).$$ (1.3)

It changes over time (in particular, it increases initially), but converges exponentially to a limit value. These features also hold for the skewness, which has a more complicated expression.

Exact solution for limit value of moments.   With a different approach, relying on properties of Pólya urns and the beta-binomial distribution, we obtained exact solutions for the steady state of x. From this distribution, we could derive exact expressions for the average and variance of x. In the limit of large N, the latter reads

$$\mathrm{Var}(x) \rightarrow \frac{2}{N}\overline{x}(1 - \overline{x}),$$ (1.4)

which is the $t \rightarrow \infty$ limit of eq. (1.3)

Experiments confirm numerical simulations following our model   Our co-authors Felix Becker and Heinrich Jung tested the numerical predictions for the steady state distribution with stochastically pipetted *Pseudomonas putida* cultures. The experimental results matched our predictions.

1.4 Conclusion

Non-selective (or neutral) evolution is a fundamental mechanism, in which stochasticity rather than selection shapes the composition of populations. Its most prominent paradigm is genetic drift: Demographic fluctuations in a heterogeneous population result in the fixation of a single trait. This seminal result was shown by M. Kimura for populations of constant size using the Wright-Fisher model.

In this project, we revisited neutral evolution from a new, more realistic perspective: **How does exponential growth affect genetic drift?** To answer this question, we replaced the Wright-Fisher model with the concept of Pólya urns. Our work showed that, in stark contrast to having a constant population size, **exponential growth freezes genetic drift**, and thus **maintains genetic diversity**. The term "freezes" becomes clear when we look at the detailed evolution of a population: Initially, the composition changes due to demographic fluctuations, but the more the population size grows, the smaller the changes become. Eventually, the composition of the population converges to a random limit value, but never fixates.

How much this "freezing" behaviour affects a distribution of compositions in an ensemble of populations depends on the initial conditions. This is again a difference from fixed population size. To investigate this dependence in more detail, we considered the special case of Poisson initial conditions, which commonly appears in natural systems. Using this initial condition, we were able to derive exact expressions for the steady state of the distribution, and approximate formulas for the time evolution of the moments. Using these expressions, we found that, for very small or very large populations, the final distribution is already largely determined by the initial compositions. For populations of intermediate size, however, the stochastic effects of growth dominate. This result highlights that randomness in populations has two sources: initial sampling and demographic fluctuations.

Because of these insights on the neutral evolution of composition distributions, we believe that our research has important implications for both experimental and conceptual studies on genetically diverse populations. Previous studies have highlighted the importance of heterogeneous populations as starting point for many aspects of evolution [18, 19]. In particular, we expect that our research is important for the study of cyclic metapopulations, in which basic concepts rely on very small, diverse sub-populations [13–15].

1.5 Publication Reprint

Non-selective evolution of growing populations

by

K. Wienand,*[1] M. Lechner,*[1], F. Becker,*[2] H. Jung,[2] and E. Frey[1]

* Contributed equally to this work

[1] Department of Physics, Arnold Sommerfeld Center for Theoretical Physics and Center for NanoScience, Ludwig-Maximilians-Universität München, Theresienstraße 37, 80333 München, Germany,

[2] Department of Biology 1, Microbiology, Ludwig- Maximilians-Universität, Grosshaderner Straße 2-4 82152, Martinsried, Germany

reprinted from

RESEARCH ARTICLE

Non-Selective Evolution of Growing Populations

Karl Wienand[1�], **Matthias Lechner**[1☺], **Felix Becker**[2☺], **Heinrich Jung**[2], **Erwin Frey**[1]*

1 Arnold-Sommerfeld-Center for Theoretical Physics and Center for NanoScience, Physics Department, Ludwig-Maximilians-Universität, Munich, Germany, **2** Department of Biology 1, Microbiology, Ludwig-Maximilians-Universität, Martinsried, Germany

☺ These authors contributed equally to this work.

* frey@lmu.de

Abstract

Non-selective effects, like genetic drift, are an important factor in modern conceptions of evolution, and have been extensively studied for constant population sizes (Kimura, 1955; Otto and Whitlock, 1997). Here, we consider non-selective evolution in the case of growing populations that are of small size and have varying trait compositions (e.g. after a population bottleneck). We find that, in these conditions, populations never fixate to a trait, but tend to a random limit composition, and that the distribution of compositions "freezes" to a steady state. This final state is crucially influenced by the initial conditions. We obtain these findings from a combined theoretical and experimental approach, using multiple mixed subpopulations of two Pseudomonas putida strains in non-selective growth conditions (Matthijs et al, 2009) as model system. The experimental results for the population dynamics match the theoretical predictions based on the Pólya urn model (Eggenberger and Pólya, 1923) for all analyzed parameter regimes. In summary, we show that exponential growth stops genetic drift. This result contrasts with previous theoretical analyses of non-selective evolution (e.g. genetic drift), which investigated how traits spread and eventually take over populations (fixate) (Kimura, 1955; Otto and Whitlock, 1997). Moreover, our work highlights how deeply growth influences non-selective evolution, and how it plays a key role in maintaining genetic variability. Consequently, it is of particular importance in life-cycles models (Melbinger et al, 2010; Cremer et al, 2011; Cremer et al, 2012) of periodically shrinking and expanding populations.

Introduction

Stochastic effects play an important role in population dynamics [8–11], particularly when competition between individuals is non-selective. Most previous theoretical analyses have studied how a non-selectively evolving trait can spread and eventually replace all other variants (fixate) under conditions in which the population size remains constant [2, 12, 13]. However, both natural and laboratory populations frequently experience exponential growth. Here we show that genetic diversity in growing populations is maintained despite demographic noise,

and reaches a stationary but random limit. We used a well-controlled model system in which well-mixed co-cultures of a wild-type *Pseudomonas putida* strain and an isogenic mutant were grown under non-selective conditions. Multiple subpopulations were generated, each containing a random number of individuals of each strain. Depending on the average initial population size and the strain ratio, we observed distinct stationary probability distributions for their genetic composition. Moreover, we showed that the dynamics of growing populations can be mapped to Pólya urn models [4], permitting the observed maintenance of genetic diversity to be understood as the random limit property of a fair game between individual strains. Generalizing the Pólya urn model to include the effects of random initial sampling and exponential growth allowed us to predict the evolution of the composition distribution. Using numerical and analytical methods we found that the distribution broadens at first but quickly "freezes" to a stationary distribution, which agrees with the experimental findings. Our results provide new insights into the role of demographic noise in growing populations.

Results and Discussion

Evolutionary dynamics is driven by the complex interplay between selective and non-selective (or neutral) effects. The paradigm of non-selective evolution originates from the seminal work of Kimura [1], in which he solved the Wright-Fisher model, thus showing that non-selective effects—and specifically genetic drift—can have a determinant role in evolution. His results sparked an ongoing debate about the nature and potency of randomness as a fundamental evolutionary force [13–15]. For very small populations genetic drift is generally considered an important factor [13], as the theory successfully predicts the outcomes of neutral evolution experiments [9, 16].

In most theoretical analyses, constant (or effectively constant) population sizes are assumed, and the role of population growth is neglected. Bacterial populations, however, often undergo rapid growth—especially when they are small. For example, as few as 10 individuals of some highly virulent pathogens (e.g. enterohemorrhagic *Escherichia coli* or *Shigella dysenteriae*) suffice to initiate a deadly infection in a human host [17, 18]. Another case of small, growing populations are water-borne bacteria that feed on phytoplankton products. Due to nutrient limitation in open water, these bacteria typically live in small populations in close proximity to the planktonic organism [19]. During spring blooms, the phytoplankton releases more organic material, boosting the bacterial growth rate [19–21]. In nature, such small populations often form by adventitious dispersal from a larger reservoir population [22]. A typical example is the spreading of pathogens from host to host. This random "sampling" from a reservoir yields small populations whose genetic compositions differ from that of the reservoir (a phenomenon known as the *founder effect* [23]). Recent studies also showed that the combination of population growth and stochastic fluctuations can have a major impact on the evolution [5–7, 24] and genetics [25] of small populations.

To probe how population growth shapes genetic diversity, we used a well-characterized microbial model system, namely the soil bacterium *Pseudomonas putida* KT2440 [3, 29, 27]. The wild-type strain KT2440 produces pyoverdine, an iron-scavenging molecule that supports growth when iron becomes scarce in the environment. Here we consider co-cultures of two genetically distinct strains: the wild-type, pyoverdine-producing strain KT2440 (strain *A*) and the mutant non-producer strain 3E2 (strain *B*). We set up conditions of non-selective competition between these strains by using an iron-replete medium (casamino acids, supplemented with 200 μM $FeCl_3$). In this medium, production of pyoverdine is effectively repressed [27], such that both strains have the same growth rate and neither has an advantage (see S2 Table). Producer (KT2440 wild type) and non-producer (3E2) strains were first mixed and diluted to
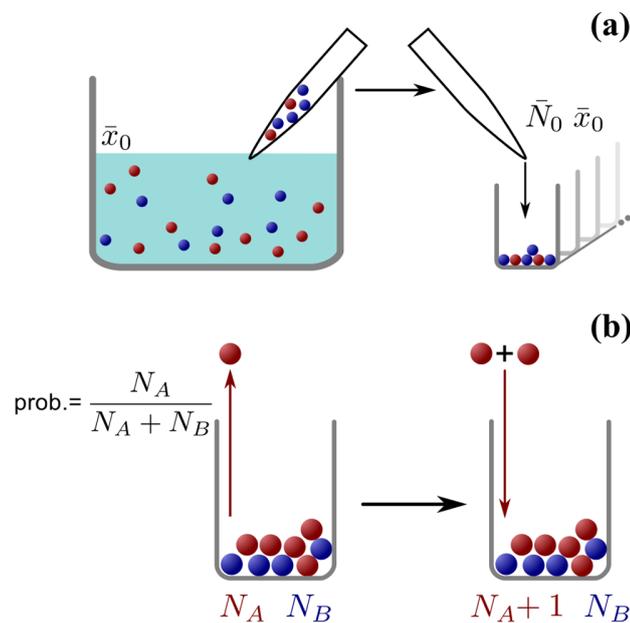
Fig 1. Schematic depiction of urn sampling and growth. (a) Schematic illustration of the random initial conditions. An infinite reservoir contains a diluted mixture of bacteria, a fraction $\bar{x}_0$ of which are of strain *A*. We draw small volumes of liquid from the reservoir containing small, random numbers of individuals, which conform to a Poisson distribution with mean (determined by the dilution of the reservoir population). A certain fraction of this initial population is of strain *A*. The mean value of this fraction is equal to $\bar{x}_0$. We use these individuals to initiate populations in the wells of a microtiter plate, so that each population starts with a random size N_0 and a random fraction of A-individuals x_0. **(b)** Illustration of the Pólya urn model. If a bacterial population is represented as an urn, each individual as a marble and each bacterial strain as a color, this urn model captures the essentials of bacterial reproduction in our populations. At each iteration, a marble is drawn at random and returned to the urn, together with another one of the same color. The probability of extracting a marble of either color is determined solely by its relative abundance, making the process non-selective (since no strain has inherent advantages, see S2 Table). The rate of growth in population size can be rendered exponential (see S2 Fig) by letting the waiting time between successive iterations be exponentially distributed (also known as Poissonization).

doi:10.1371/journal.pone.0134300.g001

yield Poisson dilution conditions. Then we initiated a large number of subpopulations from this reservoir by pipetting aliquots of the cell suspension into the wells of a 96-well plate, thereby generating a large ensemble of subpopulations with a random distribution of initial cell number N_0 and producer fraction x_0 (Fig 1). Use of shaken liquid cultures ensured homogeneous well-mixed conditions for all cells in the same well (access to nutrients, oxygen, etc.), and exponential growth was observed in all cases (see S2 Fig).

This experimental setting is well described within the mathematical framework of a *Pólya urn model*. Consider each bacterium in the population as a marble in an urn, and its genotype as the color of the marble (e.g. red for strain *A*, and blue for strain *B*). Population growth results from single reproduction events in which an individual randomly divides. This is mathematically equivalent to a stochastic event in which a marble is chosen at random from the urn and put back, together with another one of the same color. This random process, introduced by Eggenberger and Pólya [4], exhibits several important properties [28–31]. It is *self-reinforcing*: each time a marble is extracted, another one of the same color is added, increasing the likelihood of extracting a marble of that color again. In the context of bacterial populations, this means that every birth event for one strain makes it more likely that further birth events of that same strain will occur in the future. Note, however, that *fixation*, i.e., complete loss of one type

of marble from the population, cannot occur, simply because in the Pólya urn model marbles are neither removed nor do they change their color. This fully reflects the experimental conditions: During exponential growth, rates of cell death are negligible, and within the observation time mutations will be extremely rare, given the population sizes considered. The bacteria in each well reproduce randomly at a per-capita (average) rate μ. To translate this to the urn model, drawing of a marble is assumed to be a stochastic Poisson process, with a "per-marble" rate μ (a procedure known as *Poissonization* or *embedding* [32, 33]). Mathematically, the growth process in then described by a Master Equation: The time evolution for the probability $P(N_A, t)$ of finding N_A individuals of strain A at time t reads

$$\frac{d}{dt}P(N_A; t) = (N_A - 1)P(N_A - 1; t) - N_A P(N_A; t),$$ (1)

where we have set the growth rate to $\mu = 1$ in order to fix the time scale (for an introduction to the mathematical concepts see, e.g., [34]); the corresponding Master equation for individuals of strain B is of identical form. To study the composition of the populations, we use the more convenient quantities $N = N_A + N_B$ (total size) and $x = N_A/N$ (fraction of individuals of strain A).

To start the experiment, we inoculated the wells of 96-well-plates by drawing small volumes of diluted liquid bacterial culture from a large reservoir (Fig 1(a)). Each volume contains a random number of bacteria whose mean value is controlled by the dilution of the reservoir. The fraction of bacteria of strain A (wild type) in that volume is also random, with its mean value $\bar{x}_0$ given by the fraction of strain A in the reservoir. In the mathematical formulation, this setup corresponds to stochastic initial conditions for the Pólya urn model: the initial population size N_0 for each well is given by a Poisson distribution with mean $\bar{N}_0$, and each individual is assigned to strain A or B with probability $\bar{x}_0$ and $1 - \bar{x}_0$), respectively. This procedure is also equivalent to treating the initial numbers of A- and B-individuals as independent, Poisson-distributed random variables with mean values $\bar{N}_0 \bar{x}_0$ and $\bar{N}_0(1 - \bar{x}_0)$), respectively [6].

Fig 2 shows a time series of the histogram for the composition x of all subpopulations considered, as obtained from a stochastic simulation of the Master Eq (1) for a given random initial condition (with $\bar{N}_0 = 10$ and $\bar{x}_0 = 0.33$). Surprisingly, the distribution first broadens, but then quickly "freezes" to a steady state (see S1 Video). This is genuinely different from Kimura"s result for populations with constant size [1] (or similar results with effectively constant size [2]) where the balance between stochastic birth and death events leads to genetic drift, and thereby eventually to the extinction of one of the two strains. In contrast, for a growing population, death events are negligible, and therefore there is no fixation of the population during growth. Instead, fixation arises as a direct consequence of the initial sampling process, as can be seen from the heights of the black bins in the histogram (at $x = 0$ and $x = 1$), which remain constant over time (Fig 2). During growth, the composition of each subpopulation, instead of drifting to fixation at either $x = 0$ or $x = 1$, reaches a stationary limit value x^*, where it remains thereafter [35]. This limit value is random: starting several subpopulations (urns) from the exact same initial composition of strain A and B (blue and red marbles), each reaches a limit, but in general these limits differ from one another. Once all of the subpopulations in an ensemble reach their limit, the distribution of the population composition freezes to a steady state, which is equal to the probability distribution of x^*. Similar random limit properties appear in other fields, with *lock-in* in economics as the best-known example [30].

The inset in Fig 2 shows approximate solutions for the time evolution of mean, standard deviation, and skewness of the composition x, which we obtained by analytically solving the Master Eq (1) (see S2 Text). The analytical results agree well with their numerical counterparts. In particular, the mean value remains constant over time, as it must for a non-selective process.
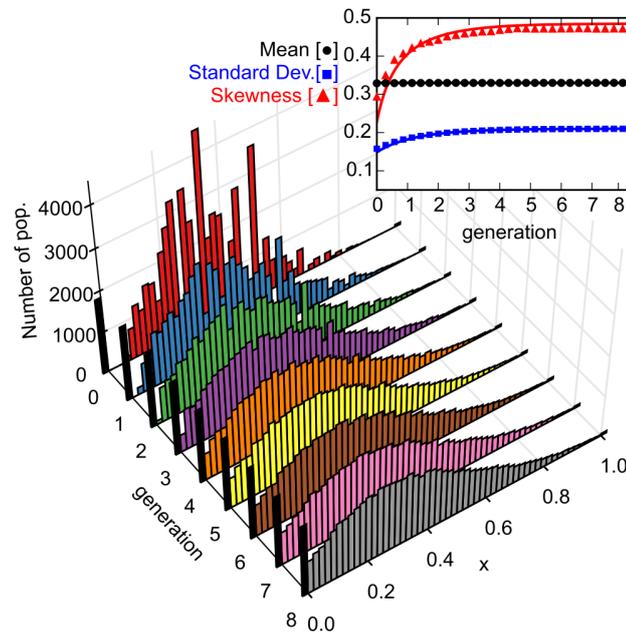
Fig 2. Time series for the simulated distribution of the population composition *x*. The distribution initially broadens, then freezes to a steady state (see S1 Video). The fraction of populations that have *x* = 0 or *x* = 1 (indicated by the black bins) remains constant during the time evolution, as expected for a Pólya urn process, and in contrast to expectations from genetic drift (see S1 Table). In each well the population follows a stochastic path and reaches a (random) limit composition, and the distribution freezes only when all populations reached their limit. The parameter values used in the simulation are $\bar{N}_0 = 10$ and $\bar{x}_0 = 0.33$ The inset shows the mean, standard deviation and skewness as a function of the number of generations, with symbols denoting numerical simulations, and the solid lines corresponding to the theoretical prediction of Eq (2) (and also those in S2 Text). Analytical and numerical values agree. The mean $\langle x \rangle$ remains constant throughout the evolution, as expected for a non-selective process; standard deviation and skewness saturate to limit values, confirming the freezing of the distribution.

doi:10.1371/journal.pone.0134300.g002

For the time evolution of the variance, which is a measure for the spread of a distribution, we obtain to leading order in population size

$$\mathrm{Var}_{\mathrm{poi}}[x](t) = \frac{2 - \mathrm{e}^{-t}}{\bar{N}_0} \bar{x}_0 (1 - \bar{x}_0) \,. \tag{2}$$

The broadening and freezing of the distribution is reflected in the exponential decay term of the variance. Note that the skewness increases as well, because growth is self-reinforcing (see inset in Fig 2). To further test the validity of the stochastic simulations, we also calculated the limit values of the average and variance after extended periods of evolution exactly, and found that they match the numerical solutions of the Master Equation perfectly (see S1 Text).

We tested these theoretical predictions using *P. putida* as a bacterial model system. We mixed the wild-type and mutant strains in order to obtain different initial fractions $\bar{x}_0$. The degree of dilution of the mixture determines the average initial cell number $\bar{N}_0$, with which we inoculated 120 wells per experiment (96-well plate format). In order to compare the experimental data with our model, we set up simulations that matched the experimental configuration by initializing $\bar{N}_0$ and $\bar{x}_0$ with the same values as measured in the experiments. We simulated the time evolution of about 10^4 populations, grouped in "virtual plates" of 120 wells each. Every virtual plate produced a histogram like the one we obtained from experiments. We then generated an average histogram of the virtual plates and used its values to compute the
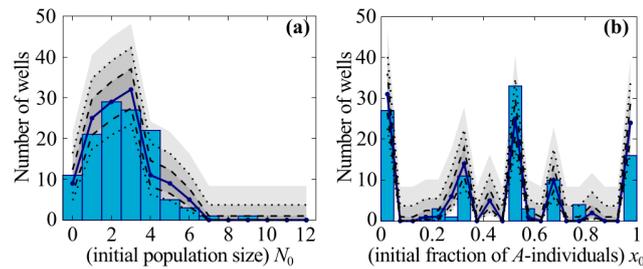
Fig 3. Initial distributions for population size N_0 and composition x_0 (parameter values $\bar{N}_0 = 2.55$, $\bar{x}_0 = 0.45$). The experimental distributions (bars) for N_0 (panel **(a)**) and x_0 (panel **(b)**) are measured from 120-well ensembles. The average $\bar{N}_0$ and $\bar{x}_0$ calculated from the measured values determine the parameters for the simulated distributions. The theoretical average distribution (solid blue line) is the average of the same distributions generated for 84 sets of 120 wells. Using that average we calculate the Wilson binomial confidence intervals (gray areas) for 68% (between dashed lines), 95% (between dotted lines) and 99.73% confidence. The measured and simulated distributions agree well within statistical error, confirming our assumption that individuals of strain A and B in the experiments start Poisson-distributed with mean $\bar{N}_0\bar{x}_0$ and $\bar{N}_0(1-\bar{x}_0)$, respectively. The ragged distribution of x_0 derives from a small-number effect, and disappears at larger values of N_0 (see main text, and also S3 Fig).

doi:10.1371/journal.pone.0134300.g003

binomial confidence intervals [36] for the count in each bin, and compared those with the experimental distribution.

Fig 3(a) shows a representative experimental histogram of the initial population sizes N_0 for strong dilution with $\bar{N}_0 = 2.55$. It is well approximated by a Poisson distribution, and agrees with the simulation results within statistical errors (blue line and shaded gray areas in Fig 3(a)). Fig 3(b) shows the probability distribution of the corresponding initial compositions x_0 of the populations, where again theoretical and experimental values agree well within statistical error. Note also that in every well the composition x_0 must be a simple fraction; this means that only a few numerical values are possible for small initial population sizes N_0. This small-number effect explains why the distribution of x_0 in Fig 3(b) is so ragged. The distribution becomes much smoother for larger initial population sizes (see S3 Fig). Taken together, these results for the distribution of initial population size and composition confirm that the inoculation of the individual wells is a stochastic sampling process with Poissonian statistics.

Next, we were interested in how the composition of the bacterial population would evolve under non-selective (neutral) growth conditions. To this end we let the 120 populations grow for an 11-hour period, during which they remained in exponential growth phase (see S2 Fig). Then we measured the population size $N(t)$ in each well by counting colony-forming units, and $x(t)$ by counting the pyoverdine-producing colonies (see Materials and Methods). Fig 4 shows the final outcome for four different initial conditions, i.e. combinations of the initial average population size $\bar{N}_0$ and composition $\bar{x}_0$. We first wanted to know what determines the number of wells that contain only individuals of either strain A or strain B, i.e. that are fixated. To this end we compared the experimentally observed values with the corresponding predictions from the numerical simulations of the Pólya urn model (Fig 4). Since both results agree within statistical error, we conclude that fixation of a population is a consequence of the initial sampling process and is not due to fixation during population growth (see also S1 Table). This is especially obvious for small average initial population size or compositions close to $x = 0$ or $x = 1$, where a large fraction of the wells contains cells of strain A or B only (Fig 4(a) and 4(d)). Next we wished to learn how the final distribution of the population composition (i.e. the random limits, x^*) depends on the initial average composition $\bar{x}_0$. For $\bar{x}_0 = 0.5$, we observed both by experiment and theoretically that the initial distribution significantly broadened (by a factor
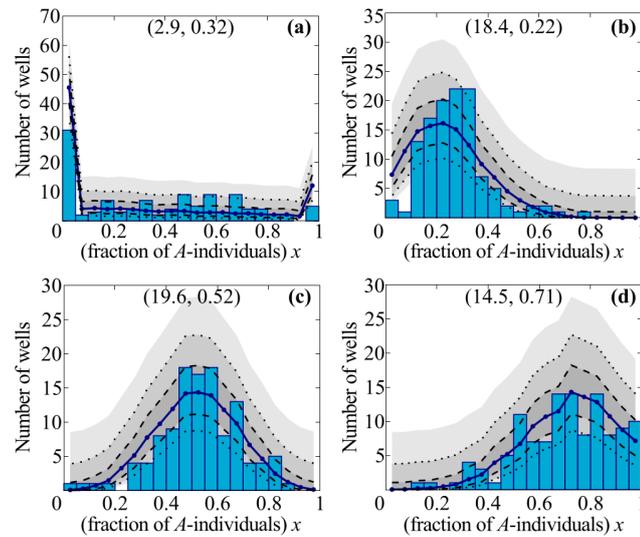
Fig 4. Steady-state distributions of population composition *x* for different initial conditions. The experimental distribution (bars) is the result of growth on 120 independent wells. We use the measured average x_0 and N_0 from the experiments to initialize the simulations of several 120-well ensembles. After growth, we compute the histogram for each of these ensembles and obtain the average theoretical distribution (blue line). Using the values from this distribution, we compute the three confidence intervals (shaded gray areas) for each bin for 68% (between dashed lines), 95% (between dotted lines) and 99.73% confidence. The two sets of data match: most experimental data agree with the first prediction confidence region, practically all with the second one. The limit distributions are clearly different from the initial ones (see S1 Fig). The importance of growth in changing the distributions depends on the initial size N_0 (see main text, and S1 Fig). Parameter values: $\bar{N}_0 = 2.9, \bar{x}_0 = 0.32$ (panel **(a)**); $\bar{N}_0 = 18.4, \bar{x}_0 = 0.22$ (panel **(b)**); $\bar{N}_0 = 19.6$, $\bar{x}_0 = 0.52$ (panel **(c)**); $\bar{N}_0 = 14.5, \bar{x}_0 = 0.71$ (panel **(d)**).

doi:10.1371/journal.pone.0134300.g004

of $\sqrt{2}$) but remained symmetrical (Fig 4(c) and S1 Fig). In contrast, starting from distributions with average values below or above 0.5 caused the final distribution to broaden and also become skewed towards smaller or larger values of *x*, respectively (Fig 4(b) and 4(d)). Moreover, we found quantitative agreement between experiment and numerical simulations within statistical errors in all analyzed parameter regimes (see blue lines and shaded areas in Fig 4): most experimental histograms fall within the first confidence interval of the prediction (darkest gray areas, between dashed lines), and almost all fall within the 99.73% confidence interval.

Taken together, our combined theoretical and experimental analysis gives a coherent picture of evolution during non-selective (exponential) growth. We have shown, experimentally and by analogy with the Pólya urn model, that for each well-mixed population the composition of the population reaches a random stationary limit, and, unlike populations with constant size, generally does not fixate. For a large ensemble of populations, this implies that the probability distribution for the population composition converges to limit distributions (Figs 2 and 4), which are nothing like Kimura"s result for constant-sized populations. Our result is also quite different from that obtained in range expansion experiments [37] or other settings featuring population growth without death on two-dimensional substrates. There, monoclonal sectoring patterns arise as a consequence of random genetic drift, which drives population differentiation along the expanding fronts of bacterial colonies, unlike our well-mixed populations that freeze to coexistence.

Our study also shows that, in a growing population with stochastic initial conditions, demographic noise has two possible sources: the initial sampling process by which subpopulations are formed, and the subsequent growth process. The initial average population size $\bar{N}_0$ sets

their relative weight (see S3 Text and S1 Fig). For very small $\bar{N}_0$, of the order of one or two individuals, the formation process already determines the final composition distribution: most populations start off fixated, many with just a single founder individual, and the composition of each well remains the same during growth. For very large $\bar{N}_0$, of the order of a few hundreds, the sampling process is again central: the composition distribution changes very little before freezing, and growth generates only a very limited amount of variation. In these two limiting regimes, neglecting stochastic effects during growth leaves the evolutionary outcome practically unchanged. In contrast, for small founder colonies such as those typically found during population bottlenecks [18, 19, 38] ($\bar{N}_0 \sim 10$), population growth is responsible for the major part of the variation observed in the final distribution.

Moreover, our results reveal that a growing population reaches a random limit composition much faster than genetic drift leads to fixation in populations of constant size. Typical fixation times for genetic drift increase logarithmically with the population size [11], while the time scale for freezing is independent of population size. This has important consequences for the role of stochastic effects when a population passes from exponential growth phase to stationary phase, in which growth rate and death rate are equal. Then, the composition of the population shows both freezing and fixation, albeit at quite distinct times because the relevant time scales differ markedly. During growth the composition distribution quickly freezes, as described above. Once the population reaches its stationary size, it slowly drifts to fixation, following Kimura-like dynamics.

We also believe that our results have a broad range of applications since growing populations are ubiquitous in nature. For example, experimental studies of *P. aeruginosa* [22, 39] have shown that typical life cycles pass through different steps with regularly occurring dispersal events being followed by the formation of new colonies. As initial colony sizes are typically small, such dispersal events coincide with *population bottlenecks* and subsequent exponential growth. During these phases of the life cycle, population dynamics is often selectively neutral and hence falls within the framework of the present work. The degree of diversity generated during these population bottlenecks has been shown to be crucial for some proposed mechanisms for the evolution of cooperation under selective pressure [5–7, 40–42]. Our analysis quantifies the ensuing degree of diversity and points to the relative importance of sampling versus growth for long-term behavior of the reservoirs. This may have important consequences for the degree of genetic diversity observed in natural populations with life-cycle structures [38].

Materials and Methods

Strains and cultivation conditions

The *P. putida* strains KT2440 (wild type) and 3E2 (mutant with defective pyoverdine synthesis) [3] were used as pyoverdine producers and non-producers, respectively. Cells were grown in casamino acid medium (CAA) containing per liter: 5 g casamino acids, 0.8445 g K_2HPO_4, 0.1404g $MgSO_4 \bullet (H_2O)$ [3]. The CAA medium was supplemented with 200 μM $FeCl_3$ (CAA-Fe) to suppress pyoverdine production (see S2 Table). Overnight cultures of the individual strains in CAA-Fe medium were adjusted to an OD_{600} of 1, diluted 10^{-2} fold, mixed to yield the desired producer fraction, and further diluted to create Poisson distribution conditions. Producer/non-producer co-cultures were started by inoculating the central 60 wells of two 96-well plates thereby adjusting the average initial cell number $\bar{N}_0$ to values between 2 and 25 cells/150 μL. Wells at the border of the plates were filled with water to minimize evaporation from central wells. For non-selective growth, co-cultures were grown in CAA-Fe medium shaking at 30°C for given periods of time. Due to the random distribution of initial cell number N_0 and producer faction x_0 in the 120 wells, each experiment was unique. An experiment was

limited to 120 wells to allow initiation of the analysis of the subpopulations in the individual wells without uncontrolled changes of growth parameters during analysis. The experiment duration was set to 11h to allow evolution to act for a significant number of generations (see S1 Table), while leaving bacteria in exponential growth phase (see S2 Fig)

Determination of growth parameters

Cell numbers N_0 and $N(t)$ were determined by counting the colony forming units (cfu) of individual wells. For this purpose $100\mu L$ aliquots of the individual wells were plated on cetrimide [43] or King's B agar (contains per liter: 20 g peptone, 10 g glycerol, 1.965 g $K_2HPO_4(3H_2O)$, 0.842 g $MgSO_4(H_2O)$ [44]. Producer fractions x_0 and $x(t)$ were determined based on the capability of cells to produce the green fluorescent pyoverdine either by direct counting of fluorescent and non-fluorescent colonies on the plates or after growth in iron-limited CAA medium. The fraction of dead cells was determined by life/dead staining with propidium iodide [45], and was always <0.02 of the total cell number under the experimental conditions.

Simulation of growing populations

We performed simulations of 10080 wells using a Gillespie algorithm [46]. The initial numbers of "cells" per well were drawn at random from a Poisson distribution with a mean value of $\bar{N}_0$ measured in the corresponding experiment. The strain assigned to every individual in each well was determined by the outcome of a Bernoulli trial (i.e., coin-flip-like process) and the probability of assignment to strain A was set to the value of $\bar{x}_0$ measured in the experiment. After initialization, wells were grouped into 84 virtual 120-well "plates", and a random waiting time was selected for each well, drawn from an exponential distribution with the population size as parameter. The Gillespie algorithm was run until the average size across all wells matched the average size measured at the end of the growth experiments, or until a specified time had elapsed (see S2 Fig).

Supporting Information

S1 Video. Time evolution of composition distribution. The distribution of compositions x first broadens due to demographic noise, but soon "freezes" to a steady state. The steady state form is maintained as long as the populations grow. Parameter values are $\bar{N}_0 = 10, \bar{x}_0 = 0.33$ (as for Fig 2).
(MP4)

S1 Text. Exact calculations for steady-state composition distribution and moments. Using the theory of Pólya urns, we calculate exactly the steady state values of: (i) the distribution of population compositions x, (ii) its mean value, and (iii) its variance.
(PDF)

S2 Text. Approximate calculations for the time evolution of the distribution moments. We use the Master equation of the growth process (Eq (1)) to determine the time evolution of variance and skewness of the composition distribution. These values are used in Eq (2) and Fig 2.
(PDF)

S1 Fig. Initial and steady state distributions, relative entropy. Panels (a),(b),(c): Initial and final distributions of x for three regimes of $\bar{N}_0$. When $\bar{N}_0$ is very small or very large (panels **(a)** and **(b)**), the evolutionary fate of the population is largely determined by the initial population sampling. Therefore, the initial distribution (red bars) and the steady-state one (green bars) look qualitatively very similar. For intermediate values of $\bar{N}_0$, however, population growth

becomes more important, and the distributions look very different. The amount of composition values the population can access through growth can be quantified looking at the "unpredictability" of the steady-state composition, once the initial one is known: the more unpredictable, the more are made accessible by growth. Mathematically, the measure for this is called *conditional entropy*: the higher the entropy, the more unpredictable the outcome. Panel **(d)** shows the conditional entropy as function of $\bar{N}_0$. Indeed, very small or very large initial populations experience little to no additional noise from growth, while in populations with intermediate values of $\bar{N}_0$ ($\bar{N}_0 \simeq 15$) growth is a major source of demographic noise. (Parameter values: $\bar{N}_0 = 2$ **(a)**, $\bar{N}_0 = 2000$ **(b)**, $\bar{N}_0 = 20$ **(c)**; $\bar{x}_0 = 0.25$ in all panels)
(TIF)

S3 Text. Comparison of initial and steady-state distributions of x, and entropy of the steady state distribution conditioned on the initial one. We use conditional entropy to analyze the impact of growth on the distribution of compositions *x*. The results are also depicted in S1(d) Fig
(PDF)

S2 Fig. Growth curve of a mixed population. The population consists of pyoverdine producer (*P. putida* KT2440) and non-producer (*P. putida* 3E2) under non-selective (iron replete) conditions. Individual precultures of the strains were mixed and diluted in iron replete medium to yield $\bar{N}_0 = 4$ (in 150 μL), and $\bar{x}_0 = 0.5$. Cells were grown aerobically at 30°C for 24 hours. The dots represent the mean $N(t)$ of three independent replications, the bars the corresponding standard deviation. After a lag phase of about 2 hours, the cells start to grow exponentially and reach the stationary phase after about 14 h of growth. For the non-selective growth experiments used to test the predictions of the Pólya urn model, cells were grown for 11.5 h to ensure exponential growth conditions.
(TIF)

S3 Fig. Additional initial conditions measurements. The experimental distributions (bars) are measured from 120-well ensembles, the average N_0 and x_0 from those sets the parameters for the simulated distributions. The theoretical average distribution (solid line) is the average of the same distributions generated for 84 sets of 120 wells. Using that average we calculate three Wilson binomial confidence intervals (gray areas). Experiments and theory agree within statistical error: the distribution of sizes (panels **(a)** and **(c)**) follows a Poisson distribution. The raggedness of the distribution of *x* for at small $\bar{N}_0$ (see panel **(b)** and Fig 3(b) in main text) is due to a small size effect: since *x* must be a simple fraction, when N_0 is small only a few values are available (see main text). This effect disappears for average initial sizes $\bar{N}_0 \simeq 10$ (see panel **(d)**). Parameter values: $\bar{N}_0 = 5.75$, $\bar{x}_0 = 0.43$ **(a)** and **(b)**; $\bar{N}_0 = 26.49$, $\bar{x}_0 = 0.45$ **(c)** and **(d)**.
(TIF)

S1 Table. Comparison between results from our experiments and those in [9]. While experiments for constant-sized populations of *Drosophila* observe significant fixations within the first tens of generations, we instead observe freezing of the probability distribution for the population composition, without any fixation.
(PDF)

S2 Table. Comparison of growth and pyoverdine production per cell of *P. putida* KT2440 and 3E2. Separate cultures of producer (*P. putida* KT2440) and non-producer (*P. putida* 3E2) were grown in iron-limiting (no addition of FeCl$_3$) and iron-replete medium (addition of 200 μM FeCl$_3$) at 30°C. The cell density was measured at 600 nm, and specific growth rates were calculated from density values of the exponential phase. The pyoverdine production was

determined by fluorescence emission measurements (excitation 400 nm, emission at 460 nm). The pyoverdine production per cell represents the ratio of pyoverdine fluorescence and optical density measured after 24 h of growth. The values in the table are averages over a minimum of five experiments, with the corresponding standard deviation. The fluorescence value for the non-producing mutant in iron-limiting medium is 0 because the culture failed to grow. (PDF)

Acknowledgments

Author Contributions

Conceived and designed the experiments: HJ FB. Performed the experiments: HJ FB. Analyzed the data: FB HJ ML KW EF. Contributed reagents/materials/analysis tools: EF ML KW. Wrote the paper: EF HJ ML KW. Designed theoretical analysis: EF ML KW. Performed theoretical analysis: ML KW.

References

1. Kimura M (1955) Solution of a process of random genetic drift with a continuous model. PNAS. PMID: 16589632

2. Otto SP, Whitlock MC (1997) The probability of fixation in populations of changing size. Genetics 146: 723–733. PMID: 9178020

3. Matthijs S, Laus G, Meyer JM, Abbaspour-Tehrani K, Schäfer M, et al. (2009) Siderophore-mediated iron acquisition in the entomopathogenic bacterium pseudomonas entomophila l48 and its close relative pseudomonas putida KT2440. Biometals 22: 951–964. doi: 10.1007/s10534-009-9247-y PMID: 19459056

4. Eggenberger F, Pólya G (1923) Über die statistik verketteter vorgänge. ZAMM—Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik 3: 279–289.

5. Melbinger A, Cremer J, Frey E (2010) Evolutionary game theory in growing populations. Phys Rev Lett 105: 178101. doi: 10.1103/PhysRevLett.105.178101 PMID: 21231082

6. Cremer J, Melbinger A, Frey E (2011) Evolutionary and population dynamics: a coupled approach. Phys Rev E 84: 051921. doi: 10.1103/PhysRevE.84.051921

7. Cremer J, Melbinger A, Frey E (2012) Growth dynamics and the evolution of cooperation in microbial populations. Sci Rep 2: 281. doi: 10.1038/srep00281 PMID: 22355791

8. Crow JF, Kimura M (1970) An Introduction to Population Genetics. Harper and Row.

9. Hartl DL, Clark AG (1989) Principles of population genetics. Sinauer, Sunderland, MD, USA, 2nd edition.

10. Frey E, Kroy K (2005) Brownian motion: a paradigm of soft matter and biological physics. Ann Phys 14: 20–50. doi: 10.1002/andp.200410132

11. Blythe RA, McKane J (2007) Stochastic models of evolution in genetics, ecology and linguistics. J Stat Mech: 07018. doi: 10.1088/1742-5468/2007/07/P07018

12. Caballero A (1994) Developments in the prediction of effective population size. Heredity 73: 657–679. doi: 10.1038/hdy.1994.174 PMID: 7814264

13. Gillespie JH (2001) Is the population size of a species relevant to its evolution? Evolution 55: 2161–2169. doi: 10.1554/0014-3820(2001)055%5B2161:ITPSOA%5D2.0.CO;2 PMID: 11794777

14. Hahn MW (2008) Toward a selection theory of molecular evolution. Evolution 62: 255–265. doi: 10.1111/j.1558-5646.2007.00308.x PMID: 18302709

15. Hurst LD (2009) Fundamental concepts in genetics: genetics and the understanding of selection. Nat Rev Genet 10: 83–93. doi: 10.1038/nrg2506 PMID: 19119264

16.   Buri P (1956) Gene frequency in small populations of mutant drosophila. Evolution 10: 367–402. doi: 10.2307/2406998

17.   Center for Food Safety and Applied Nutrition Causes of Foodborne Illness: Bad Bug Book—BBB—Escherichia coli O157:H7 (EHEC). Provides basic information about Escherichia coli O157:H7.

18.   Kurjak A, Chervenak FA (2006) Textbook of Perinatal Medicine. Informa UK Ltd, 2nd edition.

19.   Grossart HP, Levold F, Allgaier M, Simon M, Brinkhoff T (2005) Marine diatom species harbour distinct bacterial communities. Environmental Microbiology 7: 860–873. doi: 10.1111/j.1462-2920.2005.00759.x PMID: 15892705

20.   Bird DF, Kalff J (1984) Empirical relationships between bacterial abundance and chlorophyll concentration in fresh and marine waters. Can J Fish Aquat Sci 41: 1015–1023. doi: 10.1139/f84-118

21.   Buchan A, LeCleir GR, Gulvik CA, González JM (2014) Master recyclers: features and functions of bacteria associated with phytoplankton blooms. Nat Rev Micro 12: 686–698. doi: 10.1038/nrmicro3326

22.   Stoodley P, Sauer K, Davies DG, Costerton JW (2002) Biofilms as complex differentiated communities. Annual Review of Microbiology 56: 187–209. doi: 10.1146/annurev.micro.56.012302.160705 PMID: 12142477

23.   Levin SA (1974) Dispersion and population interactions. The American Naturalist 108: 207–228. doi: 10.1086/282900

24.   Sanchez A, Gore J (2013) Feedback between population and evolutionary dynamics determines the fate of social microbial populations. PLoS biology 11: e1001547. doi: 10.1371/journal.pbio.1001547 PMID: 23637571

25.   Parsons TL, Quince C, Plotkin JB (2010) Some consequences of demographic stochasticity in population genetics. Genetics 185: 1345–1354. doi: 10.1534/genetics.110.115030 PMID: 20457879

26.   Buckling A, Harrison F, Vos M, Brockhurst MA, Gardner A, et al. (2007) Siderophore-mediated cooperation and virulence in pseudomonas aeruginosa. FEMS Microbiol Ecol 62: 135–41. doi: 10.1111/j.1574-6941.2007.00388.x PMID: 17919300

27.   Cornelis P, Matthijs S, Van Oeffelen L (2009) Iron uptake regulation in pseudomonas aeruginosa. Biometals 22: 15–22. doi: 10.1007/s10534-008-9193-0 PMID: 19130263

28.   Cohen JE (1976) Irreproducible results and the breeding of pigs (or nondegenerate limit random variables in biology). BioScience 26: 391–394.

29.   Schreiber SJ (2001) Urn models, replicator processes, and random genetic drift. SIAM Journal on Applied Mathematics 61: 2148–2167. doi: 10.1137/S0036139999352857

30.   Sornette D (2009) Why Stock Markets Crash: Critical Events in Complex Financial Systems. Princeton University Press.

31.   Pemantle R (2007) A survey of random processes with reinforcement. Probability Surveys 4: 1–79. doi: 10.1214/07-PS094

32.   Athreya KB, Karlin S (1968) Embedding of urn schemes into continuous time markov branching processes and related limit theorems. The Annals of Mathematical Statistics 39: 1801–1817. doi: 10.1214/aoms/1177698013

33.   Kotz S, Mahmoud H, Robert P (2000) On generalized pólya urn models. Statistics & Probability Letters 49: 163–173.

34.   Allen L (2003) An Introduction to Stochastic Processes with Biology Applications. Prentice Hall.

35.   Arthur WB, Ermoliev YM, Kaniovski YM (1987) Path-dependent processes and the emergence of macro-structure. European Journal of Operational Research 30: 294–303. doi: 10.1016/0377-2217(87)90074-9

36.   Wilson EB (1927) Probable inference, the law of succession, and statistical inference. Journal of the American Statistical Association 22: 209–212. doi: 10.1080/01621459.1927.10502953

37.   Hallatschek O, Hersen P, Ramanathan S, Nelson DR (2007) Genetic drift at expanding frontiers promotes gene segregation. PNAS 104: 19926–19930. doi: 10.1073/pnas.0710150104 PMID: 18056799

38.   Hammerschmidt K, Rose CJ, Kerr B, Rainey PB (2014) Life cycles, fitness decoupling and the evolution of multicellularity. Nature 515: 75–79. doi: 10.1038/nature13884 PMID: 25373677

39.   Hall-Stoodley L, Costerton JW, Stoodley P (2004) Bacterial biofilms: from the natural environment to infectious diseases. Nat Rev Microbiol 2: 95–108. doi: 10.1038/nrmicro821 PMID: 15040259

40.   Chuang JS, Rivoire O, Leibler S (2009) Simpson's paradox in a synthetic microbial system. Science 323: 272–5. doi: 10.1126/science.1166739 PMID: 19131632

41.   Oliveira NM, Niehus R, Foster KR (2014) Evolutionary limits to cooperation in microbial communities. PNAS 111: 17941–17946. doi: 10.1073/pnas.1412673111 PMID: 25453102

42.   Melbinger A, Cremer J, Frey E The emergence of cooperation from a single cooperative mutant.

43. Brown VI, Lowbury EJL (1965) Use of an improved cetrimide agar medium and other culture methods for pseudomonas aeruginosa. J Clin Pathol 18: 752–756. doi: 10.1136/jcp.18.6.752 PMID: 4954265

44. King EO, Ward MK, Raney DE (1954) Two simple media for the demonstration of pyocyanin and fluorescin. Journal of Laboratory and Clinical Medicine 44: 301–307. PMID: 13184240

45. Suzuki T, Fujikura K, Higashiyama T, Takata K (1997) DNA staining for fluorescence and laser confocal microscopy. J Histochem Cytochem 45: 49–53. doi: 10.1177/002215549704500107 PMID: 9010468

46. Gillespie D (1976) General method for numerically simulating stochastic time evolution of coupled chemical reactions. J Comp Phys 22: 403–434.

Non-selective evolution of growing populations
Supplementary Information

Karl Wienand, Matthias Lechner, Felix Becker, Heinrich Jung, and Erwin Frey

Video: Time evolution of composition distribution.

`https://vimeo.com/108884249`

The distribution of compositions x first broadens due to demographic noise, but soon "freezes" to a steady state. The steady state form is maintained as long as the populations grow. Parameter values are $\bar{N}_0 = 10$, $\bar{x}_0 = 0.33$ (as for Fig. 2).

Exact calculations for steady-state composition distribution and moments.

Calculation of probability distribution.

Each population in the ensemble is initialized with A_0 individuals of type A and B_0 of type B. In the general case, A_0 and B_0 are independent random variables for each population with distributions $P(A_0)$ and $P(B_0)$. All populations evolve for ΔN reproduction events, of which a random amount ΔA generate new A-individuals. From the mathematical literature [35], it is well-known that ΔA follows a beta-binomial, with A_0, B_0 and ΔN as parameters. The fraction of A-individuals x, then follows the probability

$$P(x) = \sum_{A_0, B_0} P(A_0) P(B_0) P(\Delta A = x(A_0 + B_0 + \Delta N) - A_0 | A_0, B_0, \Delta N), \tag{3}$$

where the sums run over all allowed values of their respective indices. $P(\Delta A = k | A_0, B_0, \Delta N)$ is the probability of ΔA being equal to k, given the values of A_0, B_0 and ΔN. The sum may easily be performed numerically. For the moments of the distribution there are, however, also closed-form analytic expressions.

Exact calculation of asymptotic moment values

Let $\langle \cdot \rangle_0$ be the average over the initial conditions, $\langle \cdot \rangle_{\Delta A}$ be an average over ΔA, and $\langle \cdot \rangle$ be an average over both quantities. From the properties of the beta-binomial distribution we know that, for a given initial condition, we have

$$\langle \Delta A \rangle_{\Delta A} = \frac{\Delta N A_0}{A_0 + B_0}, \tag{4}$$

$$\mathrm{Var}[\Delta A] = \frac{\Delta N A_0 B_0 (A_0 + B_0 + \Delta N)}{(A_0 + B_0)^2 (A_0 + B_0 + 1)}. \tag{5}$$

For the mean of $\langle x \rangle$, one obtains

$$\langle x \rangle \overset{(4)}{=} \langle x_0 \rangle_0 = \bar{x}_0.$$

Hence, the average composition is exactly conserved throughout the time evolution of the populations. In other words, the stochastic process is a martingale.

For the variance we obtain

$$\text{Var}[x] = \left\langle \left(\frac{A_0 + \Delta A}{A_0 + B_0 + \Delta N} \right)^2 \right\rangle - \langle x_0 \rangle_0^2 \tag{6}$$

$$= \left\langle \frac{A_0^2 + 2A_0 \langle \Delta A \rangle_{\Delta A} + \text{Var}[\Delta A] + \langle \Delta A \rangle_{\Delta A}^2}{(A_0 + B_0 + \Delta N)^2} \right\rangle_0 - \langle x_0 \rangle_0^2 \tag{7}$$

$$\overset{(4)}{=} \left\langle \left(\frac{A_0}{A_0 + B_0} \right)^2 + \frac{\text{Var}[\Delta A]}{(A_0 + B_0 + \Delta N)^2} \right\rangle_0 - \langle x_0 \rangle_0^2 \tag{8}$$

$$\overset{(5)}{=} \text{Var}[x_0] + \left\langle \frac{\Delta N A_0 B_0}{(A_0 + B_0)^2(A_0 + B_0 + \Delta N)^2(A_0 + B_0 + 1)} \right\rangle_0 \tag{9}$$

$$= \text{Var}[x_0] + \langle x_0(1 - x_0) \rangle_0 \left\langle \frac{1}{N_0 + 1} \frac{\Delta N}{N_0 + \Delta N} \right\rangle_0 . \tag{10}$$

For long times (i.e., $\Delta N \gg 1$), $\Delta N + N_0 \simeq \Delta N$ and (10) reduces to

$$\text{Var}[x] \to \text{Var}[x_0] + \left\langle \frac{1}{N_0 + 1} \right\rangle_0 \langle x_0(1 - x_0) \rangle_0 . \tag{11}$$

The argument up to here is completely independent of the particular choice of initial conditions. If the initial distribution is known, we may even make the value of the variance more explicit. In particular, consider the distribution we obtain from experiments: in each well, N_0 is Poisson-distributed with mean $\bar{N}_0$. Then one gets

$$\left\langle \frac{1}{N_0 + 1} \right\rangle_0 = \frac{1 - e^{-\bar{N}_0}}{\bar{N}_0} . \tag{12}$$

Within each well of (random) size N_0 there is an initial random number A_0 of A-individuals, which follows a Binomial distribution with parameters N_0 and $\bar{x}_0$. For this choice of distribution, it is possible that $N_0 = 0$, which would lead to an undetermined value of $x_0 = A_0/N_0$, and therefore also for the average $\langle x_0 \rangle$. We can solve this problem by redefining x_0:

$$x_0 := \begin{cases} \bar{x}_0 & , N_0 = 0 \\ \frac{A_0}{N_0} & , \text{otherwise} \end{cases} \tag{13}$$

so that x_0 and its average have definite values, and $\langle x_0 \rangle_0 = \bar{x}_0$. With this we can compute the second moment of x_0:

$$\langle x_0^2 \rangle_0 = \sum_{N_0=1}^{\infty} e^{-\bar{N}_0} \frac{\bar{N}_0^{N_0}}{N_0!} \left\{ \sum_{A_0=0}^{N_0} \binom{N_0}{A_0} \bar{x}_0^{A_0}(1 - \bar{x}_0)^{N_0 - A_0} \frac{A_0^2}{N_0^2} \right\} + \bar{x}_0^2 e^{-\bar{N}_0} . \tag{14}$$

The sum inside the braces can be solved using exponential and binomial series and yields

$$\langle x_0^2 \rangle_0 = \bar{x}_0^2 + \bar{x}_0(1 - \bar{x}_0)e^{-\bar{N}_0} \sum_{N_0=1}^{\infty} \frac{\bar{N}_0^{N_0}}{N_0! N_0} . \tag{15}$$

The remaining series is an exponential integral (Ei), and therefore

$$\text{Var}[x_0] = \bar{x}_0(1 - \bar{x}_0)e^{-\bar{N}_0} \left[\text{Ei}(\bar{N}_0) - \gamma - \ln(\bar{N}_0) \right] =: \bar{x}_0(1 - \bar{x}_0)\varphi(\bar{N}_0) , \tag{16}$$

where we defined $\varphi(\bar{N}_0) := e^{-\bar{N}_0} \left[\text{Ei}(\bar{N}_0) - \gamma - \ln(\bar{N}_0) \right]$. Then the variance of x reads

$$\text{Var}[x] = \text{Var}[x_0] + \frac{1 - e^{-\bar{N}_0}}{\bar{N}_0} \langle x_0(1 - x_0) \rangle \tag{17}$$

$$= \bar{x}_0(1 - \bar{x}_0) \left[\varphi(\bar{N}_0) + \frac{1 - e^{-\bar{N}_0}}{\bar{N}_0} \left(1 - \varphi(\bar{N}_0) \right) \right] . \tag{18}$$

2

For large $\bar{N}_0$, through an asymptotic expansion [47],

$$\text{Ei} \simeq \frac{1}{\bar{N}_0} e^{\bar{N}_0} \sum_{m=0}^{\bar{N}_0-1} m! \bar{N}_0^{-m} - \frac{1}{3}\sqrt{\frac{2\pi}{\bar{N}_0}}, \tag{19}$$

$\varphi(\bar{N}_0)$ can be approximated by

$$\varphi(\bar{N}_0) \simeq \frac{1}{\bar{N}_0} \sum_{m=0}^{\bar{N}_0-1} m! \bar{N}_0^{-m} - e^{-\bar{N}_0}\left[\frac{1}{3}\sqrt{\frac{2\pi}{\bar{N}_0}} - \gamma - \ln(\bar{N}_0)\right]. \tag{20}$$

To leading order in $\bar{N}_0$, then, the variance of x becomes

$$\text{Var}[x] = \bar{x}_0(1 - \bar{x}_0)\frac{2}{\bar{N}_0},$$

in perfect agreement with our approximate results based on Master equations (Eq. (2) in main text, see also below).

Approximate calculations for the time evolution of the distribution moments

Using the Master Equation for the number of individuals of each strain (1), we are able to obtain the time evolution of the first three moments of the distribution of x. Equation (1) is sometimes called "Simple Growth Equation" and can be exactly solved (see, for example, [38]) using generating functions like

$$F(a,t) := \sum_{N_A} a^{N_A} P(N_A, t). \tag{21}$$

To approximate the time evolution of the first three moments of x, however, we do not need the full solution, but only the first three moments of N_A and N_B. To this end, we insert the Master Equation (Eq. (1) in main text) in the definition of the generating function to get the time derivative for $F(a,t)$:

$$\frac{d}{dt}F(a,t) = \left(-a + a^2\right)\partial_a F(a,t). \tag{22}$$

To obtain the time evolution of the nth moment, we apply the nth derivative with respect to a on both sides of equation (22), and solve for the corresponding moment. For the first three moments, the solution is

$$\langle N_A \rangle = e^t K_1, \tag{23}$$
$$\langle N_A^2 \rangle = e^t(e^t - 1)K_1 + e^{2t}K_2, \tag{24}$$
$$\langle N_A^3 \rangle = e^t\left(-3e^t + 2e^{2t} + 1\right)K_1 + 3e^{2t}\left(e^t - 1\right)K_2 + e^{3t}K_3. \tag{25}$$

K_1, K_2, K_3 are integration constants, which depend on the initial conditions. We consider the case of Poisson initial conditions. This means that the initial number of A is Poisson-distributed with mean value $\bar{N}_{A,0}$,

$$\langle N_A(t=0) \rangle \overset{!}{=} \bar{N}_{A,0}, \tag{26}$$

and, since for the Poisson distribution the variance equals the mean, we get

$$\operatorname{Var} N_A(t=0) \stackrel{!}{=} \bar{N}_{A,0} \,. \tag{27}$$

Employing these conditions in the solutions of the differential equations we found in Eq. (23) and (24), we get

$$
\begin{align}
\langle N_A \rangle &= \mathrm{e}^t \bar{N}_{A,0} \,, \tag{28} \\
\operatorname{Var} N_A &= \mathrm{e}^t (2\mathrm{e}^t - 1) \bar{N}_{A,0}. \tag{29}
\end{align}
$$

By the known properties of the Poisson distribution, the skewness of our initial distribution equals to $1/\sqrt{\bar{N}_{A,0}}$. Using Eqs. (25), (28), and the definition of the skewness, we obtain the general time evolution of the skewness

$$v(N_A) = \frac{\bar{N}_{A,0} \left(6e^{2t} - 6e^t + 1 \right) e^t}{\left(\bar{N}_{A,0} \left(2e^t - 1 \right) e^t \right)^{3/2}} \,. \tag{30}$$

For N_B, the calculations are analogous. Note also that all calculations were exact so far.

With the moments of N_A and N_B we can find the (approximate) time evolution of variance and skewness of $x = N_A/(N_A + N_B)$. For the mean of x we have already seen in the exact calculation (see Eq. (6)) that it does not change with time, and hence its time evolution is already known.

To calculate the time evolution of the variance of x, we consider x as a function of N_A and N_B:

$$x(N_A, N_B) = \frac{N_A}{N_A + N_B}. \tag{31}$$

Using the time independence of the mean ($\langle x(N_A, N_B) \rangle = x(\langle N_A \rangle, \langle N_B \rangle)$), a bivariate Taylor expansion around $(\langle N_A \rangle, \langle N_B \rangle)$, and the time evolution of the moments, Eqs. (28) and (30), we get for the variance of x:

$$
\begin{align}
\operatorname{Var} x &= \langle [x(N_A, N_B) - \langle x(N_A, N_B) \rangle]^2 \rangle \tag{32} \\
&= \left\langle [x(N_A, N_B) - x(\langle N_A \rangle, \langle N_B \rangle)]^2 \right\rangle \tag{33} \\
&= \langle [x'_{N_A}(\langle N_A \rangle, \langle N_B \rangle)(N_A - \langle N_A \rangle) + \tag{34} \\
&\quad + x'_{N_B}(\langle N_A \rangle, \langle N_B \rangle)(N_B - \langle N_B \rangle) + \mathcal{O}\left(N_A^{-2}, N_B^{-2} \right)]^2 \rangle \tag{35} \\
&= \frac{\langle N_B \rangle^2}{\langle N \rangle^4} \operatorname{Var} N_A + \frac{\langle N_A \rangle^2}{\langle N \rangle^4} \operatorname{Var} N_B + \mathcal{O}\left(N_A^{-2}, N_B^{-2} \right) \tag{36} \\
&= \frac{(2 - \mathrm{e}^{-t})}{N_0^4} N_{B,0} N_{A,0} \left(N_{A,0} + N_{B,0} \right) + \mathcal{O}\left(N_A^{-2}, N_B^{-2} \right) \tag{37} \\
&= \frac{2 - \mathrm{e}^{-t}}{\bar{N}_0} \bar{x}_0 (1 - \bar{x}_0) \tag{38} \\
&\xrightarrow[t \to \infty]{} \frac{2}{\bar{N}_0} \bar{x}_0 (1 - \bar{x}_0) \tag{39}
\end{align}
$$

From this we obtained Eq. (2) in main text. For infinite times the approximate result for the variance matches the exact one of Eq. (11).

The skewness of the x distribution is calculated analogously:

$$v(x) = \left\langle \left(\frac{x(N_A, N_B) - x(\langle N_A \rangle, \langle N_B \rangle)}{\sqrt{\mathrm{Var}\, x}} \right)^3 \right\rangle \tag{40}$$

$$= \frac{x_0 e^{-2t} \left(12 x_0^2 e^{2t} - 12 x_0^2 e^t + 2 x_0^2 \right)}{N_0^2 \left(\frac{x_0}{N_0} \left(-2 x_0 e^t + x_0 + 2 e^t - 1 \right) e^{-t} \right)^{1.5}}$$

$$+ \frac{x_0 e^{-2t} \left(-18 x_0 e^{2t} + 18 x_0 e^t - 3 x_0 + 6 e^{2t} - 6 e^t + 1 \right)}{N_0^2 \left(\frac{x_0}{N_0} \left(-2 x_0 e^t + x_0 + 2 e^t - 1 \right) e^{-t} \right)^{1.5}} + \mathcal{O}\left(N_A^{-2}, N_B^{-2} \right) . \tag{41}$$

Comparison of initial and steady-state distributions of x, and entropy of the steady state distribution conditioned on the initial one.
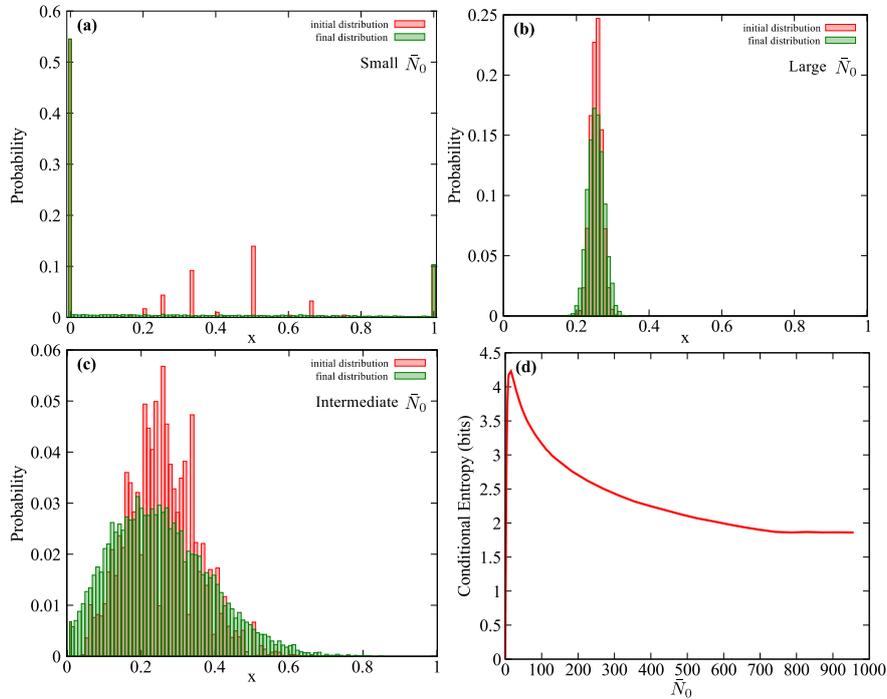
Figure 5: **Initial and steady state distributions, relative entropy. Panels (a),(b),(c)**: Initial and final distributions of x for three regimes of $\bar{N}_0$. When $\bar{N}_0$ is very small or very large (panels **(a)** and **(b)**), the evolutionary fate of the population is largely determined by the initial population sampling. Therefore, the initial distribution (red bars) and the steady-state one (green bars) look qualitatively very similar. For intermediate values of $\bar{N}_0$, however, population growth becomes more important, and the distributions look very different. The amount of composition values the population can access through growth can be quantified looking at the "unpredictability" of the steady-state composition, once the initial one is known: the more unpredictable, the more are made accessible by growth. Mathematically, the measure for this is called *conditional entropy*: the higher the entropy, the more unpredictable the outcome. Panel **(d)** shows the conditional entropy as function of $\bar{N}_0$. Indeed, very small or very large initial populations experience little to no additional noise from growth, while in populations with intermediate values of $\bar{N}_0$ ($\bar{N}_0 \simeq 15$) growth is a major source of demographic noise. (Parameter values: $\bar{N}_0 = 2$ **(a)**, $\bar{N}_0 = 2000$ **(b)**, $\bar{N}_0 = 20$ **(c)**; $\bar{x}_0 = 0.25$ in all panels)

We simulate an ensemble of populations starting from Poisson initial conditions, and track their time evolution until the x distribution freezes. Once it freezes, we can build a joint histogram of initial and final compositions, which approximates the joint distribution $P_{\text{joint}}(x_0, x_f)$. From P_{joint} we can obtain the initial and final distributions as its marginal distributions, integrating over all values of x_f and x_0, respectively. The joint information (Shannon) entropy is defined as [48]

$$H_{\text{joint}}(x_0, x_f) = - \int_0^1 \mathrm{d}x_0 \mathrm{d}x_f P_{\text{joint}}(x_0, x_f) \log(P_{\text{joint}}(x_0, x_f)). \tag{42}$$

The marginal entropies $H(x_0)$ and $H(x_f)$ are defined, analogously, through integrals only of $P(x_0)$ over x_0, and $P(x_f)$ over x_f, respectively. The conditional entropy of the final distribution given the initial is defined as

$$H(x_f|x_0) = H_{\text{joint}}(x_0, x_f) - H(x_0). \tag{43}$$

It measures the amount of information necessary to describe the final distribution, once all information about the distribution of x_0 is known. Therefore, $H(x_f|x_0)$ provides a measure of how entropic (or "noisy") growth itself is [49]—or, in other words, how many different final compositions are possible given the initial condition. Figure 5(d) shows $H(x_f|x_0)$ from repeated simulations, all with the same initial distribution form, the same $\bar{x}_0$, but different $\bar{N}_0$. For very small $\bar{N}_0$ (of the order of one or two individuals) the group formation almost completely determines the fate of populations: most populations start fixated, many with just a single founder individual, and the composition of each well remains the same during growth. The path followed by x in each population during time is a straight line, as the compositions stay constant. Therefore, x for different populations follow in time paths that do not cross or "mix". Growth produces very little demographic noise, and its conditional entropy tends to zero. For very large $\bar{N}_0$ (of the order of a few hundreds), the group sampling is again central to determine the final distribution. Very large populations, in fact, all start with similar compositions (according to the Law of Large Numbers), and their compositions are difficult to change, as each individual event has little impact. The composition distribution changes very little before freezing; time evolution paths of different populations "mix" very little. Entropy in this regime saturates for increasing initial sizes, and is rather low. Between the small size regime (where paths do not "mix") and the large size regime (where size limits "mixing"), we find a window where populations are small enough to significantly change their composition, but also large enough to not start fixated. This is the region where the conditional entropy peaks, and growth is the most important in determining the final distribution.

Intuitively, the difference in variance between initial and final distribution could provide an alternative measure of the noise introduced by growth. However, of all x distributions between 0 and 1 with fixed $\bar{x}_0$, the one with maximal variance is the one for which x is only 0 or 1, i.e., when all populations start off fixated. In this case, the compositions never change during growth and the variance stays constant. Moreover, independently on the choice of initial distribution, the difference between initial and steady-state variance decreases for increasing $\bar{N}_0$ (see Eq.(11)). Therefore, all considerations on noise sources based on variance would indicate that growth matters more when initial populations are smaller, in contrast with our observations.

	Drosophila	*P.Putida*
# of populations	107	120
Initial pop. size	16	~ 10
Max. # of generations	19	16
Pop. size	Constant	Growing
Outcome	Increasing number of fixations	No fixation, freezing

Table 1: **Comparison between results from our experiments and those in [9].** While experiments for constant-sized populations of *Drosophila* observe significant fixations within the first tens of generations, we instead observe freezing of the probability distribution for the population composition, without any fixation.
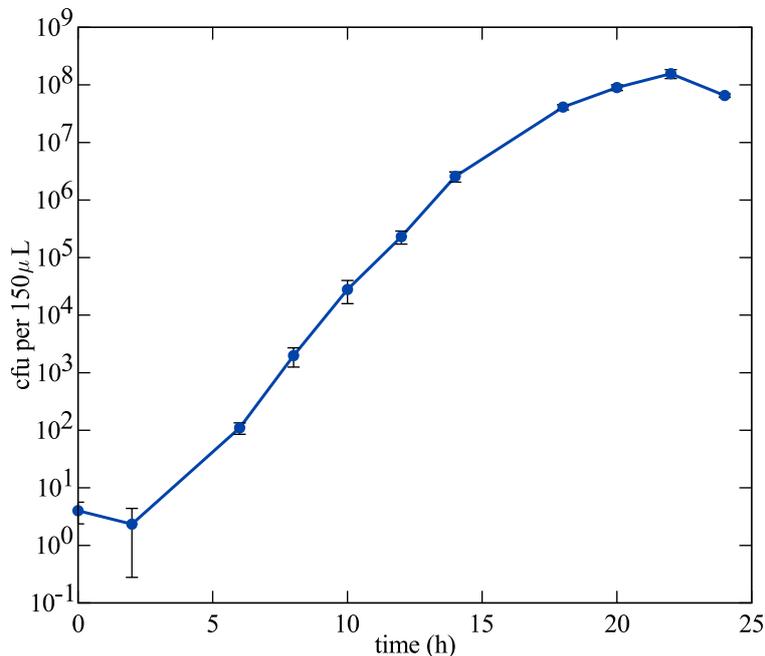
Figure 6: **Growth curve of a mixed population.** The population consists of pyoverdine producer (*P. putida* KT2440) and non-producer (*P. putida* 3E2) under non-selective (iron replete) conditions. Individual precultures of the strains were mixed and diluted in iron replete medium to yield $\bar{N}_0 = 4$ (in 150 μL), and $\bar{x}_0 = 0.5$. Cells were grown aerobically at 30°C for 24 hours. The dots represent the mean $N(t)$ of three independent replications, the bars the corresponding standard deviation. After a lag phase of about 2 hours, the cells start to grow exponentially and reach the stationary phase after about 14 h of growth. For the non-selective growth experiments used to test the predictions of the Pólya urn model, cells were grown for 11.5 h to ensure exponential growth conditions.

	Specific growth rate (h^{-1})		Fluorescence per cell (a.u.)	
Iron conc. (μM)	*KT2440*	*3E2*	*KT2440*	*3E2*
0	0.058±0.006	*no growth*	244.00 ± 21.3	0 ± 0
200	0.152±0.026	0.146±0.017	1.56 ± 0.27	0.93 ± 0.10

Table 2: **Comparison of growth and pyoverdine production per cell of *P. putida* KT2440 and 3E2.** Separate cultures of producer (*P. putida* KT2440) and non-producer (*P. putida* 3E2) were grown in iron-limiting (no addition of FeCl$_3$) and iron-replete medium (addition of 200 μM FeCl$_3$) at 30°C. The cell density was measured at 600 nm, and specific growth rates were calculated from density values of the exponential phase. The pyoverdine production was determined by fluorescence emission measurements (excitation 400 nm, emission at 460 nm). The pyoverdine production per cell represents the ratio of pyoverdine fluorescence and optical density measured after 24 h of growth. The values in the table are averages over a minimum of five experiments, with the corresponding standard deviation. The fluorescence value for the non-producing mutant in iron-limiting medium is 0 because the culture failed to grow.

Figure 7: **Additional initial conditions measurements.** The experimental distributions (bars) are measured from 120-well ensembles, the average N_0 and x_0 from those sets the parameters for the simulated distributions. The theoretical average distribution (solid line) is the average of the same distributions generated for 84 sets of 120 wells. Using that average we calculate three Wilson binomial confidence intervals (gray areas). Experiments and theory agree within statistical error: the distribution of sizes (panels **(a)** and **(c)**) follows a Poisson distribution. The raggedness of the distribution of x for at small $\bar{N}_0$ (see panel **(b)** and Fig. 3**(b)** in main text) is due to a small size effect: since x must be a simple fraction, when N_0 is small only a few values are available (see main text). This effect disappears for average initial sizes $\bar{N}_0 \simeq 10$ (see panel **(d)**). Parameter values: $\bar{N}_0 = 5.75$, $\bar{x}_0 = 0.43$ **(a)** and **(b)**; $\bar{N}_0 = 26.49$, $\bar{x}_0 = 0.45$ **(c)** and **(d)**.

Part II

Phenotypic Heterogeneity in ColicinE2 Release

2

Post-transcriptional Regulation of ColicinE2 Expression in Escherichia Coli

Project Abstract and Contributions   The secretion of toxins is a potent mechanism for bacteria to kill other strains in the competition for resources. In this project, I analysed (together with Mathias Schwarz and Erwin Frey) the regulatory network (see 2.1.2) of the toxin ColicinE2. The ColicinE2 system is activated by stochastic SOS responses (see 2.1.3), and regulates the production and release of the toxin using three hierarchically ordered components (see 2.1.4). We developed a rate equation model for this system by introducing effective quantities and reducing the complexity of the regulatory network with a time scale separation approach. A parameter analysis of the resulting system revealed that the hierarchical nature of the network reduces internal fluctuations, and allows for a fine-tuned response to SOS signals. Moreover, we coupled our reduced system to a stochastic SOS response model and reproduced broad lysis time distributions. These distributions are also found in experiments from the group of Madeleine Opitz. Our work on this model and its analysis have been published in the research paper "Hierarchical Post-transcriptional Regulation of Colicin E2 Expression in *Escherichia coli*", which is reprinted in section 2.5. In a subsequent experimental study on five different strains containing the ColicinE2 system, Alexandra Götz and Madeleine Opitz found experimental evidence that single stranded DNA (ssDNA) in the cells acts as fourth regulative component. Together with Erwin Frey, I extended the model for ColicinE2 regulation with this fourth component, and could show that the additional regulator is indeed necessary to reproduce the experimental results. Moreover, we complemented the model analysis by investigating the influence of the different plasmid compositions in the investigated strains. The extension of the model with ssDNA, as well as the analysis of the experimental and theoretical results, are detailed in the paper draft "CsrA and its regulators control the time-point of ColicinE2 release in *Escherichia coli*", which is reprinted in section 2.6 (and also in the PhD thesis of Alexandra Götz).

A: ColicinE2 production and release system:

B: hierarchical regulation:

Figure 2.1 | Graphical abstract of this project: (A) SOS responses (see 2.1.3) trigger transcription in the ColicinE2 system (for details, see 2.1.4), which produces two different mRNA transscripts. **(B)** This transcript is hierarchically regulated by CsrA (see 2.1.4) and, as we show later, also sRNA. Starting from these interactions, we developed the **(C)** reduced ColicinE2 model (see 2.3), which we finally **(D)** coupled to a stochastic SOS response system. Triggering this system with an SOS signal resulted in a broad distribution of lysis times (see 2.3).

2.1 Background

2.1.1 Gene Expression in Prokaryotes

Proteins play a fundamental role in virtually every process in living cells [20]. Despite the large number of different proteins, all are built from only 20 different amino acids. The "building plan" to create a protein from these "building blocks" is stored on DNA in form of genes. In bacteria, which are our object of study here, DNA is either located in the nucleoid, or in form of small DNA rings called *plasmids*. For many functions in cells, several proteins are required. Therefore, the necessary genes are often grouped together on the DNA, and form a functional unit known as *operon*.

plasmid

operon

The synthesis of proteins from the genetic information occurs in a complex process, which is called *gene expression*. It comprises two main steps: First, the gene sequence is *transcribed* to a mRNA molecule, which then is *translated* into the protein. Transcription starts when the enyzme RNA polymerase binds to the so called *promoter* region, which lies at the beginning of every operon. It then moves along the DNA, producing a mRNA transcript of the DNA sequences it passes. The RNA polymerase continues until it reaches a transcriptional terminator, where the mRNA is released to the cytosol. The translation of this mRNA to a protein then happens in ribosomes [20]

transcription

translation

2.1.2 Gene Regulation

The aforementioned fundamental role of proteins makes it obvious that their production needs to be regulated. To make sure that the right amount of proteins is synthesised at the right time, a plethora of gene regulatory mechanisms exists. These mechanisms fall into two categories: They either act transcriptional, or post-transcriptional [20].

Transcriptional regulation controls the production of mRNA transcripts. An important example are *repressor* proteins, which bind to the promoter region and thus prevent or impede the polymerase from binding to the DNA. Post-transcriptional mechanisms, however, act on already produced mRNA. A common example are proteins binding to mRNA, often referred to as *sequestration*. The resulting mRNA-protein complexes then cannot be translated to proteins anymore.

transcriptional

repressor

post-transcriptional

sequestration

In general, several regulatory mechanisms act in concert to synthesize the correct amount of proteins needed in the cell. Once produced, the proteins can act in turn as regulators of gene expression, including their own. The interactions of various regulating proteins can become very complex, and include structures like feedback loops. An example for such a so called *regulatory network* is the SOS Response System.

regulatory network

2.1.3 SOS Response System

Gene regulation mechanisms play a crucial role in the reaction to environmental signals like nutrient availability, radiation or population density. Here, we consider the SOS response system [22], which is activated when the cell suffers DNA damage due to severe environmental stress, and triggers repair and defence mechanisms in the cell (for an example, see 2.1.4). The system involves two proteins, RecA and LexA [23, 24]. LexA acts as a transcriptional repressor to both itself and RecA, but it also represses the expression of several repair and defence systems (see Fig. 2.2). Therefore, when the cell is in a normal non-SOS state, LexA keeps its level just high enough to keep RecA as well as the repair and defence proteins on a low basal level.

SOS response

In the case of DNA damage, however, RecA proteins become catalysts for LexA cleavage [21, 23], causing the LexA levels to go down. Consequently, the cell starts to react to the stress, as the levels of LexA-repressed proteins go up. This includes RecA, thus reinforcing the catalytic effect. As LexA is autoregulated, its production also increases. Eventually, LexA production becomes so high that the RecA catalysis saturates, and LexA pushes the synthesis of all involved proteins back to the initial level. This cycle continues as long as the DNA is
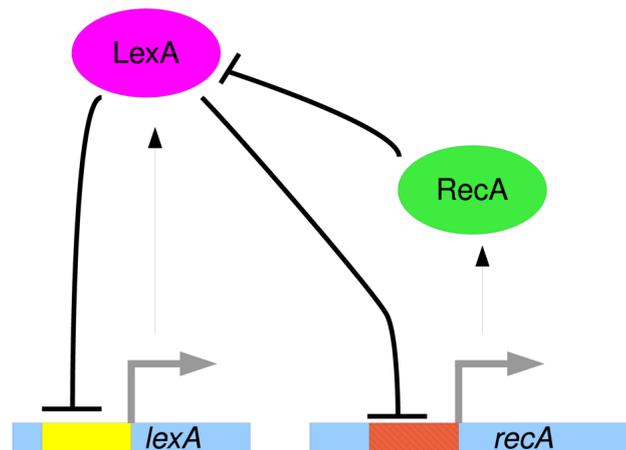
Figure 2.2 | Diagram of the regulatory network of the SOS response system: The regulatory network consists of the proteins LexA and RecA, and the corresponding genes, *lexA* and *recA*. Flat-headed arrows represent negative regulation, normal arrows represent production. For a detailed description of the dynamics, see 2.1.3. Image detail taken from Ref. [21]

damaged.

From these regulatory interactions we can see the interesting property that, during SOS responses, the systems repressed by LexA are not activated continuously, but in bursts. Experiments [25] (and ensuing mathematical models [21]) also showed that these bursts indeed occur stochastically when the SOS system is triggered.

2.1.4 The Bacteriocin ColicinE2 and its Production and Release System

In the subsection 2.1.3 we discussed how SOS responses trigger repair and defence mechanisms. An important class of defence mechanisms is the release of toxins, referred to as bacteriocins [26]. These are proteins (or peptides) produced by bacteria and released to the environment to kill closely related bacteria. Along with the toxin, every bacteriocin producing cell also has to synthesize immunity proteins, which prevent them from poisoning themselves and cells of their own strain.

In this project, we consider the bacteriocin ColicinE2, which is produced by some strains of *E. coli* [27]. The genetic information for the ColicinE2 system is located in a single operon on specific plasmids[28] (see also Fig. 2.1A). The promoter of the ColicinE2 operon is repressed by LexA, and thus connected to the SOS response system. Downstream from the promoter, the operon contains three genes: First, the gene for the toxin itself, second a gene for the immunity protein, and third a gene for a protein inducing cell lysis [29]. The latter is necessary as ColicinE2 cannot be secreted through the cell wall (like many other colicins). This means that ColcinE2 producing cells must die when they want to release their toxin into the environment. Apart from the promoter and the genes, the operon also has two transcriptional terminators, located before and after the lysis gene. Therefore, two different mRNA are produced during

transcription: A *short mRNA* with the toxin and immunity protein genes, and a *long mRNA* containing all three genes [30]. Most of the transsscripts are short mRNA.

As the transcription of the long mRNA has fatal consequences for the cell, its translation is strongly regulated (in contrast to the short mRNA). This happens via a specific DNA sequence (Shine-Dalgarno sequence), to which the regulator protein CsrA binds [30] (see also Fig. 2.1B). CsrA is highly abundant in *E. coli* cells, as it plays a major role in many regulatory networks[31, 32]. Previous studies found that the main sequestering element of CsrA are short RNA molecules (*sRNA*), which thus act as regulator of CsrA [31–36]. This means that long mRNA is subject to a hierarchical form of regulation: It is regulated by CsrA, which itself is regulated by sRNAs [37].

2.2 Motivation and Research Question

In the group of Madeleine Opitz, various aspects of ColicinE2 producing bacteria have been studied, for instance in the context of range expansions of cyclic dominant strains [38].The regulatory network was also investigated, with a particular focus on how the amount of released toxin is connected to the lysis protein [39]. These studies highlighted that, in many situations, ColicinE2 release exhibits stochastic effects like heterogeneous lysis times. Moreover, as already mentioned in subsection 2.1.3, the SOS response system is known to produce stochastic signals to the cell, rather then continuous activation. It was not known how such a stochastic signal is transduced through the regulatory network of ColicinE2, and what consequences this has for the release of the toxins. This raised the first main research question of this project:

How does stochasticity affect ColicinE2 regulation?

As pointed out in 2.1.4, the genetic code for ColicinE2 is located on plasmids. During the experiments for different ColicinE2-related projects, strains with different plasmid compositions were created. Interestingly, these strains exhibited differences in the delays between toxin and lysis protein production, even though they were genetically identical otherwise. This result suggested that ssDNA, which accumulates as an intermediate in the plasmid replication [40], could be a so far unknown regulator in the ColicinE2 system, leading to our second research question:

What is the role of ssDNA in ColicinE2 regulation?

2.3 Summary of Results

Simplified, comprehensive mathematical model. We collected the known interactions of the ColicinE2 regulatory network and formulated it as coupled rate equations of the regulatory elements and their complexes. As the resulting system of equations is very large and unhandy, we simplified it in several steps. First, we determined the mean number of occupied binding sites of the sRNAs, and broke the sRNAs up into effective sRNAs with a single binding site. Second, we assumed fast complex equilibration, and introduced a coupled degradation rates, k_M and k_S, for the complex formation with CsrA. This accounts for the fact that components

bound in complexes cannot function in regulation any more. The result was a simplified model involving only three dynamical components (long mRNS M, CsrA A, and effective sRNA S), but which still captures the hierarchical regulation:

$$\partial_t M = \alpha_M - \delta_M M - k_M M \cdot A,$$
$$\partial_t S = \alpha_S - \delta_S S - k_S S \cdot A,$$
$$\partial_t A = \alpha_A - \delta_A A - k_M M \cdot A - k_S S \cdot A.$$

The degradation rates $\delta_{A,M,S}$ and $k_{M,S}$ are already known from previous work or could be measured, such that only the production rates α needed to be fitted to experimental abundance measurements [41]. Note that some parameter values are different in 2.5 and 2.6; this originates from more precise measurements for $k_{M,S}$ in the latter publication. Moreover, we added an A-dependent growth rate for S in 2.6. This step was motivated by studies reporting this regulatory interaction [41] and provided more realistic fitting results.

Threshold behaviour in steady state.   We employed the three-component model to calculate the steady state abundances for a broad range of long mRNA and effective sRNA synthesis rates (with fixed CsrA production). This parameter study resulted in a threshold behaviour which is already known from other systems [42, 43]: The long mRNA production rate needs to exceed a threshold value for free long mRNA to exist in the system. More specifically, below a threshold mRNA production rate, the long mRNA production is so low that almost immediately free CsrA binds to it; above the threshold, the roles reverse, and only free long mRNA exists. The threshold is reduced with increasing sRNA production, as the sRNA is also binding free CsrA. Adjusting the sRNA production thus means adjusting the threshold.

Third component dampens fluctuations.   To investigate the intrinsic fluctuation of the three-component model, we reformulated the rate equations as master equations. We then used a VAN-KAMPEN approximation approach [44, 45] to calculate the first two moments, which allowed us to to obtain the FANO factor (variance divided by mean) [3] of the long mRNA abundance (for the same range of long mRNA and effective sRNA synthesis rates as we used for our steady state results). We find that, close to the threshold long mRNA production rate, fluctuations are the largest, which agrees with previous results on gene expression thresholds [46, 47]. When the sRNA production is increased, this general behaviour remains, but the fluctuation size decreases. This means that the presence of the CsrA regulator sRNA helps reducing internal fluctuations.

Broad lysis time distributions   We created a stochastic simulation for the three component model [48, 49], and coupled it to the stochastic model for the SOS response by SHIMONI ET AL [21]. For our simulation runs, we considered a system that starts in normal state and is then subjected to an SOS state for a well-defined, prolonged time, after which it switches back to normal state. In the normal state, few long mRNA molecules are produced, but noise in the SOS system creates short and random production peaks. These are filtered out by the regulatory network, as it takes time to reduce existing free CsrA buffer. During the SOS signals,

the buffer shrink quickly. Consequently, stochastic bursts of activity create significant peaks of long mRNA abundance, which we consider as large enough for lysis. Due to the stochasticity of the peaks and stochastic abundances in the ColicinE2 system, each simulation run exhibits a different lysis time. We analysed an ensemble of runs, resembling an ensemble of bacteria, and found a broad lysis time distribution that qualitatively agrees with experimental results.

Plasmid composition affects delay time distributions The studies on ssDNA as regulator required the creation of five genetically identical strains, which only differ in their plasmid composition. We recreated the five different strains in our model by adjusting the plasmid abundance parameter, and repeated the above simulation scheme for a SOS response. To be able to replicate the experimental results, we did not compare the lysis times, but the delay between the production of toxin and lysis protein. Increasing plasmid abundance leads to more CsrA sequestering long mRNAs being produced, which results in a decrease of the average delay time. However, it was impossible to bring the experimental observations in accordance with simulation results, in particular for strains containing the original ColicinE2 plasmid. This suggested the existence of an additional, unknown regulator.

ssDNA is another regulator of CsrA Experimental results indicated that ssDNA, which are intermediates in ColicinE2 plasmid replication, could also sequester CsrA. To analyse the role of ssDNA as a CsrA-regulator, we introduced it as a fourth component to our system of rate equations. To this end, we made the same assumption of fast complex equilibration as we did for long mRNA and sRNA. We also included the ssDNA accordingly to our stochastic model, and performed ensemble simulations to obtain distributions for the delays between toxin production and lysis. These delay distributions allowed to be compared with experimental measurements, which agreed both qualitatively and quantitatively. Moreover, the simulations could predict delay distributions for wild-type bacteria, which are experimentally inaccessible.

2.4 Conclusions

The production and release of toxins is a powerful mechanism for bacteria to increase their fitness against competitors. A prominent example for this mechanism is the ColicinE2 system, in which stochastic SOS signals trigger production of toxins that can only be released by cell lysis. Because of this fatal form of toxin secretion, the release is tightly regulated on both the transcriptional and post-transcriptional level in a well-studied regulatory network: Long mRNA, coding for the lysis protein, is regulated by the protein CsrA, which itself is sequestrated by sRNAs. However, the dynamical behaviour of the regulatory components was unclear, especially in context of the stochastic SOS signals and intrinsic noise. The first research question of this project thus was: **How does stochasticity affect ColicinE2 regulation?**

Stochasticity in form of noise is potentially harmful for toxin producers, as it might result in an erroneous or premature release of toxins. This is particularly the case for ColicinE2, as its release entails the death of its producer. One source of such unwanted noise are fluctuations of component abundances in the regulatory network. We studied the stochastic dynamics

of the ColicinE2 system by developing a simplified, yet comprehensive model for the three components long mRNA, CsrA and sRNA. Specifically, the model captures the hierarchical order of the components: CsrA binds long mRNA, and sRNA binds CsrA, rendering sRNA as the "regulator's regulator".

Our steady state analysis of this model showed that, for the expression of lysis proteins, the long mRNA production needs to exceed a threshold, which is adjusted by sRNA production. This threshold mechanism is known from similar regulatory systems [43, 50], and protects the cell from unwanted lysis due to fluctuations or short-lived SOS signals. However, this mechanism is prone to strong fluctuations just above its threshold value [46, 47]. When investigating the sRNA dependence of long mRNA fluctuations, we found that, in addition to the threshold mechanism, **the hierarchical regulation reduces intrinsic fluctuations** in the component abundances. More precisely, our results show that the effectiveness of this novel mechanism increases with the sRNA production rate, and that it disappears in the absence of sRNA (and thus the absence of hierarchical regulation). The fluctuation-dampening effect is particularly strong above the threshold, indicating that it has two functions here: First, to preserve the "sharpness" (that is, the position) of the threshold, and second, to reduce the fluctuations beyond the scope of the threshold mechanism.

These effects do not rely on specific properties of the regulating agents. We therefore expect that hierarchical regulation poses a general mechanism to safeguard critical regulatory networks against internal noise. The component CsrA, for instance, is a global regulator in E. coli cells, and binds to hundreds of mRNA targets. It is therefore reasonable to assume that the effects of hierarchical regulation are also relevant to many other post-transcriptional regulation systems, in particular in combinations with gene expression thresholds. Moreover, we expect that more complex hierarchical networks exist, involving more than two layers of regulation or more than one target. Finding such networks and investigating the fluctuations in them could be the starting point of future research.

To investigate the additional stochastic effects of the SOS response system, we implemented a stochastic simulation of our ColicinE2 system, and connected it to a stochastic model for the LexA-RecA system. Simulating a SOS response using this combined system then showed that the **random SOS signal peaks** from the SOS system are transduced through the regulatory network, and **together with stochastic abundances in the ColicinE2 system create heterogeneous lysis times in a population of bacteria**. Thus, in the context of our overarching question, this poses another mechanism that creates diverse populations. Previous studies already indicated this connection between the SOS response system and heterogeneous expression [51, 52]; however, by simulating an ensemble of bacteria, our model allows a more quantified view on the lysis times: We find that the lysis times in fact are broadly distributed. As the release of toxins is triggered by environmental stress, this broad distribution enables differentiated reactions like bet hedging to external signals. The lysis time distribution was also the study subject of a follow-up project, see Chapter 3.

In the second part of this project, our model for the ColicinE2 system helped to identify ssDNA as novel regulator for CsrA. Specifically, we were interested in the question: **What is the role of ssDNA in ColicinE2 regulation?** As the individual lysis times are hard to measure in experiments, we considered the delay between the production of toxin and lysis protein in this part of the project.

As a first step, we used stochastic simulations of our ColicinE2 model to obtain the delay distributions (and thus, also the average delay times) of five different strains. These strains were genetically identical, but contained different numbers of reporter and ColicinE2 plasmids. In experiments, the five strains differed significantly in their average delays. These differences could not be recreated by our model in the simulations. More specifically, the strains containing ColicinE2 plasmids exhibited particularly short delays, indicating that this plasmid creates a so far unknown regulatory interaction. Our conjecture was confirmed, as experiments identified ssDNA to be a possible regulator of CsrA, which is an intermediate in the replication of ColicinE2 plasmids. We introduced ssDNA as fourth component to our model, and repeated the simulations for the five strains. The relations and values of the resulting average delay times were in good agreement with the experiments, and therefore support the experimental findings that **ssDNA binds and thus regulates CsrA**. Our results show that experimental measurements allow predictions from numerical models, which then in turn can support experimental hypotheses. It thus constitutes a nice example for a fruitful cooperation between theory and experiment. However, while the extended model predicted abundances that were within plausible ranges, it was not able to exactly reproduce all quantitative relations found in previous experimental studies. We attribute this to the fact that the model is limited to a small subsystem of *E. coli* gene regulation; many of its components are integral parts of other regulatory networks. This particularly holds true for CsrA, which is a major regulator for several functions in *E. coli*. Consequently, the experimental results are also affected by interactions that our model does not account for.

To our knowledge, this study is the first time that ssDNA has been reported to be part of a regulatory network. This discovery thus has possible implications for many CsrA regulated systems in cells containing ColicinE2 plasmids. Moreover, we expect the regulatory effects of ssDNA also to be relevant for other post-transcriptionally regulated systems that involve plasmids with the same replication mechanism [53, 54] as ColicinE2; this is because ssDNA contains the same genes and thus the same binding motifs as the mRNA transcripts.

The results from this project suggest that the abundance of plasmids can play a crucial role in regulatory networks. For simplicity, we always kept the number of plasmids in our analysis constant. However, even though the plasmid copy number controlled by a separate system [55], the interplay of a growing plasmid number with the ColicinE2 system can have a significant effect on the lysis time distribution. Investigating these consequences of plasmid dynamics thus poses an interesting challenge for future studies.

2.5 Publication Reprint

Hierarchical Post-transcriptional Regulation of Colicin E2 Expression in **_Escherichia Coli_**

by

M. Lechner,[*][1], M. Schwarz,[*][2], M. Opitz,[1] and E. Frey[1]

[*] Contributed equally to this work
[1] Arnold-Sommerfeld-Center for Theoretical Physics and Center for NanoScience,
Ludwig-Maximilians-Universität, Munich, Germany,
[2] Institute for Biological and Medical Imaging, Technische Universität München and
Helmholtz Zentrum München, Neuherberg, Germany

reprinted from

RESEARCH ARTICLE

Hierarchical Post-transcriptional Regulation of Colicin E2 Expression in *Escherichia coli*

Matthias Lechner[1]◉, Mathias Schwarz[1,2]◉, Madeleine Opitz[1], Erwin Frey[1]*

1 Arnold-Sommerfeld-Center for Theoretical Physics and Center for NanoScience, Ludwig-Maximilians-Universität, Munich, Germany, **2** Institute for Biological and Medical Imaging, Technische Universität München and Helmholtz Zentrum München, Neuherberg, Germany

◉ These authors contributed equally to this work.
* frey@lmu.de

Abstract

Post-transcriptional regulation of gene expression plays a crucial role in many bacterial pathways. In particular, the translation of mRNA can be regulated by trans-acting, small, non-coding RNAs (sRNAs) or mRNA-binding proteins, each of which has been successfully treated theoretically using two-component models. An important system that includes a combination of these modes of post-transcriptional regulation is the Colicin E2 system. DNA damage, by triggering the SOS response, leads to the heterogeneous expression of the Colicin E2 operon including the *cea* gene encoding the toxin colicin E2, and the *cel* gene that codes for the induction of cell lysis and release of colicin. Although previous studies have uncovered the system's basic regulatory interactions, its dynamical behavior is still unknown. Here, we develop a simple, yet comprehensive, mathematical model of the colicin E2 regulatory network, and study its dynamics. Its post-transcriptional regulation can be reduced to three hierarchically ordered components: the mRNA including the *cel* gene, the mRNA-binding protein CsrA, and an effective sRNA that regulates CsrA. We demonstrate that the stationary state of this system exhibits a pronounced threshold in the abundance of free mRNA. As post-transcriptional regulation is known to be noisy, we performed a detailed stochastic analysis, and found fluctuations to be largest at production rates close to the threshold. The magnitude of fluctuations can be tuned by the rate of production of the sRNA. To study the dynamics in response to an SOS signal, we incorporated the LexA-RecA SOS response network into our model. We found that CsrA regulation filtered out short-lived activation peaks and caused a delay in lysis gene expression for prolonged SOS signals, which is also seen in experiments. Moreover, we showed that a stochastic SOS signal creates a broad lysis time distribution. Our model thus theoretically describes Colicin E2 expression dynamics in detail and reveals the importance of the specific regulatory components for the timing of toxin release.

Author Summary

Gene expression is a fundamental biological process, in which living cells use genetic information to synthesize functional products like proteins. To control this process, cells

make use of many different mechanisms. A well-studied example is the binding of expression intermediates by a cellular component in order to delay the synthesis. This mechanism is known to regulate the stress-induced release of the toxin colicin E2 by *E. coli* bacteria. However, experimental studies have shown that this system is not regulated by just one component, but the interplay of several cellular components, in which the hierarchically ordered main components interact. Here, we create a mathematical model for the interaction network of colicin E2 release, and study how the component levels evolve. We show that the system is able to delay the release of the toxin. Additional components allow to fine-tune the delay and dampen fluctuations in gene expression that would lead to premature toxin release. A comprehensive analysis of the time evolution reveals a broad distribution of toxin release times, which is also observed in experiments. This rich dynamical behavior emerges from the interplay of regulatory components, and, due to its generality, may also be transferred to similar regulatory networks, in particular toxin expression systems.

Introduction

Regulation of gene expression occurs at transcriptional and post-transcriptional levels, and has been studied intensively both experimentally and theoretically [1–10]. Bacterial stress responses, such as the well-studied production and release of the toxin colicin E2 in *Escherichia coli*, represent one setting in which post-transcriptional control is crucial [11–15].

Colicins are toxic proteins produced by certain *E. coli* strains in response to stress as a means to kill bacteria that compete with them for the same resources. More specifically, colicin E2 is a bacteriocin, which damages the DNA of bacterial cells that absorb it (a DNAse). Once synthesized, colicin E2 forms a complex with an immunity protein, thus protecting its producer from its otherwise lethal action [14, 16, 17]. The toxin is released only upon cell lysis, which is triggered by the synthesis of a dedicated lysis protein [15, 18–20]. As this inevitably entails the death of the producer cell [19], it is vital for the persistence of the population that only a fraction of its members actually releases the toxin [14]. The genes for the colicin, immunity protein and lysis protein are organized into the colicin E2 operon, which is depicted in Fig 1, together with the interaction network that controls colicin E2 expression and release.

Each of the three components is encoded by a single gene—the colicin by *cea*, the colicin-specific immunity protein by *cei*, and the lysis factor by *cel*— and three regulatory regions control their transcription: an SOS promoter upstream of the *cea* gene [21], and two transcriptional terminators T_1 and T_2, located upstream and downstream of the *cel* gene, respectively [22]. The key transcriptional regulator of the SOS operon is the LexA protein (reviewed in [23]), marked in orange in Fig 1. LexA dimers repress the SOS promoter region of the ColE2 operon, but also block the transcription of over 30 other SOS genes [24, 25], many of which play an important role in DNA repair [26]. In the event of DNA damage, the LexA dimer undergoes auto-cleavage upon interaction with RecA [27], and the transcription of SOS genes begins. The presence of the two transcriptional terminators in the ColE2 operon results in the production of two different mRNAs: A shorter transcript (short mRNA, marked purple in Fig 1) that encompasses only the genes for the toxin colicin E2 and the immunity protein, and a longer transcript (long mRNA, marked green in Fig 1), which additionally includes the lysis gene [14, 28, 29–32]. Hence, lysis can only be initiated after the translation of long mRNA [18], and this crucial operation is regulated at the post-transcriptional level, as described below.
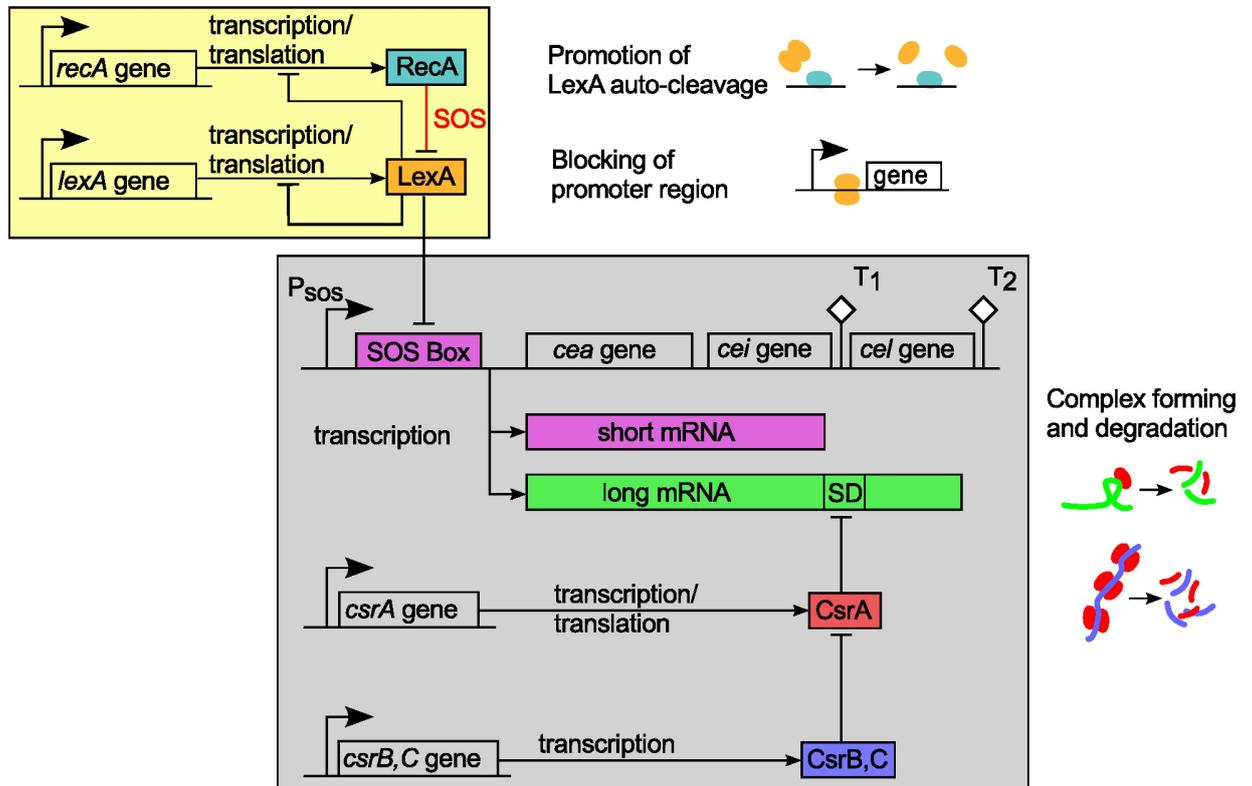
Fig 1. Regulation of colicin E2 expression and release. The interaction scheme is a generalized adaption of that presented by Yang [28]. Under normal conditions, the SOS response system (yellow box) maintains a constant level of LexA dimers, which repress the SOS promoter of the colicin E2 system (gray box). In the event of DNA damage, RecA is activated and promotes auto-cleavage of LexA. This permits the transcription of two different mRNAs: Short mRNA codes for components of colicin immunity complexes (colicin gene *cea*, immunity gene *cei*), whereas long mRNA additionally encodes the protein that triggers cell lysis. Translation of long mRNA is regulated by binding of the protein CsrA to its Shine-Dalgarno sequence (SD). CsrA itself is regulated by the two sRNAs CsrB and CsrC. Other elements: P_{sos}: SOS promoter; T_1 and T_2: transcriptional terminators.

doi:10.1371/journal.pcbi.1005243.g001

Post-transcriptional regulation makes use of many different mechanisms. Recent studies emphasize the particular importance of non-coding sRNAs [33] for various processes in *E. coli*, especially because of their ability to introduce delays and set up thresholds for translation [34–37]. This is done either directly, by sRNAs pairing with their target mRNA (sRNA-mRNA interaction), or indirectly, by sequestering of specific mRNA-binding proteins (mRNA-protein interaction) [2, 38, 39]. For the latter form of regulation, recent studies highlighted the importance of the production rates of regulatory components [40]. In the case of the ColE2 system, the translation of the long mRNA is regulated by the carbon storage regulator protein CsrA [28], marked red in Fig 1. CsrA dimers destabilize target mRNAs by binding to a region that includes the ribosome-binding site (Shine-Dalgarno sequence) [41]. Masking of the ribosome-binding site by CsrA thus not only represses translation of the lysis gene but also promotes degradation of the long mRNA. However, CsrA is also recognized by two specific sRNAs, CsrB and CsrC [42], marked blue in Fig 1. These sRNAs can therefore sequester CsrA dimers, preventing them from binding to target mRNAs [43–45]. Thus, translation of the ColE2 lysis gene is indirectly regulated by sequestration of CsrA. This process, also known as "molecular titration", exhibits ultrasensitive thresholds and has been extensively studied [46, 47].
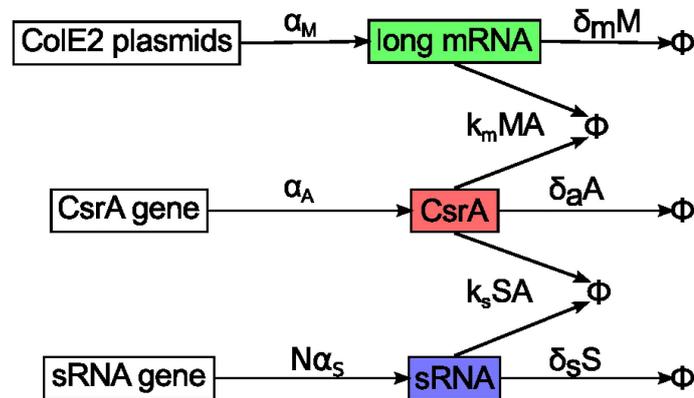
Fig 2. Simplified interaction scheme for post-transcriptional regulation of long mRNA. M,A,S: molecule numbers of free long mRNA, free CsrA dimers and the free effective sRNA; α: production rates; δ: degradation rates; k: effective rate of coupled degradation. The interaction network (see S1 Fig) of the regulatory system depicted in Fig 1 was reduced to a three component system. In both figures, the corresponding components have the same colors. In particular, we combined the complex dynamics (binding, dissociation, degradation) into an effective coupled degradation. The dynamics of sRNA complexes with N binding sites for CsrA and production rate α_S were simplified to the dynamics of an effective sRNA with one CsrA binding site but N-times higher production rate (S1 Text).

doi:10.1371/journal.pcbi.1005243.g002

The basic interaction network that controls the ColE2 regulatory network has been studied in great detail in previous works [48–51], and many of its functional characteristics, in particular the threshold behavior, were described for a wide range of both bacterial and eukaryotic systems [52]. However, a detailed theoretical description of the dynamics leading to the release of colicin is still missing, in particular the role of the hierarchically ordered regulation involving CsrB and CsrC. In this work, we have formulated this post-transcriptional network in a detailed mathematical model, constructed by analogy to studies of simpler sRNA-regulated systems (for example, [33, 34, 36]). We then simplified the model by assuming fast complex equilibration, and combining the sRNAs CsrB and CsrC into a single, effective sRNA (see S1 Text for details). This reduced the regulation network to three relevant components: free long mRNA, free CsrA and the effective sRNA (see Fig 2). We then analyzed this simplified network in detail. In contrast to previous work [36], we give a general analytical solution for the three component system, and derive a precise approximation for fast and clear analysis. This analytic solution exhibits a pronounced threshold in mRNA production due to CsrA-dependent regulation, which was also confirmed using numeric simulations. We investigated, how this threshold depends on system parameters, and how it affects the actual biological system. Furthermore, we have analyzed the role of fluctuations in the post-transcriptional regulation network and how fluctuations in long mRNA expression may be dampened by sRNA. Finally, we extended our model by including the transcriptional regulation, and analyzed how the system behaves during a realistic SOS response. Previous studies have shown discrete activation peaks in LexA-repressed promoters [26] that can lead to large fluctuations close to the threshold of mRNA expression [9]. In a stochastic simulation of the complete model, we were able to reproduce this phenomenon. Comparison with experimental data on lysis time distributions [48] also shows that our model can explain the delayed and broadly distributed release times of colicin complexes. This underlines the importance of stochasticity for the heterogeneous expression of colicin E2 in *E. coli* populations.

Results

A mathematical model for post-transcriptional regulation of colicin E2 release

For our theoretical analysis, we initially developed a detailed mathematical model for the post-transcriptional regulation of colicin E2 release. To this end, we derived a set of coupled, deterministic rate equations from the interaction network depicted in Fig 1, with the corresponding rates for transcription, degradation, binding interactions etc. as parameters. In the following, we briefly review how we reduced the network to its core components, which comprise the theoretical model. The interaction scheme underlying the complete model is presented in S1 Fig and further explanations can be found in the Supporting Information, where we also detail how our model can account for sequestration by other targets of the global regulator CsrA.

As we wished to study the post-transcriptional regulation of colicin E2 expression, we included in the model only those components that are relevant at that stage. The model therefore omits the short mRNA and its products. However, the rate of transcription of the long mRNA is a crucial parameter, which is influenced by the kinetics of activation of the SOS promoter, and thus by the processing of its repressor LexA. Upon DNA damage, RecA promotes auto-cleavage of LexA dimers, thus removing inhibition of the SOS response (marked in red in Fig 1). The LexA-RecA interaction network has recently been modeled stochastically [53]. Before including this detailed network in our final model, we focused on understanding the post-transcriptional dynamics. To this end, we initially assumed that activation of the SOS promoter occurs rapidly relative to the rates of production and degradation of the long mRNA [54], which allowed us to approximate the transcription rate of long mRNA by an effective rate α_M (Materials and Methods). With respect to post-transcriptionally relevant components, we were then left with long mRNA, CsrA, and the two sRNAs CsrB and CsrC, and the mRNA-CsrA-, CsrA-CsrB-, and CsrA-CsrC-complexes.

CsrB and CsrC regulate CsrA by forming complexes with it. The two sRNAs each have several (on average: N) CsrA binding sites, and if every occupation state of the sRNAs were to be modeled as a separate component, a large number of coupled rate equations would need to be added to the model. However, due to the fast dynamics of the CsrA-CsrB- and CsrA-CsrC-complexes, and their virtually identical biochemical behavior, we were able to reduce the sRNA interaction to a single equation for an effective sRNA, with only one binding site and transcription rate $N\alpha_S$ (see Materials and Methods). As a result, the mechanisms of complex formation, dissociation and degradation are replaced by an effective coupled degradation of complex partners. Despite the different processes that are integrated to effective ones, the effective sRNA still resembles the dynamical behavior of CsrB/CsrC. A detailed derivation of the simplified system of rate equations can be found in S1 Text. The final post-transcriptional model is thus reduced to a set of three coupled, deterministic rate equations that capture the behavior of the free long mRNA (M), free CsrA dimers (A), and an effective free sRNA (S) component with a single CsrA binding site:

$$\dot{M} = \alpha_M - \delta_M M - k_M MA, \tag{1}$$

$$\dot{A} = \alpha_A - \delta_A A - k_M p_M MA - k_S p_S AS, \tag{2}$$

$$\dot{S} = N\alpha_S - \delta_S S - A k_S S, \tag{3}$$

where $(1 - p_M)$ and $(1 - p_S)$ are the probabilities for CsrA to survive the coupled degradation. A graphical illustration of this differential equation system is depicted in Fig 2. Note that in the

model the quantities M, A and S represent the abundance of the corresponding *free* components. Once a long mRNA, sRNA, or CsrA dimer binds to some other component, it loses its function and is thus removed from the model system.

For the analysis of our model, we had to determine production, degradation and binding rates. The particular values used are listed in S1 Table. As far as possible, we chose values that are measured in studies on either the same or comparable systems (see S1 Text for details). In the other cases, we tried to derive plausible parameters from known factors that influence the particular rate. A detailed motivation and derivation of these rates is given in chapter 2 of S1 Text.

Post-transcriptional regulation yields a tunable rate threshold in mRNA abundance

We analyzed the reduced post-transcriptional model by first calculating its steady state. In order to obtain a cleaner and simpler result, we derived an approximation (see Materials and Methods) for the steady state solution, which agreed very well with the results of numerical simulations (see S2 Fig). Using these simplified equations, we then investigated the impact of the rates of production of long mRNA (α_M) and sRNA (α_S) on the levels of the three components. The results (see Fig 3) reveal a linear threshold that appears at the same position for all three components. The threshold divides the parameter space into two regimes, in which either CsrA or long mRNA and sRNA have a non-zero abundance. This is due to the coupling between the degradation of CsrA and the abundance of both long mRNA and sRNA, such that the presence of CsrA dimers excludes that of long mRNA and sRNA, and *vice versa*. This mechanism in turn controls the release of colicin-immunity complexes, since a sufficiency of CsrA dimers ensures reliable repression of the long mRNA and prevents synthesis of the lysis protein.

From the aforementioned analytic solution we calculated the threshold position as a function of the system parameters (S1 Text). We found that the threshold for non-zero levels of long mRNA lies exactly at the point where the production rate of CsrA α_A is equal to the sum of transcription rates for long mRNA α_M and sRNA α_S (S1 Text). Thus, we observed no expression of long mRNA in the regime $\alpha_M + \alpha_S < \alpha_A$, as shown in Figs 3A and 4. We find the threshold to be sharp, and attribute this to the very slow degradation of CsrA compared to long mRNA and sRNA [55, 56].

Apart from the threshold itself, we find that the levels of free CsrA and free sRNA predicted by our steady state analysis are consistent with experimental in-vivo values determined by previous studies [43, 57]. Moreover, our results are also consistent with the total amount of CsrA as well as its ratio to sRNA (S1 Text).

So far, we have demonstrated that our three-component system is capable of producing a threshold behavior. However, it has been shown previously that a mutually exclusive production of sRNA and a target mRNA is possible with just two components [36]. The question thus arises why a third component is needed at all. One possible explanation is that the sRNA makes it easier to trigger lysis, as an increase in sRNA production induces an increase in the abundance of long mRNA (Fig 3).

After SOS signals, the sRNA controls and accelerates the degradation of CsrA (see section on expression dynamics below), eventually leading to the expression of the lysis protein.

sRNA controls fluctuations close to thresholds

In a next step, we analyzed the stochastic dynamics of the post-transcriptional regulation network. To this end, we switched to a stochastic description, calculated the Fano factor
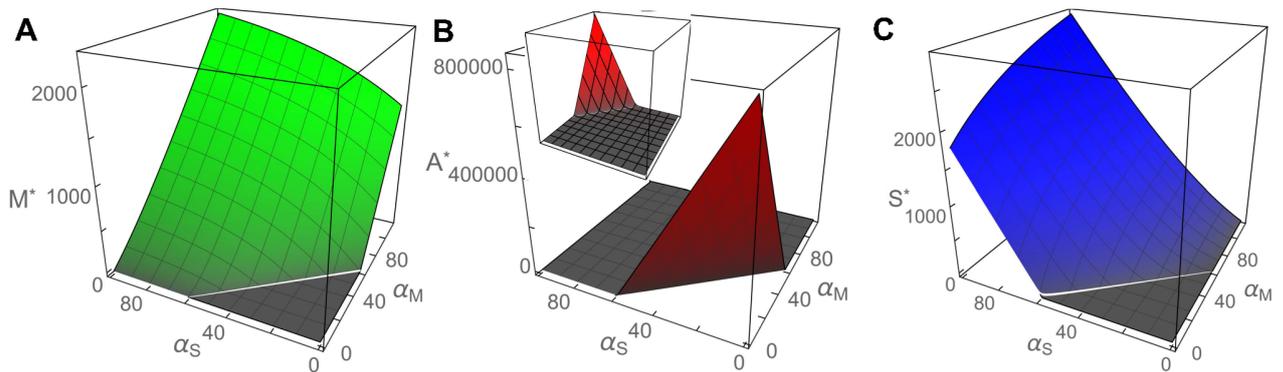
Fig 3. Approximate stationary solutions for (A) long mRNA, (B) CsrA dimers and (C) sRNA. The stationary solutions are given as a function of the effective transcription rate α_M of long mRNA and the production rate α_S of sRNA. The production rate of CsrA dimers was set to $\alpha_A = 58.52$. All other system parameters are given in S1 Table. For values of α_M and α_S below the threshold, the abundances of free long mRNA and sRNA are zero, as any newly produced component quickly forms a complex with the highly abundant CsrA. At sufficiently large production or transcription rates, sRNA and long mRNA titrate all available CsrA molecules and can thus attain non-zero molecule numbers, The white line gives the transition between two approximate analytical solutions (Materials and Methods).

($\mathrm{Var}M/\langle M\rangle$) for the abundance of long mRNA (see Materials and Methods), and depicted it as heatmap in Fig 4. The Fano factor measures the relative magnitude of fluctuations, and has already been applied to gene regulatory networks in previous studies [58]. It can also be understood as a quantified comparison with the pure birth process (Poisson process), which has the Fano factor $F = 1$.

We found that fluctuations in mRNA were most pronounced close to the threshold position, with the largest fluctuations occurring slightly above the threshold (Fig 4). Moreover, Fig 4 also shows that the fluctuations became larger as sRNA production decreases. Thus, the
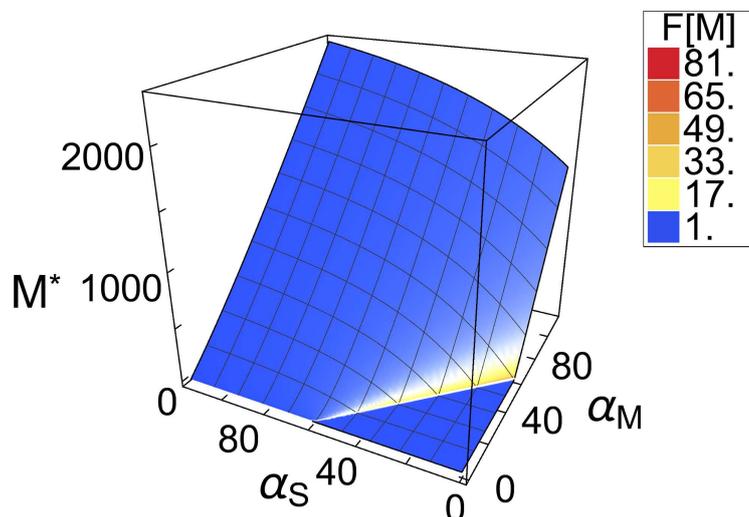
Fig 4. Fluctuations in long mRNA abundance. The fluctuations are quantified by the Fano factor (see main text) and depicted as heatmap in the plot. They are most pronounced at the threshold, and fade for parameter sets above the threshold. With an increase in sRNA production ($N\alpha_S$), the fluctuations become smaller and more localized to the threshold. This illustrates how the third component sRNA acts as a means to reduce intrinsic fluctuations. The production rate of CsrA dimers was again set to $\alpha_A = 58.52$, and all other system parameters are given in S1 Table.

third component (sRNA) in the post-transcriptional regulation network also enables significant dampening of fluctuations in long mRNA.

To understand why the fluctuations are localized to the region near threshold, one must take the characteristics of this parameter regime into account. Around the threshold, molecule numbers are close to zero, which has a direct affect on the relative size of fluctuations: the lower the abundance, the larger the fluctuations (stochastic regime). Moreover, the threshold is the only regime in which all three components, CsrA, mRNA and sRNA, can coexist and interact with each other: An increase in the level of CsrA will lead to a decrease in the abundance of long mRNA and sRNA, owing to increased complex formation and subsequent degradation. Analogously, an increase in long mRNA and sRNA molecule numbers leads to a decrease in CsrA abundance. Therefore, the abundance of CsrA dimers is anti-correlated with the abundance of both long mRNA and sRNA. It has been shown for a two-component system, that anti-correlated components can create anomalously large fluctuations [59] if degradation rates are small compared to turnover (ratio of production rate to abundance). For long mRNA, this is exactly the case close to threshold, where the long mRNA abundance is still very low.

These results show that a third component can reduce intrinsic fluctuations of a hierarchically ordered regulatory network.

Modeling colicin E2 expression dynamics in response to an SOS signal

To study the dynamical response of the ColE2 system to an SOS signal, we extended the post-transcriptional network by including the LexA-RecA regulatory network [53] (Fig 1). LexA not only represses the SOS promoter, it is also an auto-repressor, as well as being a repressor of RecA production. As outlined in the Introduction, RecA forms filaments after DNA damage, which then induce auto-cleavage of LexA dimers. Consequently, the levels of RecA, LexA and the colicin mRNAs increase, as repression due to LexA is relaxed. A stochastic model of this network has been introduced recently [53]. In that study, promoter activity in the LexA-RecA system was found to occur in ordered bursts that result from fluctuations and the particular structure of the RecA-LexA feedback loop.

In our analysis of the ColE2 post-transcriptional regulation network so far (see above), we have assumed the dynamics of SOS promoter activation to be so fast that we could use an effective transcription rate α_M for long mRNA. To link the LexA regulatory network to the post-transcriptional regulation network, we must drop this assumption and explicitly model the dynamics of LexA dimers, which connect the two networks. In the biological system, this involves the binding and dissociation of LexA dimers to and from the SOS promoter in the ColE2 operon. Long mRNA and short mRNA are transcribed only from the derepressed promoter at rates α_{M_l} and α_{M_s}, respectively. Thus, the transcription rates of long mRNA and short mRNA are proportional to the number of open SOS promoters in the bacterium. The majority of transcripts are short mRNAs. The mathematical implementation of the integrated regulation network is again a system of coupled rate equations, which we describe in S1 Text. The additional parameters of the LexA-RecA regulation network are to be found in S2 Table.

We simulated the SOS signal by temporarily up-regulating the coupling parameter c_p, which quantifies the ability of RecA to induce cleavage of LexA (Fig 1). In the uninduced state before and after the SOS signal, the auto-cleavage parameter was set to $c_p = 0$. Under SOS stress c_p was increased to $c_p = 6$. This increase in c_p subsequently boosts the long mRNA production, and therefore relates to a transition from a sub-threshold state (gray area below the white line in Fig 3A) to a super-threshold state (green area above the white line in Fig 3A). Due to the stochasticity in the LexA-RecA network and the resulting stochastic promoter dynamics, the overall transcription rate α_{M_l} of long mRNA is not constant, but fluctuates about a mean value.

The production rate of sRNA was held constant at α_S = 57.5. Fig 5 shows the dynamics of short and long mRNA levels and the abundance of CsrA dimers and sRNA in response to transient SOS signaling. When we compared a stochastic realization using Gillespie simulations (Materials and Methods) with a numerical solution of the deterministic rate-equation system, we observed significant qualitative and quantitative differences. First, the stochastic realization exhibited significant fluctuations that manifested themselves in abrupt, short-lived changes in the abundance of short mRNA over the whole time-course (Fig 5A). Second, the average over 500 stochastic realizations deviated from the deterministically predicted value. Both phenomena arise from the intrinsic stochasticity of the LexA-RecA-regulatory network, as explained by Shimoni [53]. Fluctuations may lead to a spontaneous dip in the number of LexA dimers which releases all LexA-regulated genes, including the *lexA* gene itself, from repression. This consequently leads to a sudden rise in the abundance of short mRNA. The open *lexA* and *recA* promoters will then generate a burst of newly produced LexA and RecA proteins, which block and regulate the promoters for the next burst.

Focusing on the dynamics of mRNA transcription, we found that, due to initial simulation parameters, only small numbers of the short mRNA are produced in the uninduced state. After up-regulation of the LexA auto-cleavage parameter c_p at t = 200 min, the abundance of short mRNA rises and the aforementioned large bursts appear. The amount of long mRNA, however, follows a completely different trajectory, conditioned by post-transcriptional regulation. Before the SOS signal, expression of long mRNA is almost completely repressed by CsrA (Fig 5B). Even the bursts of SOS promoter activity reflected in fluctuating amounts of the short mRNA have little or no impact on the long mRNA. This filtering effect is biologically relevant, as it ensures that noisy promoter activity does not erroneously trigger lysis. After induction of the SOS signal, the deterministic dynamics of the underlying rate equations predicted that, after a delay of about 40 min, the abundance of long mRNA should rapidly rise to a saturation value (black dashed line in Fig 5B). However, a mean of 500 realizations deviated significantly from this prediction (Fig 5B). In particular, the average number of long mRNA molecules increased more slowly than predicted by deterministic dynamics. Hence the abundance saturated at a much lower value. An appreciable delay between SOS signal induction and expression of long mRNA was still observed, but lasted for only 15 min.

Studying the dynamics of a single stochastic realization, we observed that the number of long mRNA molecules underwent large fluctuations, which were followed by periods of no expression at all. Moreover, the timing of these bursts varied considerably between different realizations. This constitutes a significant qualitative difference compared to the average over 500 realizations and to the deterministic dynamics (Fig 5), both of which exhibit a smooth and continuous temporal behavior. Fig 5B and 5C indicates the origin of this behavior: The abundance of long mRNA can only grow if the number of free CsrA dimers is low. The same holds for the abundance of sRNA, which supports the degradation of CsrA and also can only reach non-zero abundances if there is no CsrA left (Fig 5D). Thus, before any long mRNA can be expressed, the free CsrA concentration must drop to very low values due to degradation or complex formation. The delay between SOS signal induction and the first burst of long mRNA synthesis therefore depends on the amount of CsrA available. We went on to study the precise timing of the first burst in long mRNA abundance, since it is crucial for the time-point of release of colicin-immunity complexes. To this end, we calculated the probability distribution for the first peak from an ensemble of 500 stochastic realizations. The probability of a peak in long mRNA abundance rose quickly and reached its maximum approximately 60 min after induction of the SOS signal (Fig 6A). This phenomenon is also seen in experimental systems: time-lapse studies with colicin-producing bacteria revealed that their lysis time is broadly distributed [48]. The distribution depicted in Fig 6A matches qualitatively with comparable
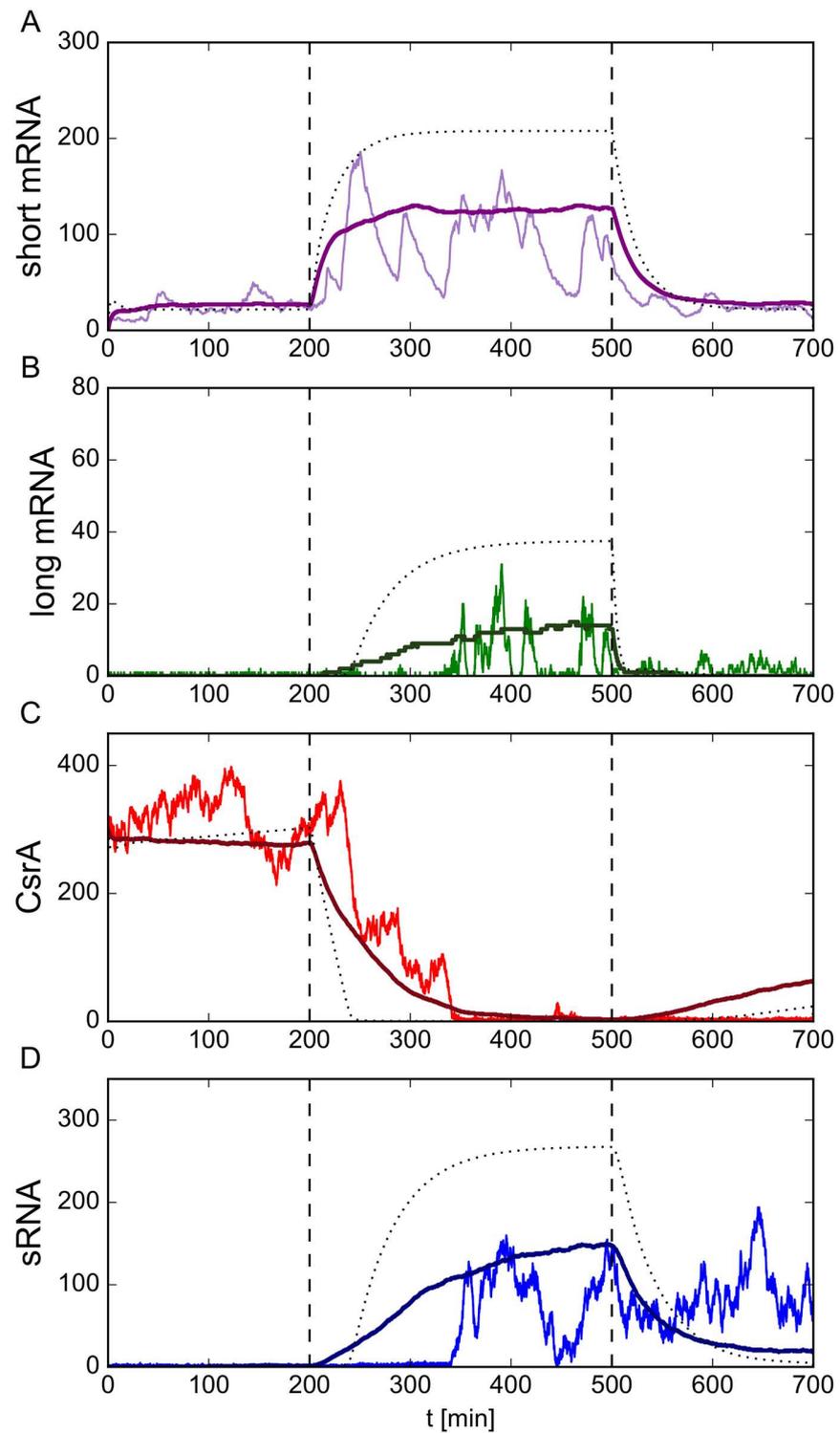
Fig 5. Dynamical behavior before and after a realistic SOS response. We simulated an SOS signal by temporarily up-regulating the LexA auto-cleavage parameter from $c_p = 0.0$ to $c_p = 6.0$ between the two dashed vertical lines at $t = 200$ min and $t = 500$ min. The parameter c_p gives the rate at which LexA dimers degrade due to the presence of RecA. During the simulation, we tracked the abundance of (A) free short mRNA, (B) free long mRNA, (C) free CsrA dimers and (D) free sRNA over time. In each panel, the fluctuating colored

curve represents a single realization of the stochastic system as implemented by a Gillespie simulation. The smoother darker-colored curve shows the average of 500 different realizations. The black dashed curve depicts the results found by numerical integration of the deterministic rate equations, which neglects fluctuations. In general, the stochastic realizations deviated significantly from both the simulation average and the deterministic solution, as they exhibited large spontaneous bursts. As the short mRNA is not post-transcriptionally regulated, its abundance level can serve as a proxy for the SOS promoter activity. Comparing the free short mRNA abundance with free long mRNA shows that short promoter activity peaks were reliably filtered out by post-transcriptional regulation. After an up-regulation of the LexA auto-cleavage parameter c_p at $t = 200$ min, the abundance of short mRNA rose and is expressed in large bursts. After some time delay, during which all newly produced long mRNAs immediately sequestered CsrA dimers, discrete bursts of free long mRNA are seen, which were followed by periods of no production at all. The timing of the bursts varied considerably between different realizations. A comparison with (C) shows that the abundance of free long mRNA is anti-correlated with the molecule number of all free CsrA. Hence, free long mRNA is only present if the number of free CsrA dimers is low. In the simulation, the production rate of CsrA dimers was set to $\alpha_A = 58.52$ and the transcription rate of sRNA to $\alpha_S = 57.5$. All other parameters are given in S1 and S2 Tables.

doi:10.1371/journal.pcbi.1005243.g005
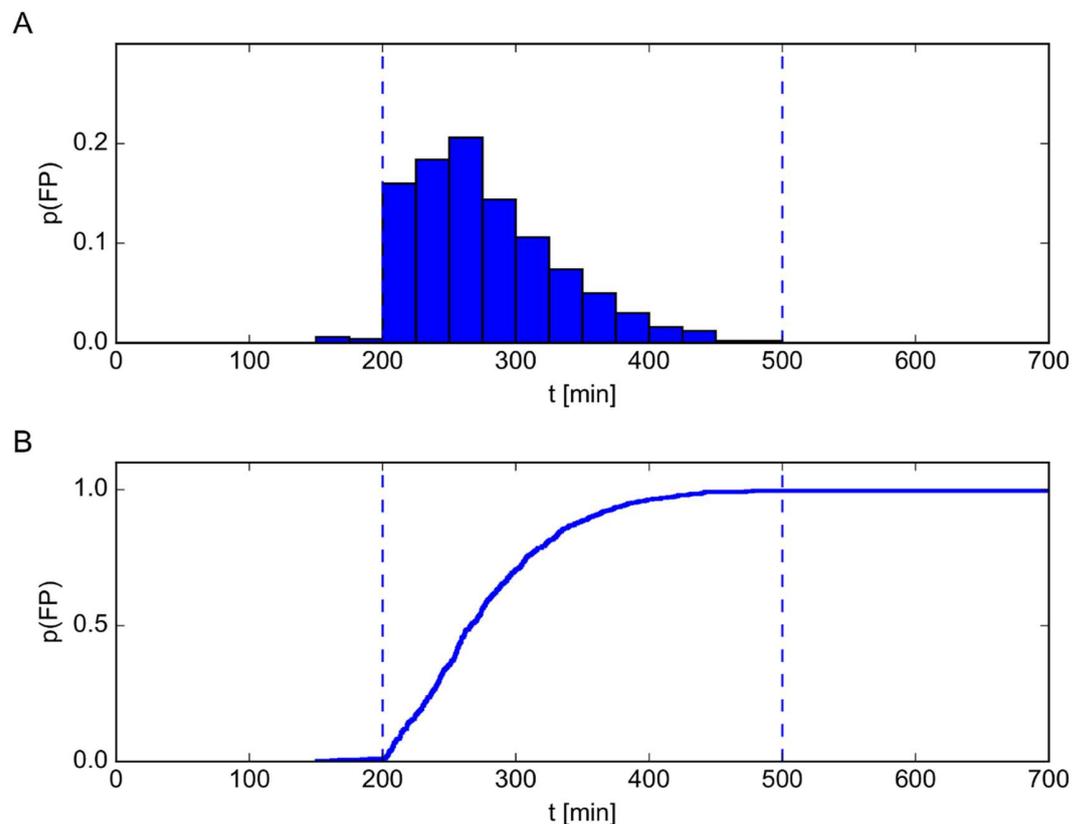
Fig 6. Probability distribution of the first peak in long mRNA abundance and survival function. We simulated an SOS signal by temporarily up-regulating the LexA auto-cleavage parameter from $c_p = 0.0$ to $c_p = 6.0$ between the two dashed vertical lines at $t = 200$ min and $t = 500$ min (see also Fig 5). The parameter c_p gives the rate at which LexA dimers degrade due to the presence of RecA. (A) With the parameters defined in S1 and S2 Tables, the timing of the first peak in long mRNA abundance is broadly distributed with maximal probability approximately 60 min after induction of the SOS signal. (B) The survival function is defined as the fraction of *E. coli* cells in a population that exhibited no peak in long mRNA abundance, and thus would not release colicin. The fraction of cells releasing colicin increased smoothly after induction up to 100%. This heterogeneous response of a bacterial population to an SOS signal is also observed in nature. In the simulation, the production rate of CsrA dimers was set to $\alpha_A = 58.52$ and the transcription rate of sRNA to $\alpha_S = 57.5$. All other parameters are given in S1 and S2 Tables.

doi:10.1371/journal.pcbi.1005243.g006

datasets from these experiments. Moreover, our model is able to numerically predict average lysis times in dependence on different SOS signal strengths (see S5 Fig). From the probability distribution of the timing of the initial peak in long mRNA abundance we calculated the survival function, i.e. the probability with respect to time that a cell will not release toxin. Here we assumed that this first burst provides enough long mRNA in the cell to produce the lysis protein, which then induces its lysis with concomitant release of colicin-immunity complexes into the surrounding medium. The function of lysed cells plotted in Fig 6B shows that the number of cells that release the toxin rises with the duration of the SOS signal.

Incorporation of the LexA-RecA regulatory network allowed us to model the colicin E2 expression dynamics in response to a realistic SOS signal, and the results presented above highlight the importance of CsrA for colicin release.

Discussion

Gene expression is a process that allows for various forms of regulation at all levels. In theoretical studies of post-transcriptional regulation of several biological systems, modulation of mRNA production by proteins or sRNA has been shown to create, for instance, temporal thresholds for mRNA translation [9, 35, 36]. Proteins have also been shown to regulate the expression of the toxin colicin E2 [28] in the context of an SOS response to environmental stress. Experimental studies have elucidated the detailed interaction network responsible for the production and release of the colicin [28]. However, the dynamics of this system, in particular at the post-transcriptional level, have remained elusive. In close analogy to previous two-component models, we developed a mathematical model for this hierarchically ordered post-transcriptional regulation of colicin E2 release. Interestingly, the known interaction network for this system necessitated the modeling of three, not two, components: the long mRNA which is necessary for colicin release, its negative regulator CsrA, and sRNA, which in turn negatively regulates CsrA. Contrary to previous studies [9, 35, 36, 60], the sRNAs do not regulate the mRNA directly, but control the level of the regulator protein CsrA. Thus, the sRNA acts as the "regulator's regulator".

In our analysis of the model, we used rate constants that were determined from experimental systems (see chapter 2 of S1 Text for details). Comparing the predicted CsrA levels before the SOS signal (see Fig 5C) with in-vivo measurements of *E. coli* [57] shows that our model results in a pre-SOS free CsrA abundance that agrees with actual bacterial systems (for other abundances, see S1 Text). Moreover, the model is not just able to predict steady state abundances, but also reproduces the reaction to varying external stress levels as seen in experiments (see S5 Fig).

Investigation of the dynamics revealed that the model exhibits a time delay in the production of free long mRNAs. This delay is due to the high abundance of CsrA in the non-SOS state of the cell, which causes CsrA to quickly bind to free long mRNA and thus prevents its transcription. Only during an SOS signal, which indicates external stress for the cell, the level of CsrA gets steadily reduced. The time this process takes to get CsrA levels so low that fluctuations in long mRNA production result in free long mRNA, causes a delay in colicin release. As colicin release is coupled to cell lysis, the delay is therefore a mechanism for filtering out transient SOS signals that might erroneously lead to synthesis of the lysis protein. Moreover, also intrinsic fluctuations, for instance in sRNA production, are filtered out by this mechanism: Even if a large and sudden burst in sRNA were strong enough to drop CsrA abundance close to zero, the CsrA buffer gets restored quickly due to the large production rate of CsrA. This rate is only effectively lowered during a SOS signal, which increases the production of the CsrA-sequestering long mRNA. The fact that lysis is regulated by a

threshold mechanism of a global regulator protein like CsrA might also be a guarding mechanism for the cell: only prolonged extreme situations will cause the abundance of these regulators to drop to low molecule numbers.

However, delays and similar threshold behavior also emerge in two-component systems, raising the question why a third component is necessary here. Strikingly, we found that the third component (sRNA) in the post-transcriptional interaction network enables the cell to tune the duration of the delay by sequestering CsrA. In the case of the ColE2 system, this means that cells are able to adjust the (average) time between a SOS signal and the onset of cell lysis leading to colicin release.

Furthermore, previous studies of systems with slow, bursting promoter kinetics have also uncovered a major limitation of two-component sRNA-based regulation compared to regulation based on transcription factors: Two-component systems are subject to significantly higher levels of intrinsic noise [9]. However, Fig 4 (panels A,C,D) shows that, in the post-transcriptional regulation of colicin E2 release, fluctuations become smaller at higher values of α_S. The sRNA might therefore allow for significant dampening of these fluctuations. This idea is supported by the fact that the relatively high degradation rate of sRNA makes it less susceptible to induced fluctuations.

In bacteria, these mechanisms could have several functions: First, a comparison of different sRNA production rates (S4 Fig) indicates that the sequestration of CsrA by the sRNA could indeed be crucial for fast release of the colicin, as CsrA degradation rates cannot be arbitrarily increased in bacterial systems. Second, they can tune the reaction to external stress at the population level. Experimental studies have shown that, in the absence of stress, 3% of colicin producing cells release the toxin during the stationary phase; but this fraction can be increased up to eventually 100% if an external SOS stress is applied [14, 48]. Previous experimental studies also found that colicin systems exhibit heterogenous expression times, which originate from the stochasticity of the SOS signal [49, 50]. Recent time-lapse experiments with colicin E2 producing bacteria showed that this lysis time distribution also depends on the strength of the SOS signal [48]. We reproduced these experiments with stochastic simulations, in which we created different stress levels by different values of the RecA degradation rate parameter c_p. Our predictions for lysis time distributions (Fig 6A and S5 Fig) show qualitative agreement with these time-lapse experiments. Moreover, the ability of the sRNA to tune the average duration of the delay might serve as a mechanism to adjust the cell lysis to different stress levels. Altering the sRNA level could be an additional mechanism, apart from the stochastic SOS signal, by which bacterial populations can adjust the fraction of cells releasing the toxin depending on the strength and duration of the external stress. Finally, the co-option of sRNA makes the cells less susceptible to lysis due to adventitious fluctuations in promoter activity. This is particularly important considering the bursting behavior and large-scale fluctuations seen in the LexA-RecA-regulatory system, which are readily observed in experiments and reproduced by stochastic models [53].

In order to focus on the interplay between the LexA-RecA system and the hierarchical regulation of long mRNA by CsrA and sRNA, we kept the plasmid number constant. If we considered random, Poisson-distributed plasmid numbers instead, the effect would be very small, as shown in S4B Fig. This fact demonstrates that the colicin plasmid copy number only has minor influence on the lysis time distribution (see S1 Text for details).

In conclusion, we have provided here the first detailed theoretical description of colicin E2 production and release, and used it to study the dynamical behavior of this system. Moreover, the general three-component model described here should be applicable to many other systems of toxin production in microorganisms.

Materials and Methods

Derivation of effective long mRNA transcription rate α_M

In most models of prokaryotic gene expression, it is assumed that promoter kinetics are fast compared to RNA production and degradation rates. In that case, the promoter state is well approximated by its steady state [54]. In the analysis of the post-transcriptional regulation network, the promoter status affects the transcription rate of the (long) mRNA. Thus, we replaced it by an effective transcription rate for (long) mRNA, which takes into account the probability of a gene being blocked. In the literature this procedure is referred to as "adiabatic elimination of fast variables" (see for example [61]). For this effective rate we also took into account that the colicin operon is located on a plasmid [62], of which approximately 20 copies exist in each cell [14] (see S1 Text).

Reduction of CsrB and CsrC to an effective sRNA

The two sRNAs CsrB and CsrC regulate CsrA via complex formation. More specifically, each CsrB molecule has approximately 22 binding sites for CsrA, with 9 CsrA dimers being attached on average [63, 64]. CsrC interacts in the same way, but has fewer CsrA binding sites [63]. As a first step, we therefore replaced the two sRNA types by a single effective one, which has N binding sites. However, all of the $N + 1$ sRNA configurations still enter the interaction network as separate components, since the binding and dissociation probabilities change with the number of free binding sites. By investigating the dynamics of the CsrA-sRNA complexes, we discovered that the probability distribution for occupied CsrA binding sites on the sRNAs reaches its stationary state on a time scale that is proportional to the rate of complex-(un)binding. Since binding and unbinding events are biochemically much simpler processes than transcription, translation or degradation, it is very likely that the dynamics of CsrA-sRNA complexes is much faster than all other reaction rates in the system. Following the line of Levine [36] and Legewie [34], we therefore assumed rapid complex dynamics, and replaced the different binding site occupations by an effective sRNA, with only one binding site and transcription rate $N\alpha_S$ (see S1 Text for details on the calculation).

Approximate solution of the reduced three-component model

For the calculations of the abundances of the three components (for example, to obtain the plots of Fig 3), we began by assuming the stationary state. Solving for the abundance of one component then gives a cubic equation, for which the exact, general solution is very lengthy and cumbersome to analyze. Therefore, we considered the cubic equation for the cases of very large and very small molecule numbers, and ignored terms that became negligible. This resulted in two easily solvable quadratic equations. Comparisons with numerical solutions of the cubic equations proved that the quadratic solutions approximate the general solution well in their respective abundance regime. Equating the terms omitted in the approximation yields a criterion for the transition between the two approximations (see S1 Text). The transition is depicted as a white line in Fig 3. That this transition lies close to the threshold is coincidental. Comparison with the exact, numerical solutions showed that the threshold is not an approximation artifact. S2 Fig illustrates the precision of the approximation by comparing its prediction for long mRNA abundance to that from numerical simulations.

Calculation of the Fano factor using linear noise approximation

We started the analysis of noise properties by reformulating the simplified three-component system as a Master equation. As Master equations are typically impossible to solve

analytically, we performed a general van Kampen expansion in multiple variables (components). Our analysis included all higher orders, and not only lowest order terms as is commonly found in textbooks [61, 65]. With van Kampen's expansion we were able to derive general formulas for the first up to the fourth moment of the random variable representing the fluctuations of the system around the stationary solution of the rate equations. The terms of each equation were classified in first order terms (dominant terms) and higher order terms (second order, third order, etc), according to the scaling behavior of each term with the system size. We used different methods to calculate the Fano factor for long mRNA. The most reliable results were obtained by implementing only first order terms in the calculations of second moments. This reproduced the shape of the Fano factor well, but it overestimates fluctuations in the vicinity of the threshold. S3 Fig illustrates the degree of agreement between analytical calculations of the Fano factor agree with the results from Gillespie simulations.

Gillespie simulations

To verify how well our analytical results of the deterministic rate equations coincide with the actual mean molecule numbers, we set up a Gillespie simulation [66]. The Gillespie algorithm generates a statistically correct realization of the master equation behind the rate equations. The core of the algorithm lies in using random numbers to determine which next reaction will occur and the waiting time prior to the succeeding reaction. The reactions simulated by the Gillespie approach are listed in S1 Text. To quantify the delay between SOS signal induction and the first burst in long mRNA abundance, we defined the beginning of the first peak as the point when the number of long mRNA molecules exceeds 8 for the first time. The time of the peak itself was set to the point at which that number reached a maximum. We then calculated the probability distribution from an ensemble of 500 stochastic realizations, using the parameters defined in S1 and S2 Tables.

Supporting Information

S1 Table. Parameter values for the post-transcriptional dynamics modeled by the rate equations and Gillespie simulations. Rates are given in molecules per cell volume V_{EC} = $0.65 \mu m^3$ per minute. The number of ColE2 plasmids is n_{sos} = 20. The literature values can be found in [28, 55, 56].
(PDF)

S2 Table. Additional parameter values for the SOS response network, modeled by the rate equations and Gillespie simulations. Rates are given in molecules per cell volume V_{EC} = $0.65 \mu m^3$ per minute. The number of ColE2 plasmids is n_{sos} = 20. R, Le, Col, L: number of RecA proteins, LexA dimers, colicin proteins and lysis proteins. M_l, M_r, M_s, M: number of *lexA, recA*, short mRNAs and long mRNAs. $B_l B_r B_{sos}$: number of LexA dimers bound to the *lexA, recA* and SOS promoters. All literature values are taken from [53].
(PDF)

S1 Fig. Detailed interaction scheme of post-transcriptional regulation network. The interaction scheme is mathematically formulated as $N + 5$ coupled rate equations. M,A,S,L and C_{ma} give the numbers of long mRNA, CsrA dimers, sRNA, lysis protein and long mRNA-CsrA complexes. C_n gives the number of sRNA molecules with n CsrA dimers bound. The rates of a reaction is expressed by the formula next to the arrows. α: production rates; δ: degradation rates; v^-, k^-: Complex dissociation rates; v^+, k^+: Complex formation rates. To illustrate the complex dynamics between CsrA dimers and sRNA we depict the

reaction rates of CsrA with an sRNA that has already bound $n \in [0, 1, \ldots, N]$ CsrA dimers. For more details see S1 Text.
(EPS)

S2 Fig. Comparison of the stationary solution for long mRNA abundance M^* with the time-average $\langle M \rangle$. The stationary solution M^* was calculated using rate equations, the time-average $\langle M \rangle$ was obtained by Gillespie simulations. We show two cuts through the surface of Fig 3A at $\alpha_S = 20$ and $\alpha_S = 40$. The points indicate the result of Gillespie simulations, whereas the lines show the analytical result obtained from the approximated steady state equations. The production rate of CsrA dimers was chosen to be $\alpha_A = 58.52$, all other parameters are given in Table S1 Table.
(EPS)

S3 Fig. Comparison of the analytically calculated Fano factor with corresponding Gillespie simulations. The production rate of CsrA dimers was set to $\alpha_A = 58.52$. All other parameters are given in S1 Table. For both parameter sets, $\alpha_S = 20$ and $\alpha_S = 40$, the analytic calculations using van Kampen's system size expansion reproduced the shape of the fluctuations obtained by Gillespie simulations well. In the threshold regime the analytic result overestimated the fluctuations slightly.
(EPS)

S4 Fig. Effects of parameters on the lysis time distribution. (A) shows the lysis time distribution as in Fig 6A for comparison. (B) This distribution hardly changes if the number of plasmids, n_{SOS}, follows a Poisson distribution. (C) Lowering the sRNA production rate to $\alpha_S = 56$ shifts the lysis distribution towards later times, whereas (D) doubling it to $\alpha_S = 58$ causes several cells to lyse even before (and hence independent of) the SOS signal. This illustrates that the sRNA is a possible means of controlling cell lysis.
(EPS)

S5 Fig. Average lysis times for different stress levels. To illustrate the predictive possibilities of our three component model, we compare the results of numerical simulations using our model with experimental data [48]. The experiment measured the average lysis time for three different concentrations of the antibiotic Mitomycin C (0.05, 0.25 and 0.70 μg/ml). In the numerical simulations, we used the parameter set defined in S1 and S2 Tables, and varied the parameter c_p (values: 1, 3, 6, 12, 15, 20, 30, 90) to emulate the stress levels. To fit the data, we only applied a scaling factor to map the Mitomycin concentration to values of c_p, and shifted the theoretical delays by a constant value. The last step is necessary, as the numerical simulations also account for the constant time between SOS signal and first appearance of short mRNA, which is not the case in the experiments.
(EPS)

S1 Text. Supplementary information on calculations and numerical simulations. Detailed derivations of the (simplified) rate equations and the linear noise approximation, as well as the detailed reaction scheme used in the Gillespie simulations.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: ML MS MO EF.

Data curation: ML MS MO EF.

Formal analysis: ML MS EF.

Funding acquisition: MO EF.

Investigation: ML MS MO.

Methodology: ML MS MO EF.

Project administration: ML MS MO EF.

Resources: MO EF.

Software: ML MS.

Supervision: EF.

Validation: ML MS MO EF.

Visualization: ML MS.

Writing – original draft: ML MS MO EF.

Writing – review & editing: ML MO EF.

References

1. Tsimring L, Volfson D, Hasty J. Stochastically driven genetic circuits. Chaos: An Interdisciplinary Journal of Nonlinear Science. 2006; 16(2):026103–026103. doi: 10.1063/1.2209571

2. Storz G, Vogel J, Wassarman KM. Regulation by small RNAs in bacteria: expanding frontiers. Molecular Cell. 2011; 43(6):880–891. doi: 10.1016/j.molcel.2011.08.022 PMID: 21925377

3. Mitarai N, Andersson AM, Krishna S, Semsey S, Sneppen K. Efficient degradation and expression prioritization with small RNAs. Phys Biol. 2007; 4(3):164–171. doi: 10.1088/1478-3975/4/3/003 PMID: 17928655

4. Kepler TB, Elston TC. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. Biophysical Journal. 2001; 81(6):3116–3136. doi: 10.1016/S0006-3495(01)75949-8 PMID: 11720979

5. Alon U. Network motifs: theory and experimental approaches. Nat Rev Genet. 2007; 8(6):450–461. doi: 10.1038/nrg2102 PMID: 17510665

6. Shen Orr SS, Milo R, Mangan S, Alon U. Network motifs in the transcriptional regulation network of Escherichia coli. Nat Genet. 2002; 31(1):64–68. doi: 10.1038/ng881 PMID: 11967538

7. Milo R, Shen Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. Network motifs: simple building blocks of complex networks. Science. 2002; 298(5594):824–827. doi: 10.1126/science.298.5594.824 PMID: 12399590

8. Desnoyers G, Bouchard MP, Masse E. New insights into small RNA-dependent translational regulation in prokaryotes. Trends in Genetics. 2012;. doi: 10.1016/j.tig.2012.10.004 PMID: 23141721

9. Mehta P, Goyal S, Wingreen NS. A quantitative comparison of sRNA-based and protein-based gene regulation. Molecular Systems Biology. 2008; 4(1). doi: 10.1038/msb.2008.58 PMID: 18854820

10. Mitarai N, Benjamin JA, Krishna S, Semsey S, Csiszovszki Z, Masse E, et al. Dynamic features of gene expression control by small regulatory RNAs. Proc Natl Acad Sci U S A. 2009; 106(26):10655–10659. doi: 10.1073/pnas.0901466106 PMID: 19541626

11. Kerr B, Riley Ma, Feldman MW, Bohannan BJM. Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. Nature. 2002; 418(6894):171–174. doi: 10.1038/nature00823 PMID: 12110887

12. Reichenbach T, Mobilia M, Frey E. Mobility promotes and jeopardizes biodiversity in rock–paper–scissors games. Nature. 2007 Aug; 448(7157):1046–1049. doi: 10.1038/nature06095 PMID: 17728757

13. Weber MF, Poxleitner G, Hebisch E, Frey E, Opitz M. Chemical warfare and survival strategies in bacterial range expansions. J R Soc Interface. 2014; 11(96):20140172. doi: 10.1098/rsif.2014.0172 PMID: 24806706

14. Cascales E, Buchanan SK, Duché D, Kleanthous C, Lloubes R, Postle K, et al. Colicin biology. Microbiology and Molecular Biology Reviews. 2007; 71(1):158–229. doi: 10.1128/MMBR.00036-06 PMID: 17347522

15. Pugsley AP, Schwartz M. A genetic approach to the study of mitomycin-induced lysis of Escherichia coli K-12 strains which produce colicin E2. Molecular Andgeneralgenetics MGG. 1983; 190(3):366–372. doi: 10.1007/BF00331060 PMID: 6308394

16. Wojdyla JA, Fleishman SJ, Baker D, Kleanthous C. Structure of the ultra-high-affinity colicin E2 DNase-Im2 complex. J Mol Biol. 2012; 417(1–2):79–94. doi: 10.1016/j.jmb.2012.01.019 PMID: 22306467

17. Jacob F, Siminovitch L, Wollman E, et al. Biosynthesis of a colicin and its mode of action. In: Annales de l'Institut Pasteur. vol. 83; 1952. p. 295. PMID: 13017127

18. Pugsley AP, Goldzahl N, Barker RM. Colicin E2 Production and Release by Escherichia coli K12 and Other Enterobacteriaceae. Journal of General Microbiology. 1985; 131(10):2673–2686. doi: 10.1099/00221287-131-10-2673 PMID: 3934329

19. Pugsley A, Schwartz M. Expression of a gene in a 400-base-pair fragment of colicin plasmid ColE2-P9 is sufficient to cause host cell lysis. Journal of Bacteriology. 1983; 156(1):109–114. PMID: 6352670

20. Pugsley AP, Schwartz M. Colicin E2 release: lysis, leakage or secretion? Possible role of a phospholipase. The EMBO Journal. 1984; 3(10):2393. PMID: 6389120

21. Dougan G, Saul M, Warren G, Sherratt D. A functional map of plasmid ColE1. Molecular and General Genetics MGG. 1978; 158(3):325–327. doi: 10.1007/BF00267204

22. Cole ST, Saint Joanis B, Pugsley AP. Molecular characterisation of the colicin E2 operon and identification of its products. Molecular and General Genetics MGG. 1985; 198(3):465–472. doi: 10.1007/BF00332940 PMID: 3892228

23. Little JW, Mount DW, et al. The SOS regulatory system of Escherichia coli. Cell. 1982; 29(1):11. doi: 10.1016/0092-8674(82)90085-X PMID: 7049397

24. Fernández de Henestrosa AR, Ogi T, Aoyagi S, Chafin D, Hayes JJ, Ohmori H, et al. Identification of additional genes belonging to the LexA regulon in Escherichia coli. Molecular Microbiology. 2002; 35(6):1560–1572. doi: 10.1046/j.1365-2958.2000.01826.x PMID: 10760155

25. Courcelle J, Khodursky A, Peter B, Brown PO, Hanawalt PC. Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient Escherichia coli. Genetics. 2001; 158(1):41–64. PMID: 11333217

26. Friedman N, Vardi S, Ronen M, Alon U, Stavans J. Precise temporal modulation in the response of the SOS DNA repair network in individual bacteria. PLoS Biology. 2005; 3(7):e238. doi: 10.1371/journal.pbio.0030238 PMID: 15954802

27. Sassanfar M, Roberts JW. Nature of the SOS-inducing signal in Escherichia coli: The involvement of DNA replication. Journal of Molecular Biology. 1990; 212(1):79–96. doi: 10.1016/0022-2836(90)90306-7 PMID: 2108251

28. Yang TY, Sung YM, Lei GS, Romeo T, Chak KF. Posttranscriptional repression of the cel gene of the ColE7 operon by the RNA-binding protein CsrA of Escherichia coli. Nucleic Acids Research. 2010; 38(12):3936–3951. doi: 10.1093/nar/gkq177 PMID: 20378712

29. Van Den Elzen P, Konings R, Veltkamp E, Nijkamp H. Transcription of bacteriocinogenic plasmid CloDF13 in vivo and in vitro: structure of the cloacin immunity operon. Journal of Bacteriology. 1980; 144(2):579–591. PMID: 6159346

30. Chak KF, James R. Analysis of the promoters for the two immunity genes present in the ColE3-CA38 plasmid using two new promoter probe vectors. Nucleic Acids Research. 1985; 13(7):2519–2531. doi: 10.1093/nar/13.7.2519 PMID: 2987857

31. Lloubes R, Baly D, Lazdunski C. The promoters of the genes for colicin production, release and immunity in the ColA plasmid: effects of convergent transcription and Lex A protein. Nucleic Acids Research. 1986; 14(6):2621–2636. doi: 10.1093/nar/14.6.2621 PMID: 2421251

32. Lloubès R, Baty D, Lazdunski C. Transcriptional terminators in the caa-cal operon and cai gene. Nucleic Acids Research. 1988; 16(9):3739–3749. PMID: 2453841

33. Shimoni Y, Friedlander G, Hetzroni G, Niv G, Altuvia S, Biham O, et al. Regulation of gene expression by small non-coding RNAs: a quantitative view. Mol Syst Biol. 2007; 3:138. doi: 10.1038/msb4100181 PMID: 17893699

34. Legewie S, Dienst D, Wilde A, Herzel H, Axmann IM. Small RNAs establish delays and temporal thresholds in gene expression. Biophysical Journal. 2008; 95(7):3232–3238. doi: 10.1529/biophysj.108.133819 PMID: 18599624

35. Levine E, Hwa T. Small RNAs establish gene expression thresholds. Current Opinion in Microbiology. 2008; 11(6):574–579. doi: 10.1016/j.mib.2008.09.016 PMID: 18935980

36. Levine E, Zhang Z, Kuhlman T, Hwa T. Quantitative characteristics of gene regulation by small RNA. PLoS Biology. 2007; 5(9):e229. doi: 10.1371/journal.pbio.0050229 PMID: 17713988

37. Schmiedel JM, Axmann IM, Legewie S. Multi-target regulation by small RNAs synchronizes gene expression thresholds and may enhance ultrasensitive behavior. PLoS One. 2012; 7(8):e42296. doi: 10.1371/journal.pone.0042296 PMID: 22927924

38. Waters LS, Storz G. Regulatory RNAs in bacteria. Cell. 2009; 136(4):615–628. doi: 10.1016/j.cell.2009.01.043 PMID: 19239884

39. Reimmann C, Valverde C, Kay E, Haas D. Posttranscriptional repression of GacS/GacA-controlled genes by the RNA-binding protein RsmE acting together with RsmA in the biocontrol strain Pseudomonas fluorescens CHA0. J Bacteriol. 2005; 187(1):276–285. doi: 10.1128/JB.187.1.276-285.2005 PMID: 15601712

40. Lavi-Itzkovitz A, Peterman N, Jost D, Levine E. Quantitative effect of target translation on small RNA efficacy reveals a novel mode of interaction. Nucleic Acids Res. 2014; 42(19):12200–12211. doi: 10.1093/nar/gku889 PMID: 25294829

41. Liu MY, Yang H, Romeo T. The product of the pleiotropic Escherichia coli gene csrA modulates glycogen biosynthesis via effects on mRNA stability. Journal of Bacteriology. 1995; 177(10):2663–2672. doi: 10.1128/jb.177.10.2663-2672.1995 PMID: 7751274

42. Gutiérrez P, Li Y, Osborne MJ, Pomerantseva E, Liu Q, Gehring K. Solution structure of the carbon storage regulator protein CsrA from Escherichia coli. Journal of Bacteriology. 2005; 187(10):3496–3501. doi: 10.1128/JB.187.10.3496-3501.2005 PMID: 15866937

43. Gudapaty S, Suzuki K, Wang X, Babitzke P, Romeo T. Regulatory Interactions of Csr Components: the RNA Binding Protein CsrA Activates csrB Transcription in Escherichia coli. Journal of Bacteriology. 2001; 183(20):6017–6027. doi: 10.1128/JB.183.20.6017-6027.2001 PMID: 11567002

44. Liu MY, Gui G, Wei B, Preston JF III, Oakford L, Yüksel Ü, et al. The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in Escherichia coli. Journal of Biological Chemistry. 1997; 272(28):17502–17510. doi: 10.1074/jbc.272.28.17502 PMID: 9211896

45. Weilbacher T, Suzuki K, Dubey AK, Wang X, Gudapaty S, Morozov I, et al. A novel sRNA component of the carbon storage regulatory system of Escherichia coli. Molecular Microbiology. 2003; 48(3):657–670. doi: 10.1046/j.1365-2958.2003.03459.x PMID: 12694612

46. Buchler NE, Louis M. Molecular Titration and Ultrasensitivity in Regulatory Networks. J Mol Biol. 2008; 384(5):1106–1119. doi: 10.1016/j.jmb.2008.09.079 PMID: 18938177

47. Buchler NE, Cross FR. Protein sequestration generates a flexible ultrasensitive response in a genetic network. Mol Syst Biol. 2009; 5(272):1–7. doi: 10.1038/msb.2009.30 PMID: 19455136

48. Mader A, von Bronk B, Ewald B, Kesel S, Schnetz K, Frey E, et al. Amount of Colicin Release in Escherichia coli Is Regulated by Lysis Gene Expression of the Colicin E2 Operon. PLoS One. 2015; 10(3):e0119124. doi: 10.1371/journal.pone.0119124 PMID: 25751274

49. Kamenšek S, Podlesek Z, Gillor O, Žgur Bertok D. Genes regulated by the Escherichia coli SOS repressor LexA exhibit heterogenous expression. BMC Microbiol. 2010; 10(1):283. doi: 10.1186/1471-2180-10-283 PMID: 21070632

50. Mrak P, Podlesek Z, van Putten JPM, Žgur Bertok D. Heterogeneity in expression of the Escherichia coli colicin K activity gene cka is controlled by the SOS system and stochastic factors. Mol Genet Genomics. 2007 Apr; 277(4):391–401. doi: 10.1007/s00438-006-0185-x PMID: 17216493

51. Babitzke P, Romeo T. CsrB sRNA family: sequestration of RNA-binding regulatory proteins. Curr Opin Microbiol. 2007; 10(2):156–163. doi: 10.1016/j.mib.2007.03.007 PMID: 17383221

52. Mukherji S, Ebert MS, Zheng GXY, Tsang JS, Sharp PA, van Oudenaarden A. MicroRNAs can generate thresholds in target gene expression. Nat Genet. 2011; 43 VN—r(9):854–859. doi: 10.1038/ng.905 PMID: 21857679

53. Shimoni Y, Altuvia S, Margalit H, Biham O. Stochastic analysis of the SOS response in Escherichia coli. PloS One. 2009; 4(5):e5363. doi: 10.1371/journal.pone.0005363 PMID: 19424504

54. Kærn M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: from theories to phenotypes. Nature Reviews Genetics. 2005; 6(6):451–464. doi: 10.1038/nrg1615 PMID: 15883588

55. Suzuki K, Babitzke P, Kushner SR, Romeo T. Identification of a novel regulatory protein (CsrD) that targets the global regulatory RNAs CsrB and CsrC for degradation by RNase E. Genes Dev. 2006; 20(18):2605–2617. doi: 10.1101/gad.1461606 PMID: 16980588

56. Nath K, Koch AL. Protein Degradation in Escherichia coli. J Biol Chem. 1970; 245(22):2889–2900. PMID: 4912536

57. Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, et al. Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. Science (80-). 2010; 329. doi: 10.1126/science.1188308 PMID: 20671182

58. Thattai M, Van Oudenaarden A. Intrinsic noise in gene regulatory networks. Proceedings of the National Academy of Sciences. 2001; 98(15):8614–8619. doi: 10.1073/pnas.151588598 PMID: 11438714

59. Elf J, Ehrenberg M. Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. Genome Research. 2003; 13(11):2475–2484. doi: 10.1101/gr.1196503 PMID: 14597656

60. Feng L, Rutherford ST, Papenfort K, Bagert JD, Van Kessel JC, Tirrell DA, et al. A Qrr noncoding RNA deploys four different regulatory mechanisms to optimize quorum-sensing dynamics. Cell. 2015; 160 (1–2):228–240. doi: 10.1016/j.cell.2014.11.051 PMID: 25579683

61. Gardiner CW. Stochastic methods. Springer; 2009.

62. Watson RJ, Vernet T, Visentin LP. Relationships of the Col plasmids E2, E3, E4, E5, E6, and E7: restriction mapping and colicin gene fusions. Plasmid. 1985; 13(3):205–210. doi: 10.1016/0147-619X(85)90044-7 PMID: 2987999

63. Clark DP, Pazdernik NJ. Molecular biology: Understanding the genetic revolution. ACACL; 2012.

64. Romeo T, et al. Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB. Molecular Microbiology. 1998; 29(6):1321. doi: 10.1046/j.1365-2958.1998.01021.x PMID: 9781871

65. Van Kampen N. Stochastic Processes in Physics and Chemistry. North-Holland; 2007. doi: 10.1063/1.2915501

66. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry. 1977; 81(25):2340–2361. doi: 10.1063/1.2710253 PMID: 17411109

Supporting Information for: Modelling of Colicin E2 Expression

Mathias Schwarz[1,2], Matthias Lechner[1], Madeleine Opitz[1], Erwin Frey[1]*,

1 Arnold-Sommerfeld-Center for Theoretical Physics and Center for NanoScience, Physics Department, Ludwig-Maximilians-Universität, München, Germany.
2 Institute for Biological and Medical Imaging, Technische Universität München and Helmholtz Zentrum München, Neuherberg, Germany

* frey@lmu.de

Contents

Chapter 1

Rate Equations

1.1 Initial mathematical model

The precise interactions in the regulation network of Colicin E2 release were presented in the results part of the main article. Here we show how we derived the simplified rate equations from the detailed regulation network. The following assumptions underly this process:

1. CsrC interacts with CsrA in the same way as CsrB [1]. There are only minor quantitative differences: At 37°C, the half-life of small RNAs CsrB and CsrC are 1.6 min and 4.1 min, respectively [2]. Furthermore, CsrB has more binding sites than CsrC, and it is unknown if the complex-binding kinetics are different for the two sRNAs. However, we assume these differences to be so small that we can describe the qualitative regulation mechanism of the two sRNAs by one effective sRNA. The biological parameters of the effective sRNA are then adapted to the biological parameters of CsrB and CsrC.

2. For the analysis of the post-transcriptional regulation of Colicin E2 release, we will neglect the regulation of transcription and translation concerning long mRNA, CsrA and sRNAs. In most models of prokaryotic gene expression it is assumed that promoter kinetics are fast compared to production and degradation rates, such that the promoter state is well approximated by its steady state [3]. Thus, an effective transcription rate can be introduced that takes into account the probability of a promoter being blocked. The effective rate is smaller than the original rate. In literature this procedure is referred to as adiabatic elimination of fast variables, see for example [4].

3. The system is considered homogeneous, i.e. reaction rates depend only on the total amount of molecule numbers and not on the local concentration of specific molecules.

4. The exact mechanism of CsrA complex degradation is not known. To keep our model as general as possible we will allow for CsrA dimers to survive degradation of the complexes with probability $(1 - p_M)$ in the case of mRNA complexes and with probability $(1 - p_S)$ for sRNA complexes. We choose CsrA to possibly survive complex degradation, since proteins usually have a much longer lifetime.

4

Notation

- L, M, A: Number of lysis proteins, free long mRNAs and CsrA dimers.

- C_{MA}: Number of long mRNA-CsrA complexes.

- C_n: Number of CsrA-sRNA complexes with n CsrA dimers bound.

- S: Total number of sRNAs (sum over all C_n).

- $\alpha_M, \alpha_A, \alpha_S, \alpha_L$: Effective production of the component denoted by the subscript.

- $\delta_M, \delta_A, \delta_S, \delta_L, \delta_{C_{ma}}$: Degradation rates of the component denoted by the subscript.

- k_M^+, k_M^-, k^+, k^-: Binding rates (+) and unbinding rates (−) of CsrA with mRNAs and sRNAs.

In general, one sRNA has at most N binding sites for CsrA dimers. The total number of sRNAs $S(t)$ is given by the sum over all numbers of complexes $C_n(t)$ with n CsrA dimers bound:

$$S(t) = \sum_{n=0}^{N} C_n(t) \tag{1.1}$$

Rate equations

From the interaction scheme described in the main text we deduce the following rate equations:

$$\dot{L} = \alpha_L M - \delta_L L \tag{1.2}$$

$$\dot{M} = \alpha_M - \delta_M M - k_M{}^+ MA + k_M{}^- C_{MA} \tag{1.3}$$

$$\dot{A} = \alpha_A - \delta_A A - k_M{}^+ MA + k_M{}^- C_{MA} + \delta_{C_{MA}} C_{MA}(1 - p_M)$$
$$- Ak^+ \sum_{n=0}^{N} C_n(N-n) + k^- \sum_{n=0}^{N} C_n n + \sum_{n=0}^{N} \delta_S C_n n(1 - p_S) \tag{1.4}$$

$$\dot{C_{MA}} = k_M{}^+ MA - k_M{}^- C_{MA} - \delta_{C_{ma}} C_{MA} \tag{1.5}$$

$$\dot{C_n} = \alpha_S \delta_{n,0} + C_{n-1} Ak^+(N-(n-1)) + C_{n+1} k^-(n+1) \tag{1.6}$$
$$- C_n \left[Ak^+(N-n) + k^- n + \delta_S \right]$$

$$C_{-1} = C_{N+1} = 0 \tag{1.7}$$

With the definition of the total number of sRNA molecules in equation (1.1) we find:

$$\dot{S}(t) = \alpha_S - \delta_S S \tag{1.8}$$

1.2 Analysis of sRNA complex dynamics

The rate equations (1.2)-(1.7) give a precise mathematical description. Yet, the coupling of $N+1$ differential equations for sRNA complexes to the dynamics of CsrA makes it hard to analyze the system. In this section, we will calculate the first and second moment of the distribution of occupied binding sites. We will find out that the

time scale at which the stationary distribution is approached is fast compared to production and degradation processes. Consequently, we can simplify the $N + 1$ rate equations for the CsrA-sRNA complexes to one effective differential equation. A very helpful tool for this task will be the definition of a generating function.

1.2.1 Generating function

A probability distribution can be characterized by its moments (if they are finite). The moments of the probability distribution of occupied CsrA binding sites $p(n, t) = C_n(t)/S(t)$ are defined as:

$$\langle n^i \rangle = \sum_{n=0}^{N} \frac{C_n}{S} n^i \qquad \text{with } i = 1, 2, \dots \tag{1.9}$$

A powerful tool to investigate the moments of a probability distribution is to define a probability generating function. In our case, we chose:

$$G(x, t) = \sum_{n=0}^{N} \frac{C_n}{S} x^n \tag{1.10}$$

The useful property of a generating function is that it encodes the information of all $p(n)$ in one variable x. Consequently, the $N + 1$ coupled rate equations for sRNA dynamics are simplified to one differential equation of $G(x, t)$ in the variable x. Once we have found the solution of the generating function, we can calculate the mean number and the variance of occupied binding sites via:

$$\langle n(t) \rangle = \partial_x G(x, t)|_{x=1} \tag{1.11}$$

$$\langle n^2(t) \rangle = \partial_x (x \partial_x G(x, t))|_{x=1} \tag{1.12}$$

Furthermore, we can calculate the probability that n CsrA binding sites are occupied:

$$p(n) = \frac{C_n(t)}{S(t)} = \frac{1}{n!} \partial_x^n G(x, t)|_{x=0} \tag{1.13}$$

Our next goal is to set up a differential equation for $G(x, t)$ and to solve this equation. Afterwards, we will infer on the result to obtain information on the probability

distribution $p(n)$. The time evolution for $G(x,t)$ reads as follows:

$$\frac{d}{dt}G(x,t) = \sum_{n=0}^{N} \frac{d}{dt}\left(\frac{C_n}{S}\right)x^n = \sum_{n=0}^{N}\left(\frac{\dot{C}_n}{S} - \frac{C_n(\alpha_S - \delta_S S)}{S^2}\right)x^n$$

$$= \frac{1}{S}\sum_{n=0}^{N}\Bigg(\alpha_S \delta_{n,0} + C_{n-1}Ak^+(N-(n-1)) + C_{n+1}k^-(n+1)$$

$$- C_n\left[Ak^+(N-n) + k^- n + \delta_S\right] - C_n\frac{\alpha_S}{S} + C_n\delta_S\Bigg)x^n$$

$$\overset{(1.7)}{=} \frac{1}{S}\sum_{n=0}^{N}\Bigg(\alpha_S \delta_{n,0} + C_n Ak^+(N-n)x + C_n k^- n\frac{1}{x}$$

$$- C_n\left[Ak^+(N-n) + k^- n\right] - C_n\frac{\alpha_S}{S}\Bigg)x^n$$

$$= \sum_{n=0}^{N}\left(\frac{\alpha_S}{C_n}\delta_{n,0} + Ak^+(N-n)(x-1) + k^- n\left(\frac{1}{x}-1\right) - \frac{\alpha_S}{S}\right)\frac{C_n}{S}x^n$$

$$= \sum_{n=0}^{N}\left(\frac{\alpha_S}{C_n}\delta_{n,0} + Ak^+(x-1)(N-x\partial_x) + k^-(1-x)\partial_x - \frac{\alpha_S}{S}\right)\frac{C_n}{S}x^n$$

$$\frac{d}{dt}G(x,t) = \left(Ak^+(x-1)(N-x\partial_x) + k^-(1-x)\partial_x - \frac{\alpha_S}{S(t)}\right)G(x,t) + \frac{\alpha_S}{S(t)} \quad (1.14)$$

The differential equation (1.14) may be solved using the methods of characteristics. To this end we rewrite equation (1.14):

$$\underbrace{\left(-\frac{\alpha_S}{S(t)} + Ak^+(x(t)-1)N\right)G(x(t),t) + \frac{\alpha_S}{S(t)}}_{\frac{d}{dt}G(x(t),t)} =$$

$$\underbrace{(Ak^+ x(t) + k^-)(x(t)-1)}_{\frac{dx(t)}{dt}}\partial_x G(x(t),t) + \partial_t G(x(t),t)$$

We end up with two differential equations of the form:

$$\frac{dx(t)}{dt} = (Ak^+ x(t) + k^-)(x(t)-1) \quad (1.15)$$

$$\frac{d}{dt}G(x(t),t) = \left(-\frac{\alpha_S}{S(t)} + Ak^+(x(t)-1)N\right)G(x(t),t) + \frac{\alpha_S}{S(t)} \quad (1.16)$$

1.2.2 Solving the differential equation of the generating function

Without loss of generality we set $t_0 = 0$.

Differential equation in $x(t)$

The differential equation (1.15) can readily be solved by separation of variables:

$$\int_0^t dt = t = \int_{x_0}^x \frac{dx'}{(x'-1)(k^+Ax'+k^-)}$$

$$= \frac{1}{k^- + Ak^+} \log\left(\frac{|x-1|}{k^- + Ak^+x} \frac{k^- + Ak^+x_0}{|x_0-1|}\right)$$

$$x(t) = \frac{k^- + e^{t(k^-+Ak^+)}(x_0-1)k^- + Ax_0k^+}{k^- + A\left(x_0 - e^{t(k^-+Ak^+)}(x_0-1)\right)k^+} \tag{1.17}$$

$$x_0 = \frac{(x-1)k^- + e^{t(k^-+Ak^+)}(k^- + Axk^+)}{-A(x-1)k^+ + e^{t(k^-+Ak^+)}(k^- + Axk^+)} \tag{1.18}$$

Note here that both choices for $|x-1| = x-1$ or $1-x$ and $|x_0-1| = x_0-1$ or $1-x_0$ yield the same result if we use the same sign convention for x and x_0.

Differential equation in $G(x,t)$

Solving equation (1.16) is more tedious, since it contains the time-dependent inhomogeneity $\alpha_S/S(t)$. The general solution of such an inhomogeneous differential equation is given by the sum of the general solution $G_h(x,t)$ of the homogeneous differential equation, which neglects the inhomogeneity $\alpha_S/S(t)$, and a particular solution $G_p(x,t)$ of the inhomogeneous differential equation, which takes into account the inhomogeneity. Thus, we have:

$$G(x,t) = G_h(x,t) + G_p(x,t) \tag{1.19}$$

Homogeneous solution

We start by solving the homogeneous differential equation:

$$\frac{d}{dt}G_h(x(t),t) = \left(-\frac{\alpha_S}{S(t)} + Ak^+(x(t)-1)N\right)G_h(x(t),t) \tag{1.20}$$

We can set as initial condition $G_0 \overset{!}{=} x_0^{n_0}$, meaning that at $t_0 = 0$ there are only complexes with n_0 CsrAs bound to it. Since there is one degree of freedom in the choice of $G_p(x(t),t)$ (we may add $G_h(x,t)$ multiplied by an arbitrary constant), we can choose $G_p(x_0,0) = 0 \to G_0 = G_h(x_0,0) = G(x_0,0) \overset{!}{=} x_0^{n_0}$. It follows:

$$\int_{G_0}^G \frac{dG_h'}{G_h'} = \int_0^t dt'\left(-\frac{\alpha_S}{S(t')} + Ak^+(x(t')-1)N\right)$$

$$G_h(x,t) = e^{-u(t)}(k^- + Ak^+)^N x_0^{n_0}\left(\frac{e^{(k^-+Ak^+)t}(k^- + Ak^+)^2}{Ak^+(1-x) + (k^- + Ak^+x)e^{(k^-+Ak^+)t}}\right)^{-N}$$

$$= e^{-u(t)}\left(k^- + Ak^+\right)^N\left(\frac{(x-1)k^- + e^{t(k^-+Ak^+)}(k^- + Axk^+)}{A(1-x)k^+ + e^{t(k^-+Ak^+)}(k^- + Axk^+)}\right)^{n_0} \cdot$$

$$\cdot\left(\frac{e^{(k^-+Ak^+)t}(k^- + Ak^+)^2}{Ak^+(1-x) + (k^- + Ak^+x)e^{(k^-+Ak^+)t}}\right)^{-N} \tag{1.21}$$

$$u(t) = \int_0^t dt' \frac{\alpha_S}{S(t')} \tag{1.22}$$

For the integration over time we have introduced $x(t)$ explicitly. Following integration we have replaced every x_0 by the right-hand side of equation (1.18). The term $u(t)$ is the integral over the inhomogeneity that we leave untouched for the moment. When we look at the long-time limit, $u(t)$ will simplify significantly.

To simplify equation (1.21), we set $n_0 = 0$. This choice will not limit our analysis, because we are only interested in the stationary binding site distribution and in the time scale at which the stationary binding site distribution is approached. Both objects of interest are independent of n_0. Hence, the general homogeneous solution is given by:

$$G_h(x,t) = e^{-u(t)} \left(\frac{e^{t\left(k^- + Ak^+\right)}\left(k^- + Ak^+\right)}{-A(-1+x)k^+ + e^{t(k^- + Ak^+)}\left(k^- + Axk^+\right)} \right)^{-N} \tag{1.23}$$

$$G_h(x_0,t) = e^{-u(t)} \left(\frac{k^- + Ak^+}{k^- + A\left(-e^{t(k^- + Ak^+)}(x_0 - 1) + x_0\right)k^+} \right)^{N} \tag{1.24}$$

Particular solution

As commonly done, we choose the ansatz $G_p(x(t),t) = G_{pp}(t)G_h(t)$ for the particular solution of equation (1.16). This ansatz leads to a differential equation for the time-dependent parameter $G_{pp}(t)$:

$$\frac{d}{dt}G_{pp}(x(t),t) = \frac{\alpha_S}{S(t)G_h(x(t),t)} \tag{1.25}$$

Thus, we find:

$$G_{pp}(x(t),t) = \int_0^t dt' \underbrace{\frac{\alpha_S}{S(t')} e^{u(t')}}_{\frac{d}{dt'}e^{u(t')}} \left(\frac{e^{t\left(k^- + Ak^+\right)}\left(k^- + Ak^+\right)}{A(x(t)-1)k^+ + e^{t(k^- + Ak^+)}\left(k^- + Ax(t)k^+\right)} \right)^{N}$$

$$G_{pp}(x_0,t) = \int_0^t dt' \left[\frac{d}{dt'}\left(e^{u(t')}\right) \right] \left(\frac{k^- + A\left(-e^{t\left(k^- + Ak^+\right)}(x_0 - 1) + x_0\right)k^+}{k^- + Ak^+} \right)^{N}$$

$$\overset{\text{PI}}{=} \left[e^{u(t')} \left(\frac{k^- + Ak^+ x_0 + Ak^+(1 - x_0)e^{(k^- + Ak^+)t'}}{k^- + Ak^+} \right)^{N} \right]_0^t$$

$$- \int_0^t dt' e^{u(t')} \frac{d}{dt'} \left(\frac{k^- + Ak^+ x_0 + Ak^+(1 - x_0)e^{(k^- + Ak^+)t'}}{k^- + Ak^+} \right)^{N}$$

$$= \sum_{n=0}^{N} \binom{N}{n} \left(\frac{k^- + Ak^+ x_0}{k^- + Ak^+} \right)^{N-n} \left(\frac{Ak^+(1 - x_0)}{k^- + Ak^+} \right)^{n}$$

$$\left(\left[e^{u(t')}e^{(k^- + Ak^+)nt'} \right]_0^t - \int_0^t dt' e^{u(t')}(k^- + Ak^+)ne^{(k^- + Ak^+)nt'} \right) \tag{1.26}$$

In order to proceed with the calculations we have to specify $u(t)$. Since $\lim\limits_{t \to \infty} S(t) = S_\infty = \frac{\alpha_S}{\delta_S}$, we may calculate $u(t)$ as follows:

$$u(t) = \int_0^t dt' \frac{\alpha_S}{S(t)} = \int_0^t dt' \left(\frac{\alpha_S}{S_\infty} + \delta_u(t) \right) = \delta_S t + \int_0^t dt' \delta_u(t) \tag{1.27}$$

Since $S(t)$ converges exponentially fast towards S_∞, we know that $\delta_u(t)$ has to go exponentially fast to zero as well and we end up with a finite integral $\int_0^\infty dt' \delta_u(t) = \Delta_u$. Thus, for times larger than the time scale given by the degradation of sRNA, we may approximate:

$$u(t) \approx \delta_S t + \Delta_u \tag{1.28}$$

It follows that for large times larger than $1/\delta_S$:

$$
\begin{aligned}
G_{pp}(x_0, t) &= \sum_{n=0}^N \binom{N}{n} \left(\frac{k^- + Ak^+ x_0}{k^- + Ak^+} \right)^{N-n} \left(\frac{Ak^+(1 - x_0)}{k^- + Ak^+} \right)^n \\
&\quad \left(\left[e^{u(t)} e^{(k^- + Ak^+)nt} - 1 \right] - \left[\frac{(k^- + Ak^+)n}{(k^- + Ak^+)n + \delta_S} e^{\delta_S t' + \Delta_u} e^{(k^- + Ak^+)nt'} \right]_0^t \right) \\
&= \sum_{n=0}^N \binom{N}{n} \left(\frac{k^- + Ak^+ x_0}{k^- + Ak^+} \right)^{N-n} \left(\frac{Ak^+(1 - x_0)}{k^- + Ak^+} \right)^n \\
&\quad \left(1 - \frac{(k^- + Ak^+)n}{(k^- + Ak^+)n + \delta_S} \right) \left(e^{u(t)} e^{(k^- + Ak^+)nt} - 1 \right)
\end{aligned}
$$

$$\tag{1.29}$$

$$
\begin{aligned}
G_{pp}(x, t) &= \sum_{n=0}^N \binom{N}{n} \left(\frac{(k^- + Ak^+ x) e^{(k^- + Ak^+)t}}{(k^- + Ak^+ x) e^{(k^- + Ak^+)t} - Ak^+(x - 1)} \right)^{N-n} \\
&\quad \left(\frac{Ak^+(1 - x)}{(k^- + Ak^+ x) e^{(k^- + Ak^+)t} - Ak^+(x - 1)} \right)^n \\
&\quad \left(\frac{\delta_S}{(k^- + Ak^+)n + \delta_S} \right) \left(e^{u(t)} e^{(k^- + Ak^+)nt} - 1 \right)
\end{aligned}
\tag{1.30}
$$

General solution for $n_0 = 0$

Taking everything together we have:

$$G(x, t) = G_h(x, t)(1 + G_{pp}(x, t)) \tag{1.31}$$

$$G_h(x, t) = e^{-u(t)} \left(\frac{(k^- + Ak^+ x) e^{(k^- + Ak^+)t} - Ak^+(x - 1)}{e^{(k^- + Ak^+)t}(k^- + Ak^+)} \right)^N \tag{1.32}$$

$$
\begin{aligned}
G_{pp}(x, t) &= \sum_{n=0}^N \binom{N}{n} \left(\frac{(k^- + Ak^+ x) e^{(k^- + Ak^+)t}}{(k^- + Ak^+ x) e^{(k^- + Ak^+)t} - Ak^+(x - 1)} \right)^{N-n} \\
&\quad \left(\frac{Ak^+(1 - x)}{(k^- + Ak^+ x) e^{(k^- + Ak^+)t} - Ak^+(x - 1)} \right)^n \\
&\quad \left(\frac{\delta_S}{(k^- + Ak^+)n + \delta_S} \right) \left(e^{u(t)} e^{(k^- + Ak^+)nt} - 1 \right)
\end{aligned}
\tag{1.33}
$$

1.2.3 Calculating mean and variance

With the general solution of the generating function at hand, we can evaluate the mean and the variance of n occupied binding sites:

$$\langle n(t)\rangle = \partial_x G(x,t)|_{x=1} = [(1+G_{pp}(x,t))\partial_x G_h(x,t) + G_h(x,t)\partial_x G_{pp}(x,t)]_{x=1}$$

(1.34)

$$\langle (n(t))^2\rangle = \partial_x(x\partial_x G(x,t))|_{x=1} = \partial_x G(x,t)|_{x=1} + x\left[(1+G_{pp}(x,t))\partial_x^2 G_h(x,t)+\right.$$
$$\left. +2\partial_x G_h(x,t)\partial_x G_{pp}(x,t) + G_h(x,t)\partial_x^2 G_{pp}(x,t)\right]_{x=1}$$

(1.35)

Inserting equations (1.31)-(1.33) into equation (1.34) and equation (1.35) yields:

$$\langle n(t)\rangle = \frac{ANk^+}{\delta_s + k^- + Ak^+}\left(1 - e^{-(k^- + Ak^+)t - u(t)}\right)$$

(1.36)

$$\langle n_\infty\rangle = \frac{ANk^+}{\delta_s + k^- + Ak^+}$$
$$= \frac{ANk^+}{k^- + Ak^+}\left[1 - \frac{\delta_s}{k^- + Ak^+} + \mathcal{O}\left(\left(\frac{\delta_s}{k^- + Ak^+}\right)^2\right)\right]$$

(1.37)

$$\langle n_\infty^2\rangle = \frac{ANk^+\left(\delta_s + 2\left(k^- + ANk^+\right)\right)}{\left(\delta_S + k^- + Ak^+\right)\left(\delta_s + 2\left(k^- + Ak^+\right)\right)}$$
$$= \frac{ANk^+\left(k^- + ANk^+\right)}{\left(k^- + Ak^+\right)^2} - \frac{\left(ANk^+\left(2k^- + A(-1+3N)k^+\right)\right)\delta_s}{2\left(k^- + Ak^+\right)^3}$$
$$+ \mathcal{O}\left(\left(\frac{\delta_s}{(k^- + Ak^+)^2}\right)^2\right)$$

(1.38)

1.2.4 Discussion

From equation (1.36) we obtain an important result. Again taking $u(t) \approx \delta_S t + \Delta_u$, we find that the mean number of occupied binding sites relaxes with the rate $\delta_S + k^- + Ak^+$ to its stationary value, which is faster than the relaxation of sRNA abundance, happening with rate δ_S. The degradation of sRNA happens on a time scale of several minutes, whereas binding and unbinding of molecules occurs within several seconds. Thus, $\delta_S \ll k^- + Ak^+$ most likely holds, which has important consequences: *The relaxation of occupied binding sites towards its equilibrium value is so fast that it can be considered in quasi-equilibrium compared to production and degradation processes. Consequently, the dynamics of sRNA complexes may be approximated by its equilibrium distribution.*

If $\delta_S \ll k^- + Ak^+$, the 0th order term in $\delta_s/(k^- + Ak^+)$ in equation (1.37) dominates and we end up with the results for a simple random walk on N sites with hopping probability $p = \frac{Ak^+}{k^- + Ak^+}$ to the right and hopping probability $q = \frac{k^-}{k^- + Ak^+}$ to the left. For such a process, we have: $\langle n\rangle = Np$ and $\text{Var}[n] = \langle n^2\rangle - \langle n\rangle^2 = Npq$, which is reproduced by the 0th order term in equation (1.38).

The mean number of occupied binding sites decreases with increasing δ_S. This makes sense, since an sRNA is always produced with no CsrA dimer bound, i.e. the source of sRNAs pulls the mean towards lower values.
The faster sRNAs degrade, the less molecules are able to bind multiple CsrA molecules before they degrade.

1.3 Simplified mathematical model

With the results from above we can significantly simplify the rate equations (1.2)-(1.7). First, we note that the rate equation for lysis proteins (1.2) is a linear differential equation that depends only on the number of long mRNA molecules, the translation rate α_L and the degradation rate δ_L. Thus, once we understand the dynamics of M, we comprehend the dynamics of L as well. That is why we ignore equation (1.2), leaving us with the following set of equations:

$$\dot{M} = \alpha_M - \delta_M M - k_M{}^+ MA + k_M{}^- C_{MA} \tag{1.39}$$

$$\dot{A} = \alpha_A - \delta_A A - k_M{}^+ MA + k_M{}^- C_{MA} + \delta_{C_{MA}} C_{MA}(1 - p_M)$$
$$- Ak^+ \sum_{n=0}^{N} C_n(N - n) + k^- \sum_{n=0}^{N} C_n n + \sum_{n=0}^{N} \delta_S C_n n(1 - p_S) \tag{1.40}$$

$$\dot{C}_{MA} = k_M{}^+ MA - k_M{}^- C_{MA} - \delta_{C_{ma}} C_{MA} \tag{1.41}$$

$$\dot{C}_n = \alpha_S \delta_{n,0} + C_{n-1} Ak^+ (N - (n-1)) + C_{n+1} k^- (n+1)$$
$$- C_n \left[Ak^+(N - n) + k^- n + \delta_S \right] \tag{1.42}$$

$$\dot{S} = \alpha_S - \delta_S S \tag{1.43}$$

In section 1.2.4 we found that the probability distribution of occupied CsrA binding sites on sRNA is approached on the time scale $\delta_S + k^- + Ak^+$. As in the work of Levine [5] and Legewie [6] we assume now rapid equilibrium of complex dynamics and approximate the quasi-equilibrated CsrA binding sites distribution by a single, effective complex configuration that has exactly $\langle n_\infty \rangle$ molecules bound:

$$C_0, C_1, \ldots, C_N \to C_{\langle n_\infty \rangle} \qquad \text{with } \langle n_\infty \rangle = \frac{ANk^+}{\delta_S + k^- + Ak^+} \tag{1.44}$$

As all sRNAs are assumed to have this distribution, it holds

$$C_{\langle n_\infty \rangle} \equiv S \tag{1.45}$$

The number of CsrA-mRNA complexes relaxes as well on a time scale proportional to the complex binding and unbinding rates $Ak_M{}^+$ and $k_M{}^-$. Consequently, we set the left-hand side of equation (1.41) equal to zero:

$$\dot{C}_{MA} = 0: \quad C_{MA} = \frac{k_M{}^+ MA}{k_M{}^- + \delta_{C_{MA}}} = \frac{k_M MA}{\delta_{C_{MA}}} \tag{1.46}$$

For a clear notation, we defined the lumped complex dynamic parameters:

$$k_M = \frac{k_M{}^+ \delta_{C_{MA}}}{k_M{}^- + \delta_{C_{MA}}} \tag{1.47}$$

$$k = \frac{k^+ \delta_S}{k^- + \delta_S} \tag{1.48}$$

These lumped parameters can be understood as the effectiveness of coupled degradation, for it is the ratio of binding rate times degradation rate divided by the unbinding rate of the complex. k will be used later on. Taking everything together, we find:

$$\dot{M} = \alpha_M - \delta_M M - k_M MA \tag{1.49}$$

$$\dot{A} = \alpha_A - \delta_A A - p_M k_M MA - p_S \frac{\delta_S k^+}{\delta_S + k^- + Ak^+} N C_{\langle n_\infty \rangle} A \tag{1.50}$$

$$\dot{S} \equiv \dot{C}_{\langle n_\infty \rangle} = \alpha_S - \delta_S C_{\langle n_\infty \rangle} \tag{1.51}$$

Since we switched to a description in which all sRNAs have the same effective binding site occupation, we can perform another simplifying step: Instead of considering $S \equiv C_{\langle n_\infty \rangle}$ sRNAs with N binding sites each, we change to $NC_{\langle n_\infty \rangle}$ sRNAs with a single binding site. In other words, we consider each binding site a separate particle. Then, the total number of sRNAs can then be written as sum of free (unbound) sRNAs (S_{free}) and sRNA-CsrA complexes(C_{AS}).

$$NC_{\langle n_\infty \rangle} = S_{\text{free}} + C_{AS}. \tag{1.52}$$

By "adding a zero", we can write the time derivative of $NC_{\langle n_\infty \rangle}$ as:

$$NC\dot{}_{\langle n_\infty \rangle} = \dot{S}_{\text{free}} + \dot{C}_{AS} \tag{1.53}$$

$$\dot{S}_{\text{free}} = C_{AS}k^- - S_{\text{free}}Ak^+ + N\alpha_S - \delta_S S_{\text{free}} \tag{1.54}$$

$$\dot{C}_{AS} = -C_{AS}k^- + S_{\text{free}}Ak^+ - \delta_S C_{AS} \tag{1.55}$$

Assuming fast complex dynamics, we find:

$$\dot{C}_{AS} = 0 \longrightarrow C_{AS} = \frac{S_{\text{free}}Ak^+}{k^- + \delta_S} \tag{1.56}$$

$$\dot{S}_{\text{free}} = N\alpha_S - kAS_{\text{free}} - \delta_S S_{\text{free}} \tag{1.57}$$

Then it follows:

$$\frac{\delta_S k^+}{\delta_S + k^- + Ak^+} NC_{\langle n_\infty \rangle} = \frac{\delta_S k^+}{\delta_S + k^- + Ak^+}(S_{\text{free}} + C_{AS}) \tag{1.58}$$

$$= \frac{\delta_S k^+}{\delta_S + k^- + Ak^+} S_{\text{free}} \frac{k^- + \delta_S + Ak^+}{k^- + \delta_S} \tag{1.59}$$

$$= \frac{k^+ \delta_S}{k^- + \delta_S} S_{\text{free}} = kS_{\text{free}} \tag{1.60}$$

To obtain a clear and concise notation, we redefine $S_{\text{free}} \to S$ and $k \to k_S$, which leads to the very simple rate equations:

$$\dot{M} = \alpha_M - \delta_M M - k_M MA \tag{1.61}$$

$$\dot{A} = \alpha_A - \delta_A A - k_M p_M MA - k_S p_S AS \tag{1.62}$$

$$\dot{S} = N\alpha_S - \delta_S S - Ak_S S \tag{1.63}$$

1.4   Dimensionless form of the rate equations

It proved beneficial to work with a dimensionless form of the rate equations (3.9)-(3.11). We start by introducing characteristic time and molecule numbers $t = \tau t_c$, $M = m m_c$, $A = a a_c$, $S = s s_c$ and find:

$$m' = \alpha_M \frac{t_c}{m_c} - \delta_M t_c \cdot m - k_M t_c a_c \cdot ma \tag{1.64}$$

$$a' = \alpha_A \frac{t_c}{a_c} - \delta_A t_c \cdot a - p_M k_M t_c m_c \cdot ma - p_S k_S t_c s_c \cdot as \tag{1.65}$$

$$s' = \alpha_S \frac{t_c}{s_c} - \delta_S t_c \cdot s - k_S t_c a_c \cdot as \tag{1.66}$$

Building on these equations, we have various possibilities to proceed. There are 10 parameters, which we could reduce to 6 parameter combinations. Yet, we would still like to count molecule numbers in the same units, i.e. $m_c = a_c = s_c$. Hence, the number of lumped parameters will decrease by two to a final number of 8.

As a next step, we reduce the number of free parameters even further by dividing the differential equations by suitable parameters. This is also an opportunity to introduce small ratios that can be used for an expansion later on. In particular, we take advantage of the fact that sRNAs, mRNAs and proteins in the regulation network have quite different degradation rates. We choose $t_c = \frac{1}{\delta_M}$, which will lead to a small ratio $\delta_{am} := \frac{\delta_A}{\delta_M}$. Later on, we will use this ratio as an expansion parameter.

Furthermore we would like to simplify the interaction terms and choose $m_c = a_c = s_c = \frac{\delta_M}{k_M}$, which results in:

$$m' = \frac{\alpha_M k_M}{\delta_M{}^2} - m - ma \tag{1.67}$$

$$a' = \frac{\alpha_A k_M}{\delta_M{}^2} - \frac{\delta_A}{\delta_M} a - p_M am - p_S \frac{k_S}{k_M} as \tag{1.68}$$

$$s' = \frac{\alpha_S k_S}{\delta_M{}^2} - \frac{\delta_S}{\delta_M} s - \frac{k_S}{k_M} as \tag{1.69}$$

To further simplify we define $\alpha_m := \frac{\alpha_M k_M}{\delta_M{}^2}$, $\alpha_a := \frac{\alpha_A k_M}{\delta_M{}^2}$, $\alpha_s := \frac{\alpha_M k_M}{\delta_M{}^2}$ and $k_{sm} = \frac{k_S}{k_M}$. Thus, we find

$$\boxed{\begin{aligned} m' &= \alpha_m - m - ma & (1.70) \\ a' &= \alpha_a - \delta_{am} a - p_M am - p_S k_{sm} as & (1.71) \\ s' &= \alpha_s - \delta_{sm} s - k_{sm} as & (1.72) \end{aligned}}$$

The coupled equations (1.70)-(1.72) are easier to analyze, compared to the original rate equations, but we have to be aware of the dependencies of our newly defined parameters α_m, α_a and α_s on k_M and δ_M.

1.5 Stationary solution of free long mRNA abundance

We start with equations (1.71) and (1.72) and solve the resulting quadratic equation. We discard the solution with negative molecule numbers and find:

$$a^* = \frac{1}{2 k_{sm}(\delta_{am} + m^* p_M)} \Bigg[-\alpha_s k_{sm} p_S + \alpha_a k_{sm} - \delta_{am}\delta_{sm} - \delta_{sm} m^* p_M + \\ + \sqrt{(\alpha_s k_{sm} p_S - \alpha_a k_{sm} + \delta_{am}\delta_{sm} + \delta_{sm} m^* p_M)^2 + 4\alpha_a \delta_{sm} k_{sm}(\delta_{am} + m^* p_M)} \Bigg] \tag{1.73}$$

$$s^* = \frac{1}{2\delta_{sm} k_{sm} p_S} \Bigg[\alpha_s k_{sm} p_S - \alpha_a k_{sm} - \delta_{am}\delta_{sm} - \delta_{sm} m^* p_M + \\ + \sqrt{(\alpha_s k_{sm} p_S - \alpha_a k_{sm} + \delta_{am}\delta_{sm} + \delta_{sm} m^* p_M)^2 + 4\alpha_a \delta_{sm} k_{sm}(\delta_{am} + m^* p_M)} \Bigg] \tag{1.74}$$

Inserting equation (1.73) into equation (1.70) leaves us with a radical equation for m^*. Isolating the square root and taking the square on both sides yields:

$$
\begin{aligned}
0 = k_{sm}(\delta_{am} + m^* p_M)\big[&\alpha_s k_{sm} m^* p_S (\alpha_m - m^*) + \\
&(\alpha_m k_{sm} + m^*(\delta_{sm} - k_{sm}))(\alpha_m(\delta_{am} + m^* p_M) - m^*(\alpha_a + \delta_{am} + m^* p_M))\big]
\end{aligned}
\tag{1.75}
$$

Since $\delta_{am} > 0$ and $m^* \geq 0$, we can discard the solution $m^* = -\delta_{am}$ as unphysical and divide by this solution. We now expand the equation and sort the terms in the order of m^*:

$$
\begin{aligned}
0 = &(m^*)^3 p_M (k_{sm} - \delta_{sm}) + \\
&(m^*)^2 [\delta_{sm}(\alpha_m p_M - \alpha_a - \delta_{am}) + k_{sm}(\alpha_a - 2\alpha_m p_M - \alpha_s p_S + \delta_{am})] + \\
&(m^*)^1 \alpha_m [k_{sm}(\alpha_m p_M + \alpha_s p_S - \alpha_a) + \delta_{am}(\delta_{sm} - 2k_{sm})] + \\
&(m^*)^0 \alpha_m^2 \delta_{am} k_{sm} \\
=: &(m^*)^3 \mathcal{M}_3 + (m^*)^2 \mathcal{M}_2 + (m^*)^1 \mathcal{M}_1 + \mathcal{M}_0
\end{aligned}
\tag{1.76}
$$

The solutions of this cubic equation can be calculated exactly. Solutions that do not satisfy the original radical solution then have to be discarded. However, the general solution of a cubic equation is very lengthy, and its explicit form does not reveal much of physics or lead to a deeper understanding. That is why we would like to find an easier, approximate solution that allows us to analyze the analytic findings.

1.5.1 Approximation for small and large molecule numbers

For small molecule numbers of long mRNA we can neglect the cubic term in equation (1.76). This leaves us with a quadratic equation whose solution is given by:

$$
m^*_{\ll} = \frac{1}{-2\mathcal{M}_2}\left(\mathcal{M}_1 + \sqrt{\mathcal{M}_1^2 - 4\mathcal{M}_2 \mathcal{M}_0}\right)
\tag{1.77}
$$

For large m^*, we neglect the term $(m*)^0 = 1$ in equation (1.76):

$$
m^*_{\gg} = \frac{1}{2\mathcal{M}_3}\left(-\mathcal{M}_2 - \sqrt{\mathcal{M}_2^2 - 4\mathcal{M}_3 \mathcal{M}_1}\right)
\tag{1.78}
$$

1.5.2 Combined solution

Having found two solutions for the two regimes $m^* \ll 1$ and $m^* \gg 1$, the question arises how these solutions may be combined to form one function that describes the stationary solution of m^* over the whole parameter range. In the two limiting cases, we neglected for small molecule numbers the cubic term $\mathcal{M}_3(m^*)^3$, and for larger molecule numbers the term $\mathcal{M}_0$. When $\mathcal{M}_3(m^*_{\text{trans}})^3 = \mathcal{M}_0$, these two terms have exactly the same magnitude and, as a consequence, yield exactly the same result for m^*. We use this fact to introduce a transition from $m^*_{\ll}$ to $m^*_{\gg}$ and define:

$$
m^* \;:=\; \begin{cases} m^*_{\ll} & ; m^*_{\ll} < m^*_{\text{trans}} \\ m^*_{\gg} & ; \text{else} \end{cases} \qquad \text{for } \delta_{sm} < k_{sm}
\tag{1.79}
$$

$$
m^*_{\text{trans}} \;=\; \left(\frac{\mathcal{M}_0}{\mathcal{M}_3}\right)^{1/3} = \left(\frac{\alpha_m^2 \delta_{am} k_{sm}}{p_M(k_{sm} - \delta_{sm})}\right)^{1/3}
\tag{1.80}
$$

With the solution of m^* and equations (1.73) and (1.74), we obtain a^* and s^*.

There is a minor setback, since in the regime $\delta_{sm} > k_{sm}$ the transition molecule number m^*_{trans} becomes negative. That is why equation (1.79) is only defined for $\delta_{sm} < k_{sm}$.

In the regime $\delta_{sm} > k_{sm}$ we have to follow a different approach. So far, we derived a cubic equation for m^*, but we can also do the same for s^*. Following the same steps as above, while interchanging the roles of m and s, gives rise to another cubic equation:

$$
\begin{aligned}
0 = {} & (s^*)^3 \delta_{sm} k_{sm} p_S (\delta_{sm} - k_{sm}) + \\
& (s^*)^2 (k_{sm}(-\alpha_m \delta_{sm} p_M + \alpha_s p_S(k_{sm} - 2\delta_{sm}) + \alpha_a(\delta_{sm} - k_{sm})) + \delta_{am}\delta_{sm}(\delta_{sm} - k_{sm})) + \\
& (s^*)^1 \alpha_s (k_{sm}(\alpha_m p_M + \alpha_s p_S - \alpha_a) + \delta_{am}(k_{sm} - 2\delta_{sm})) + \\
& (s^*)^0 \alpha_s^2 \delta_{am} \\
=: {} & (s^*)^3 \mathcal{S}_3 + (s^*)^2 \mathcal{S}_2 + (s^*)^1 \mathcal{S}_1 + \mathcal{S}_0
\end{aligned}
$$
(1.81)

In the same way as for the cubic equation in m^*, we can now calculate two limiting solutions for $s^* \ll 1$ and $s^* \gg 1$ that we call $s^*_\ll$ and $s^*_\gg$:

$$
s^*_\ll = \frac{1}{-2\mathcal{S}_2}\left(\mathcal{S}_1 + \sqrt{\mathcal{S}_1^2 - 4\mathcal{S}_2\mathcal{S}_0}\right)
$$
(1.82)

$$
s^*_\gg = \frac{1}{-2\mathcal{S}_3}\left(\mathcal{S}_2 + \sqrt{\mathcal{S}_2^2 - 4\mathcal{S}_3\mathcal{S}_1}\right)
$$
(1.83)

The transition between these solutions takes place at:

$$
s^*_{\text{trans}} = \left(\frac{\alpha_s^2 \delta_{am}}{\delta_{sm}k_{sm}p_S(\delta_{sm} - k_{sm})}\right)^{1/3}
$$
(1.84)

which is positive for $\delta_{sm} > k_{sm}$, as opposed to m^*_{trans}. We further define:

$$
s^* := \begin{cases} s^*_\ll & ; s^*_\ll < s^*_{\text{trans}} \\ s^*_\gg & ; \text{else} \end{cases} \qquad \text{for } \delta_{sm} > k_{sm}
$$
(1.85)

From s^* we may than calculate a^*, and finally m^*. Thus, we are able to find an approximate analytic solution for the whole range of parameters ($\delta_{sm} \gtrless k_{sm}$). Comparison with the numerical solution of the cubic equation shows that the approximation is very exact.

The advantage of our approximate solution compared to the exact solution of the cubic equation (1.76) is twofold. First, due to its simple form we are able to understand the equation and are in the position to predict, for example, the dependence of the threshold on specific parameters (see section 1.6). Second, when studying fluctuations we have to deal with long equations that contain the stationary solutions m^*, a^*, s^* as parameters. The simpler the stationary solutions, the quicker are the calculations.

1.6 Threshold properties of long mRNA expression

In the main text we found a distinct threshold in the expression of long mRNA. In this section we will study the threshold properties of the stationary solution m^* in great detail. The analysis will be based on equation (1.79) in general and equation (1.77) in

particular. Equation (1.77) describes well the stationary solution below threshold, from which we can deduce threshold properties. We already defined:

$$\mathcal{M}_3 = p_M(k_{sm} - \delta_{sm}) \tag{1.86}$$

$$\mathcal{M}_2 = [\delta_{sm}(\alpha_m p_M - \alpha_a - \delta_{am}) + k_{sm}(\alpha_a - 2\alpha_m p_M - \alpha_s p_S + \delta_{am})] \tag{1.87}$$

$$\mathcal{M}_1 = \alpha_m[k_{sm}(\alpha_m p_M + \alpha_s p_S - \alpha_a) + \delta_{am}(\delta_{sm} - 2k_{sm})] \tag{1.88}$$

$$\mathcal{M}_0 = \alpha_m^2 \delta_{am} k_{sm} \tag{1.89}$$

The solution given by (1.79) is only valid in the regime $\delta_{sm} < k_{sm}$. In the regime $\delta_{sm} > k_{sm}$, we would have to work with the approximate solution of s^* (equation (1.85)), which complicates the threshold analysis in m^*. However, there is no obvious reason why the dependence of threshold properties should be different in the two parameter ranges $\delta_{sm} < k_{sm}$ and $\delta_{sm} > k_{sm}$.

1.6.1 Threshold position

For small α_M and α_S, the terms $\mathcal{M}_1$ and $\mathcal{M}_2$ are negative. The negative $\mathcal{M}_1$-term in front of the square root is compensated by the same term squared under the root. Neglecting the second term under the square root, this would result in $m^* = 0$ for $\mathcal{M}_1 < 0$ and linear increase of m^* once $\mathcal{M}_1$ becomes positive. This is how the threshold is encoded in the equations, and we find the threshold position at $\mathcal{M}_1 = 0$:

$$\alpha_{m,\text{th}} = \frac{1}{p_M}\left[\alpha_a - \alpha_s p_S + \delta_{am}\left(2 - \frac{\delta_{sm}}{k_{sm}}\right)\right] \approx \frac{1}{p_M}[\alpha_a - \alpha_s p_S] \tag{1.90}$$

$$\alpha_{a,\text{th}} = \alpha_m p_M + \alpha_s p_S - \delta_{am}\left(2 - \frac{\delta_{sm}}{k_{sm}}\right) \approx \alpha_m p_M + \alpha_s p_S \tag{1.91}$$

From this expression we can deduce that we can shift the threshold to larger α_m if

$p_M, p_S \searrow$: $(1 - p_M)$ and $(1 - p_S)$ are the probabilities that a CsrA dimer survives the degradation of an mRNA-CsrA complex and a CsrA-sRNA complex, respectively. If $p_M = 1$ ($p_S = 1$) the regulation is called non-catalytic. If $p_M = 0$ ($p_S = 0$) the regulation is called catalytic, for in this case CsrA acts as a catalyst for the degradation of its binding partners. Then, the threshold value $\alpha_{m,\text{th}}$ is proportional to $1/p_M$, since $1/p_M$ is the number of long mRNA molecules that are degraded along with 1 CsrA dimer.

$\alpha_a \nearrow$: Increasing α_A leads to a build-up of a larger CsrA buffer that has a greater capability to down-regulate long mRNA expression.

$\alpha_s \searrow$: For smaller α_S less sRNA molecules are produced. It follows that less sRNA molecules may interfere with CsrA dimers.

Since equation (1.90) is linear in all production rates, the statement above holds true not only for dimensionless rates α_s, α_a but as well for dimensionful rates α_M, α_A.

1.7 Comparison with Gillespie simulations

We compared the stationary solution of the rate equations with Gillespie simulations. In all Gillespie simulations the starting molecule numbers where set to $M = A = S = 0$.

We observed that the system needs less than $N_{\text{offset}} = 3000$ reactions to relax to equilibrium. We obtained the mean molecule number $\langle M \rangle$ by time averaging the molecule number M over one run between reaction number N_{offset} and $N_{\text{max}} = 1000000$. Each molecule number M was weighted by the waiting time to the next reaction and summed up. At the end, the result was divided by the total time it took between reaction N_{offset} and N_{max}. This can be expressed as:

$$\langle M \rangle = \frac{1}{T} \sum_{i=N_{\text{offset}}}^{N_{\text{max}}-1} \Delta t_i M_i \tag{1.92}$$

$$\text{Var}[M] = \frac{1}{T} \sum_{i=N_{\text{offset}}}^{N_{\text{max}}-1} \Delta t_i M_i^2 - \langle M \rangle^2 \tag{1.93}$$

where $\Delta t_i = t_{i+1} - t_i$ and $T = t_{N_{\text{max}}} - t_{N_{\text{offset}}}$.

The stationary solution M^* might be different from $\langle M \rangle$:

$$
\begin{aligned}
\langle \dot{M} \rangle = 0 &= \langle \alpha_M \rangle - \langle \delta_M M \rangle - \langle k_M M A \rangle \\
&= \alpha_M - \delta_M \langle M \rangle - k_M \langle [\langle M \rangle + \delta_M][\langle A \rangle + \delta_A] \rangle \\
&= \underbrace{\alpha_M - \delta_M \langle M \rangle - k_M \langle M \rangle \langle A \rangle}_{\text{stationary solution of rate equation}} - k_M [\langle M \rangle \langle \delta_A \rangle + \langle A \rangle \langle \delta_M \rangle + \langle \delta_A \delta_M \rangle]
\end{aligned}
\tag{1.94}
$$

S2 Figure shows very good agreement between the approximative analytical stationary solution of long mRNA abundance and the mean molecule number of mRNA obtained by Gillespie simulations.

1.8 Accounting for additional targets of CsrA

Our study focuses on gene regulation of Colicin E2 release. Therefore, we did not explicitly consider other targets of CsrA (or any component in the *E. coli* cell), although we are aware that CsrA alone can bind to at least over 700 different mRNA targets. The question of how to obtain a simplified biochemical network despite the thousands of different proteins in a living cell, is of very fundamental nature, and remains unsolved. This is particularly critical in our case, since CsrA is a master regulator protein in *E. coli*. However, we still think that it is possible to reduce these system, and want to illustrate, how such a reduction can be done.

In section 1.3 of this Supporting Information, we derived the reduced model from a simplified biochemical network (see S1 Fig). This network comprises five components: The regulator CsrA (A), its target long mRNA (M), the "regulator's regulator" sRNA (S), and the complexes of CsrA with both the long mRNA (C_{MA}) and the sRNA (C_{SA}). For the two types of sRNA (CsrB and CsrC), we derived an effective sRNA, which contains only a single binding site (instead of N binding sites) and thus considerably simplifies the equations (see section 1.3 of the Supporting Information). Employing the effective sRNAs, the biochemical network can be written as the following set of ordinary

differential equations:

$$\text{CsrA:} \quad \dot{A} = \alpha_A + k_S^- C_{SA} + k_M^- C_{MA} +$$
$$+ \delta_{C_{SA}} C_{SA}(1 - p_S) + \delta_{C_{MA}} C_{MA}(1 - p_M)$$
$$- \delta_A A - k_M^+ MA - k_S^+ SA, \tag{1.95}$$

$$\text{long mRNA:} \quad \dot{M} = \alpha_M + k_M^- C_{MA} - \delta_M M - k_M^+ MA, \tag{1.96}$$

$$\text{long mRNA-CsrA complex:} \quad \dot{C}_{MA} = k_M^+ MA - k_M^- C_{MA} - \delta_{C_{ma}} C_{MA}, \tag{1.97}$$

$$\text{sRNA:} \quad \dot{S} = N\alpha_S + k_S^- C_{SA} - \delta_S S - k_S^+ SA, \tag{1.98}$$

$$\text{sRNA-CsrA complex:} \quad \dot{C}_{SA} = k_S^+ SA - k_S^- C_{SA} - \delta_{C_{SA}} C_{SA}. \tag{1.99}$$

These five differential equations describe the temporal change in the abundance of the corresponding quantity. They all contain terms that describe production (α) or degradation (δ) of components, or the formation (k^+) and breaking (k^+) of complexes. Let us shortly recapitulate the biochemical significance of these terms. The first line in the dynamical equation for CsrA (eq. (1.95)) comprises the rate of CsrA production (α_A), and two terms accounting for the increase in CsrA due to CsrA-sRNA- and CsrA-mRNA-complexes breaking up, respectively. The next line contains two terms describing the CsrA increase by the degradation of these two complexes, and include the parameters p_M and p_S, which describe the probability for CsrA to be co-degraded with the complex. Finally, the last line describes terms which reduce CsrA abundance: CsrA decreases either by degradation of CsrA (δ_A), or by forming complexes with long mRNA or sRNA, respectively. The equations for long mRNA and sRNA, eq. (1.96) and (1.98), describe analogous biochemical processes. As the formation of a complex means a decrease in the abundance of the respective complex partners, we find in the equations of the complexes, eq. (1.97) and (1.99), that terms with positive sign in the dynamical equations of A, M or S appear with negative sign the equations for complexes, and vice versa.

In order to account for a new target, we assume that its qualitative behavior is that of long mRNA. This means that in the model its differential equation has the very same structure as that for the long mRNA, but of course with rate parameters specific to the corresponding target. As it would be very unhandy to add over 700 targets to the model, we introduce a single, effective target, T. This additional effective target is an "average" mRNA target, which forms complexes with CsrA. Therefore, accounting for such an effective target adds dynamic equations for the abundance of the target as well as for its complexes,

$$\text{eff. target:} \quad \dot{T} = \alpha_T + k_T^- C_{TA} - k_T^+ AT - \delta_T T \tag{1.100}$$

$$\text{target-CsrA complex:} \quad \dot{C}_{TA} = k_T^+ TA - k_T^- C_{TA} - \delta_{C_{TA}} C_{TA}, \tag{1.101}$$

and also adds new terms to the dynamic equation for CsrA, eq. (1.95):

$$\text{CsrA:} \quad \dot{A} = [\text{r.h.s. of (1.95)}] - k_T^+ TA + k_T^- C_{TA} + \delta_{C_{TA}} C_{TA}(1 - p_T). \tag{1.102}$$

As stated above, the structure of its terms is analogous to those found in the dynamics of the long mRNA: The effective target is produced at rate α_T and degraded at rate $\delta_T T$. Note that α_T is chosen such that the target abundance in the cell matches the combined abundance of the 700 different targets. As with long mRNA, the abundances of A and T get reduced by the formation of CsrA-target-complex ($-k_T^+ TA$), and increased once these complexes either break apart ($+k_T^- C_{TA}$) or get degraded ($+\delta_{C_{TA}} C_{TA}(1 - p_T)$). In the equation for the CsrA-target-complexes, the last three rates appear again with opposite signs.

Eqs.(1.96)-(1.102) now define our initial biochemical network, extended with an effective additional target and its complex with CsrA. We proceed by first considering this system in the steady state. From this state, we can calculate the component abundances. In a second step, we make the simplifying assumptions that all complexes and, in a third step, also the target abundance equilibrate fast compared to other components of the reduced model. This fast-equilibrium-assumption eliminates the affected biochemical processes, and allows us to finally reduce the model to three components.

Steady State. In order to compare the abundances predicted by the extended biochemical network with experimental data, we start by considering the steady state of the system. The steady state is defined as the state, in which no abundance is subject to changes with time. It is obtained by setting the left hand sides of eqs.(1.96)-(1.102) to zero (i.e. $\dot{A} = 0, \dot{M} = 0, ...$). We begin with the equations for the complexes, eqs. (1.97), (1.99) and (1.101). For C_{MA} (eq. (1.97)), we get from $\dot{C} = 0$,

$$C_{MA} = \frac{k_M^+ \cdot M \cdot A}{k_M^- + \delta_{C_{MA}}} = \frac{k_M \cdot M \cdot A}{\delta_{C_{MA}}}, \tag{1.103}$$

where we introduced the effective binding parameters

$$k_M := \frac{k_M^+ \delta_{C_{MA}}}{k_M^- + \delta_{C_{MA}}}. \tag{1.104}$$

As the equations of the complexes, eqs. (1.97), (1.99) and (1.101), have all the very same structure, we can find equations and effective parameters for the sRNA/CsrA- and target/CsrA-complexes analogously. Taken together, these equations read

$$C_{MA} = \frac{k_M \cdot M \cdot A}{\delta_{C_{MA}}}, \qquad k_M := \frac{k_M^+ \delta_{C_{MA}}}{k_M^- + \delta_{C_{MA}}}, \tag{1.105}$$

$$C_{SA} = \frac{k_S \cdot S \cdot A}{\delta_{C_{SA}}}, \qquad k_S := \frac{k_S^+ \delta_{C_{SA}}}{k_S^- + \delta_{C_{SA}}}, \tag{1.106}$$

$$C_{TA} = \frac{k_T \cdot T \cdot A}{\delta_{C_{TA}}}, \qquad k_T := \frac{k_T^+ \delta_{C_{TA}}}{k_T^- + \delta_{C_{TA}}}. \tag{1.107}$$

Inserting these equations to the (steady state) differential equations for A, S, M and T, we obtain an set of coupled equations that is independent of the complex abundances:

$$0 = \alpha_A - \delta_A A - k_M p_M M \cdot A - k_S p_S S \cdot A - k_T p_T A \cdot T, \tag{1.108}$$
$$0 = \alpha_M - \delta_M M - k_M M \cdot A, \tag{1.109}$$
$$0 = N\alpha_S - \delta_S S - k_S S \cdot A, \tag{1.110}$$
$$0 = \alpha_T - \delta_T T - k_T T \cdot A. \tag{1.111}$$

These four equations describe the steady state of the free components A, S, M and T. Note that by employing eqs. (1.105)-(1.107), we were able to combine for each component complex degradation and (un)binding of the complex partners to an effective "coupled degradation" term, e.g. $-k_M p_M M A$ for long mRNA. This step reduces the complexity: The equations for M, S and T now contain only three terms, one each for production (α), degradation (δ) and complex formation with CsrA (k). The dynamic equation for CsrA has the same structure, but a special coupled degradation term: As CsrA forms complexes with each of the three other components (M, S and T), it also has three coupled degradation terms.

Solving our system of equations, eqs. (1.108)-(1.111) numerically allows us to calculate the steady state abundances of the system. To this end, we need to estimate the production, degradation, and binding rates of all the components. We motivate our estimations in chapter 2. Moreover, we assumed $p_M = p_S = p_T = 1$, as CsrA dimers, even if they survive complex degradation, are unlikely to form a complex again after prolonged binding. More specifically, we used the following parameters (in the units molecules/min, 1/min, 1/(molecules · min) for α, γ, k, respectively):

$$
\begin{array}{cccccc}
\text{M:} & \alpha_M = 1 & \delta_M = 0.04 & k_M = 0.5 & p_M = 1 & \delta_{C_M} = \delta_M & (1.112) \\
\text{S:} & N\alpha_S = 57.5 & \delta_S = 0.023 & k_S = 0.5 & p_M = 1 & \delta_{C_S} = \delta_S & (1.113) \\
\text{T:} & \alpha_T = 350 & \delta_T = 0.04 & k_T = 0.5 & p_T = 1 & \delta_{C_T} = \delta_T & (1.114) \\
\text{A:} & \alpha_A = 408.45 & \delta_A = 0.00007 & & & & (1.115)
\end{array}
$$

Note that the production rates α of the components are unknown (see also chapter 2), and are thus treated as free parameters. We chose them such that the numerical solution of eqs. (1.108)-(1.111) with the other parameters in eqs. (1.112)-(1.115) results in

$$
\begin{array}{cccc}
M = 0.01, & S = 0.29, & T = 1.77, & A = 395.45, \\
C_{MA} = 24.99, & C_{TA} = 8748.23, & C_{SA} = 2499.71.
\end{array}
$$

We find that our model consistently predicts not only the abundance of free CsrA (A) as found by Taniguchi et al. [7](474 ± 191 free CsrA molecules), but also its total abundance $A + C_{MA} + C_{TA} + C_{SA}$ and the sRNA ratio, which are given by Gudapathy et al. (11.000-33.000 CsrA molecules in total, 16-32% bound to sRNA [8]). This shows that the abundances in our model reconcile with abundances found in experiments. Moreover, the model extended with the effective targets produced the same abundance of free CsrA as our reduced model. From this we learn that it is indeed justifiable to use a reduced model, which does not account for all possible targets, as the abundance of free CsrA is sufficient to describe our specific regulatory system.

Dynamics. The next steps, which reduce the number of components in the model, are more difficult, and require us to make assumptions on the speed of equilibration of a subset of biochemical processes. Specifically, we assume fast complex equilibration (i.e. $\dot{C}_{SA} = 0, ...$). This assumption is well established in the literature (see, for instance, [9]), and has already been employed in our derivation of the reduced model. It allows us to use eqs. (1.105)-(1.107):

$$
C_{MA} = \frac{k_M \cdot M \cdot A}{\delta_{C_{MA}}},
$$
$$
C_{SA} = \frac{k_S \cdot S \cdot A}{\delta_{C_{SA}}},
$$
$$
C_{TA} = \frac{k_T \cdot T \cdot A}{\delta_{C_{TA}}},
$$

This yields the following set of equations

$$
\dot{A} = \alpha_A - \delta_A A - k_M MA - k_S SA - k_T AT, \qquad (1.116)
$$
$$
\dot{M} = \alpha_M - \delta_M M - k_M MA, \qquad (1.117)
$$
$$
\dot{S} = N\alpha_S - \delta_S S - k_S SA, \qquad (1.118)
$$
$$
\dot{T} = \alpha_T - \delta_T T - k_T TA. \qquad (1.119)
$$

Note that the right hand side of these equations is identical to the right hand side of the steady state equations, eqs.(1.96)-(1.102). These four equations describe the dynamics of our reduced model (as presented in our paper), interacting with a second, effective target for CsrA.

So far, we have showed how it is possible to account for additional targets in a cell in the form of a single, effective target, and derived a reduced four-component model, eqs. (1.116)-(1.119). In order to work with these equations, we must specify the dynamics of the effective target. However, the corresponding rates for most targets are not known, and can only be estimated roughly. It is therefore not useful to explicitly account for these targets in the model. In the following paragraph, we will use the aforementioned rough estimates to reduce eqs. (1.116)-(1.119) back to our three-component model, as the additional terms for the target turn out to be constant above a threshold value of A.

Elimination of the Target Dynamics. In order to eliminate T from eqs. (1.116)-(1.119), we proceed analogous to the elimination of the complexes and make the additional assumption that also the target abundance equilibrates fast. This means that we assume $\dot{T} = 0$ in eq. (1.119), and, just as with the complexes (see, e.g., eq. (1.103)), solve for the target abundance:

$$T = \frac{\alpha_T}{\delta_T + k_T A}. \tag{1.120}$$

We then insert this solution in the differential equation for A, eq. (1.116):

$$\dot{A} = \alpha_A - \delta_A A - k_M M A - k_S S A - k_T \cdot A \cdot \frac{\alpha_T}{\delta_{CT} + k_T A}. \tag{1.121}$$

The equation for A is now independent of T, and eqs. (1.121),(1.117) and (1.118) comprise a closed system of differential equations for three components (just as in our reduced model presented in our main text). If we compare it to our reduced model, we find that the models differ only by a single degradation term in the equation for CsrA. The term reads

$$-k_T \cdot A \cdot \frac{\alpha_T}{\delta_{CT} + k_T A} = -A \cdot \frac{\alpha_T}{\frac{\delta_{CT}}{k_T} + A}, \tag{1.122}$$

and is special for two reasons: First, it is the only term that contains the parameters for the effective targets, and thus describes their influence on the dynamics. Second, its dependence on the parameter A is more complex than for the other terms in eq. (1.121), as it has it has a Langmuir-like dependence on A. Because of the Langmuir functional form, the ratio of δ_{CT} and k_T determines whether the term depends on A or not: If δ_{CT}/k_T is significantly larger than A, it dominates the denominator in eq. (1.122), and the term becomes linearly dependent on A. In the opposite case, if A dominates the denominator, it cancels with the linear A-dependence, rendering the term constant. These two limiting scenarios can be summarized as follows:

$$\frac{\delta_{CT}}{k_T} \ll A: \qquad A \cdot \frac{\alpha_T}{\frac{\delta_{CT}}{k_T} + A} \approx A \cdot \frac{\alpha_T}{A} = \alpha_T, \tag{1.123}$$

$$\frac{\delta_{CT}}{k_T} \gg A: \qquad A \cdot \frac{\alpha_T}{\frac{\delta_{CT}}{k_T} + A} \approx A \cdot \frac{\alpha_T}{\frac{\delta_{CT}}{k_T}} = A \cdot \frac{\alpha_T k_T}{\delta_{CT}}. \tag{1.124}$$

If the parameter sets used in our simulations fall into one of the two limiting scenarios, we could approximate the Langmuir-like term by either eq.(1.123) or (1.124). This

would then simplify the analysis of the equations, and allows us to absorb the term eq. (1.122) into an effective rate of production or degradation, respectively.

Using the parameters defined in eq. (1.114), we find $\frac{\delta_{CT}}{k_T} = 0.08$. For steady state calculations before an SOS signal, that is, when $A \gg 1$, we can thus assume $\frac{\delta_{CT}}{k_T} \ll A$. This means that eq.(1.123) can be applied, and eq. (1.122), becomes the constant α_T. We insert it to eq. (1.121) to get:

$$\dot{A} = \alpha_A - \delta_A A - k_M M A - k_S A S - k_T \cdot A \cdot \frac{\alpha_T}{\delta_{CT} + k_T A} \tag{1.125}$$

$$\approx \alpha_A - \delta_A A - k_M M A - k_S A S - \alpha_T \tag{1.126}$$

$$= (\alpha_A - \alpha_T) - \delta_A A - k_M M A - k_S A S \tag{1.127}$$

$$= \alpha_{A,\text{eff}} - \delta_A A - k_M M A - k_S A S. \tag{1.128}$$

In these steps, we approximated the Langmuir-like term by the constant limiting case, as described above. We then eliminated this now constant term by adding it to the (also constant) production rate α_A, thus defining a new, effective production rate

$$\alpha_{A,\text{eff}} = \alpha_A - \alpha_T.$$

Calculating this effective production rate from the parameters defined in eqs. (1.115) and (1.114), we get $\alpha_{A,\text{eff}} = 58.45$. If we use this value to numerically solve the steady state eqs. (1.128),(1.117) and (1.118), we find that $\alpha_{A,\text{eff}}$ does not reproduce the correct steady state abundances. However, to get the correct values, we have to slightly increase this value to 58.52. This slight difference stems from the approximation of the Langmuir-like term.

With the correction, we get for our three-component system

$$M = 0.01, \qquad\qquad S = 0.30, \qquad\qquad A = 386.44,$$
$$C_{MA} = 24.99, \qquad\qquad C_{SA} = 2499.7,$$

which matches the abundances found in the steady state solution with the targets again very well.

In summary, we derived the three-component system from our main text,

$$\dot{M} = \alpha_M - \delta_M M - k_M M A,$$
$$\dot{A} = \alpha_A - \delta_A A - k_M p_M M A - k_S p_S A S,$$
$$\dot{S} = N \alpha_S - \delta_S S - A k_S S,$$

from our initial biochemical network, eqs. (1.95)-(1.99), which now also accounted for additional targets, eqs.(1.100) and (1.101). We found, that the following set of rate parameters (in the units molecules/min, 1/min, 1/(molecules · min) for α, γ, k, respectively):

M:	$\alpha_M = 1$	$\delta_M = 0.04$	$k_M = 0.5$	$p_M = 1$	$\delta_{C_M} = \delta_M$
S:	$N\alpha_S = 57.5$	$\delta_S = 0.023$	$k_S = 0.5$	$p_M = 1$	$\delta_{C_S} = \delta_S$
A:	$\alpha_A = 58.52$	$\delta_A = 0.00007$			

the model is able to reproduce experimentally observed abundances.

Chapter 2

Biological parameter values in post-transcriptional regulation

The goal of this chapter is to find meaningful parameter values. We follow a three step procedure. First, literature is searched for experimental measurements. Second, if there are no experimental measurements, we will estimate the range of parameter values by looking at similar regulation systems in bacteria. Third, considering the explicit biological processes involved in expression, we will find rough estimates for parameter values and we will check if these values agree with the biological range found in step two.

2.1 Experimental values

At 37°C the half-life of small RNAs CsrB and CsrC are 1.6 min and 4.1 min, respectively [2]. Since the relation between degradation rates and half-life is $\delta = 1/\tau = \ln 2/t_{1/2}$, we find $\delta_{\mathrm{CsrB}} \approx 0.43$ min^{-1} and $\delta_{\mathrm{CsrC}} \approx 0.17$ min^{-1}. However, these values have been measured in experimental conditions that suggest the presence of CsrD, which was shown to be responsible for the degradation of the sRNAs CsrB and CsrC. Recent studies show that CsrB/C decay is activated by the presence of glucose, because glucose leads finally to activation of CsrD [10]. Another recent study demonstrated, that only the unphosphorylated form of EIIAglc (the glucose specific permease of the PTS system) is able to bind to CsrD and activate CsrB/C degradation [11]. In this regard, [12] found that in glucose media, EIIA is unphosphorylated, but phosphorylated in glycerol media. As we want to compare the results of our model to experiments which employ glycerol as the only carbon source, CsrD will not be activated. This implies that we can consider the half-life of the sRNA to be about 30 min [13], which corresponds to $\delta_S = 0.023$ min^{-1}.

Long mRNA decays with a half-life of 18 ± 1.5 min [14], which leads to $\delta_M \approx 0.04$ min^{-1}.

For proteins in *E. coli* we know that "[o]nly a limited portion of the cellular protein is subject to rapid degradation. It decays with a half-life of approximately 1 hour and constitutes 2 to 7% of the total cellular protein" [15]. This classification was refined one year later [16]:

1. $2\% - 7\%$ of all proteins in *E. coli* degrade quickly with a half-life of approximately one hour, meaning $\delta_{\text{fast}} \approx 0.012 \text{ min}^{-1}$.

2. The remainder, i.e. $93\% - 98\%$ of all proteins in *E . coli*, degrades

   (a) under starvation at a rate of $2.5\% - 6\%$ of proteins per hour. Since $N(t) = N_0 e^{-\delta t}$, it follows that
   $\delta_{\text{starve}} = -\ln(0.98 - 0.935) \text{ hr}^{-1} = (0.0034 - 0.001) \text{ min}^{-1}$.

   (b) without starvation at a rate of $0.2\% - 0.6\%$ of proteins per hour. It follows that $\delta_{\text{slow}} = (0.000034 - 0.0001) \text{ min}^{-1}$.

Since we are not interested in conditions under starvation, we have to choose either the degradation rate of fast degrading proteins with $\delta_{\text{fast}} \approx 0.01 \text{ min}^{-1}$, or that of slowly degrading proteins with $\delta_{\text{slow}} \approx 0.00007 \text{ min}^{-1}$. Since CsrA is generally described as very stable [13], we assume $\delta_A = 0.00007 \text{ min}^{-1}$.

In summary, we set the degradation rates as

$$\boxed{\delta_M = 0.04 \text{ min}^{-1} \quad | \quad \delta_A = 0.00007 \text{ min}^{-1} \quad | \quad \delta_S = 0.023 \text{ min}^{-1}}$$

2.2   Estimations from similar system

In [6] a cyanobacterial iron stress response was analyzed and analytic calculations were fitted to experimental data. This lead to an estimate of complex binding parameters. The best-fit parameters for an up-regulated system are as follows:

$$k = \frac{k_{\text{on}} \delta_C}{k_{\text{off}} + \delta_C} \approx \frac{k_{\text{on}} \delta_C}{k_{\text{off}}} \approx 4.4 \text{ nM}^{-1} \text{min}^{-1}.$$

In [5] the target gene *sodB* was regulated by an sRNA, RyhB, that is involved in iron homeostasis of *E. coli*. The complex binding parameters were estimated to

$$k = 0.02 \text{ nM}^{-1} \text{min}^{-1}.$$

The estimates for complex binding parameters deviate in the two different systems by two orders of magnitudes. We will use these results as the biological range for these parameters in our model of post-transcriptional regulation. We choose the mean order of magnitude and take:

$$\boxed{k_M = k_S = 0.5 \text{ min}^{-1} \text{molecule}^{-1}.}$$

Cooperative Binding.   We are aware of the fact that some studies (like dscussed in [17, 18]) suggest that CsrB and CsrC are subject to positive cooperative binding, that is, an increase in binding affinity of a sRNA the more CsrA molecules bind to it. However, we were not able to find reliable quantitative data, which would clearly show that our assumption of fast complex equilibration is void. Indirect ways of analyzing the binding rates, particularly the measurement of K_D values, produce highly varying results, depending on the particular experimental condition used [19, 20]. Since clear evidence for highly cooperative binding interfering with our assumptions is missing, we did not include this phenomenon in our model. In the case that future studies would show that cooperative binding effects of CsrA to sRNA are indeed crucial, our model is still valid, but has to be slightly extended: As the CsrA-sRNA-complexes cannot be considered to equilibrate fast anymore, their abundance C_{SA} must be included explicitly in the model. This would turn the three-component-model to a four-component one.

2.3   Biological estimations

Conversion of units

First, we have to find a conversion between the unit nanomolar (nM), as found in experimental papers, and molecules per cell, as we have used for our estimations. The cell volume of *E. coli* is $V_{EC} = (0.6 - 0.7)\mu m^3$ [21]. Therefore,

$$1\text{nM} = 10^{-9}\,\tfrac{\text{mol}}{\text{dm}^3} = 10^{-9}\,\tfrac{0.65}{1}\,\tfrac{6.02\cdot 10^{23}\ \text{molecules}}{0.65\cdot 10^{15}\mu m^3} \approx \tfrac{0.4\ \text{molecules}}{0.65\mu m^3} = \tfrac{0.4\ \text{molecules}}{V_{EC}}.$$

Production rates

The production rates of CsrA, the sRNAs and the long mRNA, α_A, α_S and α_M, have not been measured, and are thus unknown. In order to obtain plausible values, we fit them such that our model produces component abundances that are found in the literature.

For CsrA, Taniguchi [7] finds an abundance of free molecules of 474 ± 191 per cell. In the reduced model, this value is reached in the steady state if we set $\alpha_A = 58.52\ \text{molecules}/\text{min}$. We show in section 1.8 of this Supporting Information, that this number is in good agreement with an extended model, which also accounts for additional targets of CsrA and correctly reproduces the total CsrA abundance in the cell.

For sRNA, Gudapathy et al. [8] find an abundance of about $S_{\text{exp}} = 250$ CsrB molecules per cell. Moreover, they assume that all of these molecules have formed a complex with CsrA molecules (that is, there is no free sRNA in the system), and that for each sRNA all binding sites are occupied with CsrA. The CsrB molecule is known to have approximately 22 binding sites for CsrA, with $N \approx 10$ CsrA dimers being attached on average [1, 22]. Since our model uses an effective sRNA with only a single binding site, we have to fit the sRNA production rate, $N\alpha_S$, such that our model produces the N-fold amount of sRNAs compared to the abundance found in experiments. In order to get $N \cdot S_{\text{exp}} = 10 \cdot 250$ sRNA complexes, we need to set $N\alpha_S = 57.5\ \text{molecules}/\text{min}$.

For the long mRNA, the abundance in the cell has not been measured yet. Since the gene of long mRNA (2335 nucleotides [23]) is about one order of magnitude larger than the ones of CsrA (183 nucleotides [23]) and the sRNAs (369 nucleotides and 245 *nt* [23]), the transcription of long mRNA will take longer. Consequently, we assumed α_M to be significantly smaller than the production rates of CsrA and sRNA. In our model, we assumed it to be $\alpha_M = 1$.

In summary, we defined the following production rates for long mRNA, CsrA and sRNA:

$\alpha_M = 1\ \text{molecules}/\text{min}$	$\alpha_A = 58.52\ \text{molecules}/\text{min}$	$\alpha_S = 57.5\ \text{molecules}/\text{min}$

Plasmid numbers

Previous studies showed that in a single *E. coli* cell there are approximately $n_{\text{sos}} = 20$ copies of the plasmid containing the colicin operon. The variation in the colcin plasmid number from cell to cell is caused by colicin E2 plasmids being steadily replicated in the cell by a rolling circle replication mechanism. The plasmid copy number enters our

model via the total (per cell) production rate of long mRNA (α_M) as a multiplicative factor: $\alpha_M = \alpha_{M_l} \cdot (n_{\text{sos}} - B_{\text{SOS}})$ (where B_{SOS} denotes the number of plasmids with respressed colicin promoter). Therefore, changes in the plasmid copy number affect our model by either increasing (fewer plasmids) or decreasing (more plasmids) the delay between SOS signal and lysis. However, since most colicin promoters are repressed, even during an SOS signal, the consequences of this effect are limited, compared to changes in the rate parameters. To show the effects of varying plasmid copy numbers, we briefly discuss the lysis time distribution of a population, in which the plasmid copy number is Poisson-distributed with mean $n_{\text{sos}} = 20$ (see S4 Fig B). Compared to its counterpart with fixed n_{sos}, S4 Fig A, we find that the distribution in S4 Fig B is wider. This is due to the effects of variation in n_{sos} we described above: As the population contains cells with plasmid levels both above and below the average, the distribution gets shifted in both directions. However, the comparison with S4 Fig A also shows that the widening of the distribution is rather weak, and that the overall shape of the distribution is largely conserved. This illustrates that variations in plasmid copy number affect the lysis time distribution only weakly. Moreover, the replication of plasmids is the only mechanism that affects their copy number, and happens much slower than any other process considered by our model. Hence, the effect of variation in plasmid copy number on lysis time distributions is expected to be only minor. In order to keep the focus on effects happening on the timescale of SOS responses, we kept the number of colicin plasmids constant, and chose their abundance to be the average value.

Chapter 3

Gillespie simulations

We define the following notation:

- R, Le, Col, L: number of RecA proteins, LexA dimers, Colicin proteins and lysis proteins.

- $M_r M_l, M_s, M$: number of *lexA*, *recA*, short mRNAs and long mRNAs.

- S: number of effective sRNAs with one CsrA binding site.

- $B_r B_l, B_{\text{sos}}$: number of LexA dimers bound to the *lexA*, *recA* and SOS promoter.

- $\alpha_{M_r}, \alpha_{M_l}, \alpha_R, \alpha_{Le}, B_{\text{sos}}, \alpha_{M_s}, \alpha_M, \alpha_A, \alpha_S, \alpha_L$: Production rates ($\alpha$) of the component denoted by the subscript.

- $\delta_{M_r}, \delta_{M_l}, \delta_{Le}, \delta_R, \delta_{M_s}, \delta_M, \delta_A, \delta_S, \delta_L$: Degradation rate of the component denoted by the subscript.

- $k_r^+, k_l^+, k_{sos}^+, k_r^-, k_l^-, k_{sos}^-$: Binding rates ($+$) and unbinding rates ($-$) of LexA dimers to *recA*, *lexA* and SOS promoter sites. The subscript denotes the component.

- k_M, k_S: coupled degradation parameters for the complexes of mRNA and sRNA, respectively

- c_p: Rate of LexA auto-cleavage due to RecA protein.

- n_{sos}: number of ColE2 plasmids.

- $1 - p_M$: Probability of CsrA dimers surviving degradation of sRNA-CsrA complexes.

- $1 - p_S$: Probability of CsrA dimers surviving degradation of mRNA-CsrA complexes.

3.1 Rate equations

With the defined interaction scheme and notation we are now able to introduce the rate equations used by Shimoni [24]:

$$\dot{M_r} = \alpha_{M_r}(1 - B_r) - \delta_{M_r}M_r \tag{3.1}$$

$$\dot{M_l} = \alpha_{M_l}(1 - B_l) - \delta_{M_l}M_l \tag{3.2}$$

$$\dot{R} = \alpha_R M_r - \delta_R R \tag{3.3}$$

$$\dot{Le} = \alpha_{Le}M_l - \delta_{Le}Le - k_l^+(1 - B_l)Le + k_l^- B_l - k_r^+(1 - B_r)Le + k_r^- B_r$$
$$\quad - k_{sos}^+(1 - B_{\text{sos}})Le + k_{sos}^- B_{\text{sos}} - c_p R Le \tag{3.4}$$

$$\dot{B_r} = k_r^+(1 - B_r)Le - k_r^- B_r \tag{3.5}$$

$$\dot{B_l} = k_l^+(1 - B_l)Le - k_l^- B_l \tag{3.6}$$

$$\dot{B_{\text{sos}}} = k_{sos}^+(n_{\text{sos}} - B_{\text{sos}})Le - k_{sos}^- B_{\text{sos}} \tag{3.7}$$

$$\dot{M_s} = \alpha_{M_s}(n_{\text{sos}} - B_{\text{sos}}) - \delta_{M_s}M_s \tag{3.8}$$

$$\dot{M} = \alpha_{M_l}(n_{\text{sos}} - B_{\text{sos}}) - \delta_M M - k_M MA \tag{3.9}$$

$$\dot{A} = \alpha_A - \delta_A A - k_M p_M MA - k_S p_S AS \tag{3.10}$$

$$\dot{S} = \alpha_S - \delta_S S - A k_S S \tag{3.11}$$

3.2 Gillespie simulations

From the rate equations (3.1)-(3.11) we set up a Gillespie simulation [25] with the following reactions:

1. $M_r \xrightarrow{\alpha_{M_r}(1-B_r)} M_r + 1$

2. $M_l \xrightarrow{\alpha_{M_l}(1-B_l)} M_l + 1$

3. $R \xrightarrow{\alpha_R M_r} R + 1$

4. $Le \xrightarrow{\alpha_{Le}M_l} Le + 1$

5. $M_r \xrightarrow{\delta_{M_r}M_r} M_r - 1$

6. $M_l \xrightarrow{\delta_{M_l}M_l} M_l - 1$

7. $R \xrightarrow{\delta_R R} R - 1$

8. $Le \xrightarrow{\delta_{Le}Le} Le - 1$

9. $Le, B_r \xrightarrow{k_r^+(1-B_r)Le} Le - 1, B_r + 1$

10. $Le, B_l \xrightarrow{k_l^+(1-B_l)Le} Le - 1, B_l + 1$

11. $Le, B_r \xrightarrow{k_r^- B_r} Le + 1, B_r - 1$

12. $Le, B_l \xrightarrow{k_l^- B_l} Le + 1, B_l - 1$

13. $Le \xrightarrow{c_p R Le} Le - 1$

14. $Le, B_{\text{sos}} \xrightarrow{k_{sos}^+(n_{\text{sos}}-B_{\text{sos}})Le} Le - 1, B_{\text{sos}} + 1$

15. $Le, B_{\text{sos}} \xrightarrow{k_{sos}^- B_{\text{sos}}} Le + 1, B_{\text{sos}} - 1$

16. $M_s \xrightarrow{\alpha_{M_s}(n_{\text{sos}}-B_{\text{sos}})} M_s + 1$

17. $M_s \xrightarrow{\delta_{M_s}M_s} M_s - 1$

18. $M \xrightarrow{\alpha_{M_l}(n_{\text{sos}}-B_{\text{sos}})} M + 1$

19. $M \xrightarrow{\delta_M M} M - 1$

20. $A \xrightarrow{\alpha_A} A + 1$

21. $A \xrightarrow{\delta_A A} A - 1$

22. $S \xrightarrow{\alpha_S} S + 1$

23. $S \xrightarrow{\delta_S S} S - 1$

24. $(M, A) \xrightarrow{k_M p_M M A} (M - 1, A - 1)$

25. $(M, A) \xrightarrow{k_M (1 - p_M) M A} (M - 1, A)$

26. $(A, S) \xrightarrow{k_S p_S A S} (A - 1, S - 1)$

27. $(A, S) \xrightarrow{k_S (1 - p_S) A S} (A, S - 1)$
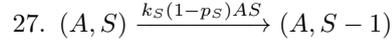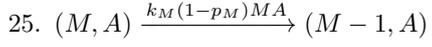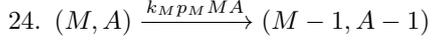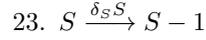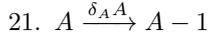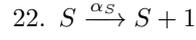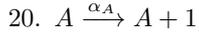
The parameter values are shown in S1 Table. The values from literature in this table were taken from [2, 14–16, 24, 26]. Estimated parameter values were chosen according to [5, 6, 24]. For the transcription rates of long mRNA, CsrA and sRNA we calculated a rough estimate using the transcription rate of RNA polymerase [27] and the length of the individual genes [23], taking into account the number of Colicin plasmids, CsrA binding sites on CsrB, and the translational burst size for CsrA. All parameters are given in the unit of molecules per cell and minute.

Chapter 4

Linear noise approximation

4.1 Definitions

The state vector $\vec{x} = (X_1, X_2, \ldots, X_N)^T$ gives the copy numbers of the N components involved.

There are M reactions with rates $\vec{W}(\vec{x}) = (W_1(\vec{x}), W_2(\vec{x}), \ldots, W_M(\vec{x}))^T$.

The matrix $\mathcal{A}$ with components a_{ij} gives the change in copy number of component i following reaction j.

4.2 Master equation and rate equation

With the definitions above the master equation is given by:

$$\frac{d}{dt} P(\vec{x}, t) = \sum_{j=1}^{M} [W_j(\vec{x} - \vec{a}_j) P(\vec{x} - \vec{a}_j, t) - W_j(\vec{x}) P(\vec{x}, t)] \qquad (4.1)$$

The Master equation 4.1 gives rise to the time evolution of the first moment:

$$\langle \dot{\vec{x}} \rangle = \langle \mathcal{A} \vec{W}(\vec{x}) \rangle \approx \mathcal{A} \vec{W}(\langle \vec{x} \rangle) \qquad (4.2)$$

In the last step we have neglected correlations. If all reaction rates in vector $\vec{W}$ were linear, an equal sign would hold true. Equation (4.2) with neglected correlations is the deterministic rate equation of the system.

4.3 Kramers-Moyal expansion and van Kampen's expansion

The master equation (4.1) is a set of N_x coupled ordinary differential equations (ODEs), where N_x is the number of states in the system. There is a large number of states, since each set of copy numbers corresponds to one individual state. This makes it very hard

to acquire useful information directly from the master equation. A master equation is often approximated by a Kramers-Moyal expansion, which converts the set of N_x coupled ODEs to one partial differential equation of order $i_{\max}$:

$$\frac{d}{dt}P(\vec{x},t) = \partial_t P(\vec{x},t) \approx \sum_{i=1}^{i_{\max}} \left(\frac{(-1)^i}{i!} \prod_{j=1}^{i} (\partial x_{k_j}) \left[\prod_{j=1}^{i} (a_{k_j l}) W_l(\vec{x}) P(\vec{x},t) \right] \right) \tag{4.3}$$

For the sake of concise notation, we define $\vec{\mathcal{X}} = \langle \vec{x} \rangle$, so $\partial_t \vec{\mathcal{X}} = \mathcal{A}\vec{W}(\vec{\mathcal{X}})$. Next, we introduce a new random variable $\vec{\xi}$, which gives the fluctuations around the deterministic trajectory given by the rate equations:

$$\vec{x} = \vec{\mathcal{X}} + \vec{\xi} \qquad \text{with } \vec{\xi} = \mathcal{O}\left(\sqrt{|\vec{x}|}\right) \tag{4.4}$$

It is important that fluctuations scale with the square root of the mean, because van Kampen's expansion is only valid if fluctuations are in the vicinity of the deterministic rate equation. Contrary to the van Kampen expansion typically found in textbooks [4, 28], we neglect the system size parameter Ω used for scaling arguments at this point. The reasons will become clear in section 4.6.

The probability density in the new random variable $\vec{\xi}$ relates to the probability density in the random variable $\vec{x}$ as

$$\pi(\vec{\xi},t) = P(\vec{x},t) = P(\vec{\mathcal{X}}(t) + \vec{\xi},t). \tag{4.5}$$

Consequently, we find:

$$\begin{aligned}
\partial_t \pi(\vec{\xi},t) &= \partial_{x_i} P(\vec{x},t) \frac{dx_i(t)}{dt} + \partial_t P(\vec{x},t) \\
&= \partial_{x_i} P(\vec{x},t) \frac{d\mathcal{X}_i(t)}{dt} + \sum_{i=1}^{i_{\max}} \left(\frac{(-1)^i}{i!} \prod_{j=1}^{i} (\partial x_{k_j}) \left[\prod_{j=1}^{i} (a_{k_j l}) W_l(\vec{x}) P(\vec{x},t) \right] \right) \\
&= \partial_{\xi_i} \pi(\vec{\xi},t) a_{ij} W_j(\vec{\mathcal{X}}) \\
&\quad + \sum_{i=1}^{i_{\max}} \left(\frac{(-1)^i}{i!} \prod_{j=1}^{i} (\partial \xi_{k_j}) \left[\prod_{j=1}^{i} (a_{k_j l}) W_l(\vec{\mathcal{X}} + \vec{\xi}) \pi(\vec{\xi},t) \right] \right) \\
&= -\partial_{\xi_i} \left(\left[a_{il} W_l(\vec{\mathcal{X}} + \vec{\xi}) - a_{il} W_l(\vec{\mathcal{X}}) \right] \pi(\vec{\xi},t) \right) \\
&\quad + \sum_{i=2}^{i_{\max}} \left(\frac{(-1)^i}{i!} \prod_{j=1}^{i} (\partial \xi_{k_j}) \left[\prod_{j=1}^{i} (a_{k_j l}) W_l(\vec{\mathcal{X}} + \vec{\xi}) \pi(\vec{\xi},t) \right] \right).
\end{aligned} \tag{4.6}$$

To perform these calculations, we used several times the equality

$$\begin{aligned}
\partial_{\xi_i} \left[\pi(\vec{\xi},t) f(\vec{x}(\vec{\xi})) \right] &= \partial_{\xi_i} \left[P(\vec{x}(\vec{\xi}),t) f(\vec{x}(\vec{\xi})) \right] = \partial_{x_j} \left[P(\vec{x},t) f(\vec{x}) \right] \underbrace{\frac{dx_j}{d\xi_i}}_{\delta_{i,j}} \\
&= \partial_{x_i} \left[P(\vec{x},t) f(\vec{x}) \right].
\end{aligned} \tag{4.7}$$

The basic assumption here is that fluctuations $\vec{\xi}$ around the mean $\vec{\mathcal{X}}$ are expected to scale with the square root of the mean as denoted in equation (4.4). It follows that for large $\vec{\xi}$ and a sufficiently smooth reaction rate vector, we may perform a Taylor expansion:

$$W_l(\vec{\mathcal{X}} + \vec{\xi}) = \sum_{u=0}^{\infty} \frac{1}{u!} \prod_{v=1}^{u} (\partial \mathcal{X}_{k_v}) W_l(\vec{\mathcal{X}}) \prod_{v=1}^{u} (\xi_{k_v}). \tag{4.8}$$

32

4.4 Simplification for post-transcriptional regulation of Colicin release

So far, the analysis holds true for all systems that can be cast into the form defined in section 4.1. Let us now turn to our specific model of post-transcriptional regulation of Colicin release. With the reactions defined in chapter 1.7 we find:

$$\mathcal{A} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & -1 & -1 \end{pmatrix} \quad (4.9)$$

$$\vec{W}(M, A, S) = (\alpha_M, \delta_M M, \alpha_A, \delta_A A, \alpha_S, \delta_S S, k_M p_M M A,$$
$$k_M(1 - p_M)MA, k_S p_S AS, k_S(1 - p_S)AS)^T \quad (4.10)$$

From equation (4.10) we see that all derivatives higher than second order must vanish, which simplifies the sum in (4.8) significantly, i.e. u can take the values $0, 1$ or 2.

4.5 Calculation of moments

With equation (4.6) we are in the position to calculate the moments of the random variable $\vec{\xi}$. We do so by integrating equation (4.6) multiplied by the random variables whose moment is calculated by parts. Terms containing the expression $\prod_{j=1}^{i}(\partial_{\xi_{k_j}})$ have to be integrated by parts i times. After integration, only a few terms are non-zero. For the first two moments we obtain:

$$\partial_t \langle\!\langle \xi_b \rangle\!\rangle = \int_{-\infty}^{\infty} d\xi_1 d\xi_2 \ldots d\xi_n \partial_t \pi(\vec{\xi}, t) \xi_b =$$
$$= a_{bl} \sum_{u=1}^{\infty} \frac{1}{u!} \prod_{v=1}^{u}(\partial_{\chi_{k_v}}) W_l(\vec{\mathcal{X}}) \langle\!\langle \prod_{v=1}^{u}(\xi_{k_v}) \rangle\!\rangle \quad (4.11)$$

$$\partial_t \underbrace{\langle\!\langle \xi_b \xi_c \rangle\!\rangle}_{\text{symm in } b,c} = \int_{-\infty}^{\infty} d\xi_1 d\xi_2 \ldots d\xi_n \partial_t \pi(\vec{\xi}, t) \xi_b \xi_c =$$

$$= a_{bl} \sum_{u=1}^{\infty} \frac{1}{u!} \prod_{v=1}^{u}(\partial_{\chi_{k_v}}) W_l(\vec{\mathcal{X}}) \langle\!\langle \prod_{v=1}^{u}(\xi_{k_v}) \xi_c \rangle\!\rangle \quad (4.12)$$

$$+ a_{cl} \sum_{u=1}^{\infty} \frac{1}{u!} \prod_{v=1}^{u}(\partial_{\chi_{k_v}}) W_l(\vec{\mathcal{X}}) \langle\!\langle \prod_{v=1}^{u}(\xi_{k_v}) \xi_b \rangle\!\rangle \quad (4.13)$$

$$+ a_{bl} a_{cl} \sum_{u=0}^{\infty} \frac{1}{u!} \prod_{v=1}^{u}(\partial_{\chi_{k_v}}) W_l(\vec{\mathcal{X}}) \langle\!\langle \prod_{v=1}^{u}(\xi_{k_v}) \rangle\!\rangle \quad (4.14)$$

To calculate the third moments, we will need the following relations:

$$\partial_{\xi_i}(\xi_b\xi_c\xi_d) = \delta_{ib}\xi_c\xi_d + \delta_{ic}\xi_b\xi_d + \delta_{id}\xi_b\xi_c \tag{4.15}$$

$$\prod_{j=1}^{2}(\partial_{\xi_{k_j}})(\xi_b\xi_c\xi_d) = \xi_b(\delta_{k_1c}\delta_{k_2d} + \delta_{k_2c}\delta_{k_1d}) + \xi_c(\delta_{k_1b}\delta_{k_2d} + \delta_{k_2b}\delta_{k_1d})$$
$$+ \xi_d(\delta_{k_1b}\delta_{k_2c} + \delta_{k_2b}\delta_{k_1c}) \tag{4.16}$$

$$\prod_{j=1}^{3}(\partial_{\xi_{k_j}})(\xi_b\xi_c\xi_d) = \sum_{\hat{P}(b,c,d)} \delta_{k_1b}\delta_{k_2c}\delta_{k_3d} \tag{4.17}$$

$$\prod_{j=1}^{4}(\partial_{\xi_{k_j}})(\xi_b\xi_c\xi_d) = 0 \tag{4.18}$$

The operator $\hat{P}(b,c,d)$ signifies all permutations in (b,c,d). It follows:

$$\partial_t \underbrace{\ll \xi_b\xi_c\xi_d \gg}_{\text{symm in } b,c,d} = \int_{-\infty}^{\infty} d\xi_1 d\xi_2 \ldots d\xi_n \partial_t \pi(\vec{\xi},t)\xi_b\xi_c\xi_d$$

$$= -\int_{-\infty}^{\infty} d\vec{\xi}\,\partial_{\xi_i}\left(\left[a_{il}W_l(\vec{\mathcal{X}}+\vec{\xi}) - a_{il}W_l(\vec{\mathcal{X}})\right]\pi(\vec{\xi},t)\right)\xi_b\xi_c\xi_d$$

$$+ \int_{-\infty}^{\infty} d\vec{\xi}\sum_{i=2}^{\infty}\left(\frac{(-1)^i}{i!}\prod_{j=1}^{i}(\partial_{\xi_{k_j}})\left[\prod_{j=1}^{i}(a_{k_jl})W_l(\vec{\mathcal{X}}+\vec{\xi})\pi(\vec{\xi},t)\right]\right)\xi_b\xi_c\xi_d$$

$$\stackrel{PI}{=} \int_{-\infty}^{\infty} d\vec{\xi}\,\partial_{\xi_i}(\xi_b\xi_c\xi_d)\left(\left[a_{il}W_l(\vec{\mathcal{X}}+\vec{\xi}) - a_{il}W_l(\vec{\mathcal{X}})\right]\pi(\vec{\xi},t)\right)$$

$$+ \int_{-\infty}^{\infty} d\vec{\xi}\prod_{j=1}^{2}(\partial_{\xi_{k_j}})(\xi_b\xi_c\xi_d)\left(\frac{1}{2}\underbrace{\prod_{j=1}^{2}(a_{k_jl})}_{\text{symm in } k_1,k_2} W_l(\vec{\mathcal{X}}+\vec{\xi})\pi(\vec{\xi},t)\right)$$

$$+ \int_{-\infty}^{\infty} d\vec{\xi}\prod_{j=1}^{3}(\partial_{\xi_{k_j}})(\xi_b\xi_c\xi_d)\left(\frac{1}{6}\underbrace{\prod_{j=1}^{3}(a_{k_jl})}_{\text{symm in } k_1,k_2,k_3} W_l(\vec{\mathcal{X}}+\vec{\xi})\pi(\vec{\xi},t)\right)$$

$$+ \int_{-\infty}^{\infty} d\vec{\xi}\prod_{j=1}^{4}(\partial_{\xi_{k_j}})(\xi_b\xi_c\xi_d)$$

$$\sum_{i=4}^{\infty}\left(\frac{(-1)^i}{i!}\prod_{j=5}^{i}(\partial_{\xi_{k_j}})\left[\prod_{j=1}^{i}(a_{k_jl})W_l(\vec{\mathcal{X}}+\vec{\xi})\pi(\vec{\xi},t)\right]\right)$$

$$= a_{bl}\sum_{u=1}^{\infty}\frac{1}{u!}\prod_{v=1}^{u}(\partial_{\mathcal{X}_{k_v}})W_l(\vec{\mathcal{X}})\ll \prod_{v=1}^{u}(\xi_{k_v})\xi_c\xi_d \gg \tag{4.19}$$

$$+ a_{cl}\sum_{u=1}^{\infty}\frac{1}{u!}\prod_{v=1}^{u}(\partial_{\mathcal{X}_{k_v}})W_l(\vec{\mathcal{X}})\ll \prod_{v=1}^{u}(\xi_{k_v})\xi_b\xi_d \gg \tag{4.20}$$

$$+ a_{dl}\sum_{u=1}^{\infty}\frac{1}{u!}\prod_{v=1}^{u}(\partial_{\mathcal{X}_{k_v}})W_l(\vec{\mathcal{X}})\ll \prod_{v=1}^{u}(\xi_{k_v})\xi_b\xi_c \gg \tag{4.21}$$

$$+ a_{cl}a_{dl}\sum_{u=0}^{\infty}\frac{1}{u!}\prod_{v=1}^{u}(\partial_{\mathcal{X}_{k_v}})W_l(\vec{\mathcal{X}})\ll \prod_{v=1}^{u}(\xi_{k_v})\xi_b \gg \tag{4.22}$$

34

$$+ a_{bl}a_{dl}\sum_{u=0}^{\infty}\frac{1}{u!}\prod_{v=1}^{u}(\partial_{\mathcal{X}_{k_v}})W_l(\vec{\mathcal{X}}) \ll \prod_{v=1}^{u}(\xi_{k_v})\xi_c \gg \tag{4.23}$$

$$+ a_{bl}a_{cl}\sum_{u=0}^{\infty}\frac{1}{u!}\prod_{v=1}^{u}(\partial_{\mathcal{X}_{k_v}})W_l(\vec{\mathcal{X}}) \ll \prod_{v=1}^{u}(\xi_{k_v})\xi_d \gg \tag{4.24}$$

$$+ a_{bl}a_{cl}a_{dl}\sum_{u=0}^{\infty}\frac{1}{u!}\prod_{v=1}^{u}(\partial_{\mathcal{X}_{k_v}})W_l(\vec{\mathcal{X}}) \ll \prod_{v=1}^{u}(\xi_{k_v}) \gg \tag{4.25}$$

The same steps of calculation can be applied to the fourth moment as well. To avoid too longish expressions, we just state the result:

$$\partial_t \underbrace{\ll \xi_b\xi_c\xi_d\xi_e \gg}_{\text{symm in } b,c,d,e} = \int_{-\infty}^{\infty} d\xi_1 d\xi_2 \ldots d\xi_n \partial_t \pi(\vec{\xi},t)\xi_b\xi_c\xi_d\xi_e$$

$$= a_{bl}\sum_{u=1}^{\infty}\frac{1}{u!}\prod_{v=1}^{u}(\partial_{\mathcal{X}_{k_v}})W_l(\vec{\mathcal{X}}) \ll \prod_{v=1}^{u}(\xi_{k_v})\xi_c\xi_d\xi_e \gg \tag{4.26}$$

$$+ a_{cl}\sum_{u=1}^{\infty}\frac{1}{u!}\prod_{v=1}^{u}(\partial_{\mathcal{X}_{k_v}})W_l(\vec{\mathcal{X}}) \ll \prod_{v=1}^{u}(\xi_{k_v})\xi_b\xi_c\xi_d \gg \tag{4.27}$$

$$+ a_{dl}\sum_{u=1}^{\infty}\frac{1}{u!}\prod_{v=1}^{u}(\partial_{\mathcal{X}_{k_v}})W_l(\vec{\mathcal{X}}) \ll \prod_{v=1}^{u}(\xi_{k_v})\xi_b\xi_c\xi_e \gg \tag{4.28}$$

$$+ a_{el}\sum_{u=1}^{\infty}\frac{1}{u!}\prod_{v=1}^{u}(\partial_{\mathcal{X}_{k_v}})W_l(\vec{\mathcal{X}}) \ll \prod_{v=1}^{u}(\xi_{k_v})\xi_b\xi_c\xi_d \gg \tag{4.29}$$

$$+ a_{bl}a_{cl}\sum_{u=0}^{\infty}\frac{1}{u!}\prod_{v=1}^{u}(\partial_{\mathcal{X}_{k_v}})W_l(\vec{\mathcal{X}}) \ll \prod_{v=1}^{u}(\xi_{k_v})\xi_d\xi_e \gg \tag{4.30}$$

$$+ a_{bl}a_{dl}\sum_{u=0}^{\infty}\frac{1}{u!}\prod_{v=1}^{u}(\partial_{\mathcal{X}_{k_v}})W_l(\vec{\mathcal{X}}) \ll \prod_{v=1}^{u}(\xi_{k_v})\xi_c\xi_e \gg \tag{4.31}$$

$$+ a_{bl}a_{el}\sum_{u=0}^{\infty}\frac{1}{u!}\prod_{v=1}^{u}(\partial_{\mathcal{X}_{k_v}})W_l(\vec{\mathcal{X}}) \ll \prod_{v=1}^{u}(\xi_{k_v})\xi_c\xi_d \gg \tag{4.32}$$

$$+ a_{cl}a_{dl}\sum_{u=0}^{\infty}\frac{1}{u!}\prod_{v=1}^{u}(\partial_{\mathcal{X}_{k_v}})W_l(\vec{\mathcal{X}}) \ll \prod_{v=1}^{u}(\xi_{k_v})\xi_b\xi_e \gg \tag{4.33}$$

$$+ a_{cl}a_{el}\sum_{u=0}^{\infty}\frac{1}{u!}\prod_{v=1}^{u}(\partial_{\mathcal{X}_{k_v}})W_l(\vec{\mathcal{X}}) \ll \prod_{v=1}^{u}(\xi_{k_v})\xi_b\xi_d \gg \tag{4.34}$$

$$+ a_{dl}a_{el}\sum_{u=0}^{\infty}\frac{1}{u!}\prod_{v=1}^{u}(\partial_{\mathcal{X}_{k_v}})W_l(\vec{\mathcal{X}}) \ll \prod_{v=1}^{u}(\xi_{k_v})\xi_b\xi_c \gg \tag{4.35}$$

$$+ a_{bl}a_{cl}a_{dl}\sum_{u=0}^{\infty}\frac{1}{u!}\prod_{v=1}^{u}(\partial_{\mathcal{X}_{k_v}})W_l(\vec{\mathcal{X}}) \ll \prod_{v=1}^{u}(\xi_{k_v})\xi_e \gg \tag{4.36}$$

$$+ a_{bl}a_{cl}a_{el}\sum_{u=0}^{\infty}\frac{1}{u!}\prod_{v=1}^{u}(\partial_{\mathcal{X}_{k_v}})W_l(\vec{\mathcal{X}}) \ll \prod_{v=1}^{u}(\xi_{k_v})\xi_d \gg \tag{4.37}$$

$$+ a_{bl}a_{dl}a_{el}\sum_{u=0}^{\infty}\frac{1}{u!}\prod_{v=1}^{u}(\partial_{\mathcal{X}_{k_v}})W_l(\vec{\mathcal{X}}) \ll \prod_{v=1}^{u}(\xi_{k_v})\xi_c \gg \tag{4.38}$$

$$+ a_{cl}a_{dl}a_{el}\sum_{u=0}^{\infty}\frac{1}{u!}\prod_{v=1}^{u}(\partial_{\mathcal{X}_{k_v}})W_l(\vec{\mathcal{X}}) \ll \prod_{v=1}^{u}(\xi_{k_v})\xi_b \gg \tag{4.39}$$

35

$$+ a_{bl}a_{cl}a_{dl}a_{el} \sum_{u=0}^{\infty} \frac{1}{u!} \prod_{v=1}^{u}(\partial_{\mathcal{X}_{k_v}})W_l(\vec{\mathcal{X}}) \ll \prod_{v=1}^{u}(\xi_{k_v}) \gg \qquad (4.40)$$

4.6 Scaling of terms in the equations of moments

We have neglected the system size parameter Ω so far, which in fact seems odd, since van Kampen's expansion is also known as the Ω-expansion. There are two important points why we have followed this procedure:

1. It is indeed possible to find a parameter Ω that fulfills the requirement needed for an Ω-expansion, namely $(\dot{M}, \dot{A}, \dot{S})^T = \vec{F}(M, A, S) = \Omega\vec{f}(M/\Omega, A/\Omega, S/\Omega)$. Ω has to be a large quantity proportional to the system size. However, inserting real parameters, we find $\Omega \ll 1$. We conclude, that the parameter Ω is not well-defined in our model, but would be only an artificial construct.

2. When looking at the scaling properties of $\langle\langle\xi\rangle\rangle, \langle\langle\xi^2\rangle\rangle, \ldots$ we cannot simply group all terms of the same order in $\sqrt{\Omega}^i$, $i \in \mathbf{N}$, since all of these terms come in combination with $\langle\langle\xi\rangle\rangle, \langle\langle\xi^2\rangle\rangle, \ldots$ terms which in turn have a specific scaling property, as we shall see.

In the following, we define the system size parameter $\tilde{\Omega}$ as the total number of molecules present in the system. With this definition we will work out the scaling properties of $\langle\langle\xi\rangle\rangle, \langle\langle\xi^2\rangle\rangle, \langle\langle\xi^3\rangle\rangle$ and $\langle\langle\xi^4\rangle\rangle$ in the system size parameter $\tilde{\Omega}$ as well as the significance of each term in the equations for the moments.

Scaling of specific terms:

- The stoichiometric matrix A_{bl} scales with $\mathcal{O}(1)$.

- The reaction rate matrix is quadratic in the molecule numbers, such that $\vec{W}(\vec{\mathcal{X}})$ scales with $\mathcal{O}(\tilde{\Omega}^2)$.

- Each derivative with respect to $\mathcal{X}_i$ introduces a factor $\mathcal{O}(1/\tilde{\Omega})$.

The equations derived for the moments (see section 4.5) may be classified into terms with equal scaling behavior. The scaling behavior depends on u (0, 1 or 2) and the order of the moment. Since we are interested in the stationary values of fluctuations, we can set all derivatives with respect to time equal to zero, and find:

Equations	u=0	u=1	u=2
(4.11)	-	$\mathcal{O}(\tilde{\Omega})\langle\langle\xi\rangle\rangle$	$\mathcal{O}(1)\langle\langle\xi^2\rangle\rangle$

Table 4.1. Scaling of terms for first moments

Equations	u=0	u=1	u=2
(4.12)-(4.13)	-	$\mathcal{O}(\tilde{\Omega})\langle\langle\xi^2\rangle\rangle$	$\mathcal{O}(1)\langle\langle\xi^3\rangle\rangle$
(4.14)	$\mathcal{O}(\tilde{\Omega}^2)$	$\mathcal{O}(\tilde{\Omega})\langle\langle\xi\rangle\rangle$	$\mathcal{O}(1)\langle\langle\xi^2\rangle\rangle$

Table 4.2. Scaling of terms for second moments

Equations	u=0	u=1	u=2
(4.19)-(4.21)	-	$\mathcal{O}(\tilde{\Omega})\langle\langle\xi^3\rangle\rangle$	$\mathcal{O}(1)\langle\langle\xi^4\rangle\rangle$
(4.22)-(4.24)	$\mathcal{O}(\tilde{\Omega}^2)\langle\langle\xi\rangle\rangle$	$\mathcal{O}(\tilde{\Omega})\langle\langle\xi^2\rangle\rangle$	$\mathcal{O}(1)\langle\langle\xi^3\rangle\rangle$
(4.25)	$\mathcal{O}(\tilde{\Omega}^2)$	$\mathcal{O}(\tilde{\Omega})\langle\langle\xi\rangle\rangle$	$\mathcal{O}(1)\langle\langle\xi^2\rangle\rangle$

Table 4.3. Scaling of terms for third moments

Equations	u=0	u=1	u=2
(4.26)-(4.29)	-	$\mathcal{O}(\tilde{\Omega})\langle\langle\xi^4\rangle\rangle$	$\mathcal{O}(1)\langle\langle\xi^5\rangle\rangle$
(4.30)-(4.35)	$\mathcal{O}(\tilde{\Omega}^2)\langle\langle\xi^2\rangle\rangle$	$\mathcal{O}(\tilde{\Omega})\langle\langle\xi^3\rangle\rangle$	$\mathcal{O}(1)\langle\langle\xi^4\rangle\rangle$
(4.36)-(4.39)	$\mathcal{O}(\tilde{\Omega}^2)\langle\langle\xi\rangle\rangle$	$\mathcal{O}(\tilde{\Omega})\langle\langle\xi^2\rangle\rangle$	$\mathcal{O}(1)\langle\langle\xi^3\rangle\rangle$
(4.40)	$\mathcal{O}(\tilde{\Omega}^2)$	$\mathcal{O}(\tilde{\Omega})\langle\langle\xi\rangle\rangle$	$\mathcal{O}(1)\langle\langle\xi^2\rangle\rangle$

Table 4.4. Scaling of terms for fourth moments

Looking at the dominant terms in the tables 4.1, 4.2, 4.3 and 4.4, we deduce:

$$\langle\langle\xi\rangle\rangle = \mathcal{O}(1/\tilde{\Omega})\langle\langle\xi^2\rangle\rangle \tag{4.41}$$

$$\langle\langle\xi^2\rangle\rangle = \mathcal{O}(\tilde{\Omega}) + \mathcal{O}(1/\tilde{\Omega})\langle\langle\xi^3\rangle\rangle \tag{4.42}$$

$$\langle\langle\xi^3\rangle\rangle = \mathcal{O}(\tilde{\Omega}) + \mathcal{O}(1)\langle\langle\xi^2\rangle\rangle + \mathcal{O}(1/\tilde{\Omega})\langle\langle\xi^4\rangle\rangle \tag{4.43}$$

$$\langle\langle\xi^4\rangle\rangle = \mathcal{O}(\tilde{\Omega})\langle\langle\xi^2\rangle\rangle + \mathcal{O}(1/\tilde{\Omega})\langle\langle\xi^5\rangle\rangle \tag{4.44}$$

Since van Kampen's expansion is only valid for small noise, we expect ξ to be of order $\mathcal{O}\sqrt{(\tilde{\Omega})}$ and thus $\langle\langle\xi^5\rangle\rangle$ is smaller than of order $\mathcal{O}(\tilde{\Omega}^3)$.

Hence, we find: $\boxed{\langle\langle\xi\rangle\rangle = \mathcal{O}(1), \ \langle\langle\xi^2\rangle\rangle = \mathcal{O}(\tilde{\Omega}), \ \langle\langle\xi^3\rangle\rangle = \mathcal{O}(\tilde{\Omega}), \ \langle\langle\xi^4\rangle\rangle = \mathcal{O}(\tilde{\Omega}^2)}$.

Using this result, we marked all terms in the tables above of order $\mathcal{O}(\tilde{\Omega}^3)$ in green, all terms of order $\mathcal{O}(\tilde{\Omega}^2)$ in blue and all terms of order $\mathcal{O}(\tilde{\Omega})$ in red. In each of the tables 4.1, 4.2, 4.3 and 4.4, we will call the dominant terms first order terms, followed by second order terms that are of order $\mathcal{O}(\tilde{\Omega})$ smaller than first order terms. Consequently, third order terms are of order $\mathcal{O}(\tilde{\Omega}^2)$ smaller than first order terms and so on.

4.7 Calculation of the Fano factor of long mRNA

The scaling behavior of all terms that are necessary to calculate fluctuations, i.e. second moments, are given in table 4.2. The two dominating terms are marked blue. Thus, in first order it is sufficient to take only these two terms into account. This procedure is in fact the standard procedure used in the literature [4, 28, 29]. Due to nonlinear reaction rates, we get $u = 2$ terms, which mediate the coupling to higher moments. If we want to calculate fluctuations to higher than just first order, we have to take into consideration both red and blue terms in table 4.2. It follows that we have to include first moments (table 4.1) and third moments (table 4.3). First moments are simple to implement, because they couple only to second moments. To calculate third moments, however, we have to consider all dominant terms (blue) in table 4.3. Unfortunately, these terms include also fourth moments, which we would have to calculate via the green terms in table 4.4.

Hence, if we want to consider higher order terms in our calculation of fluctuations, we have to either work out the coupled equations from first moments up to fourth

37

moments, or truncate the coupled equations by introducing a suitable closure-relation. Furthermore, in the threshold region, where the coupling has its largest influence, copy numbers are pretty low, which could bias all order arguments we have used so far. Nonetheless, we would like to test how well the calculated results fit to data obtained by Gillespie simulations. To this end we started from first order calculations and work our way up to higher order calculations.

We began with the standard procedure by considering only the two dominant terms (blue) in table 4.2. Comparing with the result of Gillespie simulations (see S3 Fig) shows an adequate match, which overestimates fluctuations in the vicinity of the threshold. When studying different parameter sets, it can be seen that, although fluctuations are overestimated, the shape of the surface is well matched. We continued and included higher moments, both by an adequate closure relation after the second moment, as well as by actually implementing all terms of table 4.1, all terms of 4.2, all blue terms of table 4.3 and all green terms of 4.4. However, the results for these methods were (in general) worse than those from considering the dominant terms only. Thus, we chose the first order method to calculate the Fano factor.

Bibliography

1. Clark DP, Pazdernik NJ. Molecular biology: Understanding the genetic revolution. ACACL; 2012.

2. Suzuki K, Babitzke P, Kushner SR, Romeo T. Identification of a novel regulatory protein (CsrD) that targets the global regulatory RNAs CsrB and CsrC for degradation by RNase E. Genes & development. 2006;20(18):2605.

3. Kærn M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: from theories to phenotypes. Nature Reviews Genetics. 2005;6(6):451–464.

4. Gardiner CW. Stochastic methods. Springer; 2009.

5. Levine E, Zhang Z, Kuhlman T, Hwa T. Quantitative characteristics of gene regulation by small RNA. PLoS biology. 2007;5(9):e229.

6. Legewie S, Dienst D, Wilde A, Herzel H, Axmann IM. Small RNAs establish delays and temporal thresholds in gene expression. Biophysical journal. 2008;95(7):3232–3238.

7. Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, et al. Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. Science (80-). 2010;329.

8. Gudapaty S, Suzuki K, Wang X, Romeo T, Wang XIN, Babitzke P. Regulatory Interactions of Csr Components : the RNA Binding Protein CsrA Activates csrB Transcription in Escherichia coli Regulatory Interactions of Csr Components : the RNA Binding Protein CsrA Activates csrB Transcription in Escherichia coli. J Bacteriol. 2001;.

9. Legewie S, Dienst D, Wilde A, Herzel H, Axmann IM. Small RNAs establish delays and temporal thresholds in gene expression. Biophys J. 2008;95(7):3232–3238. Available from: http://dx.doi.org/10.1529/biophysj.108.133819.

10. Vakulskas CA, Leng Y, Abe H, Amaki T, Okayama A, Babitzke P, et al. Antagonistic control of the turnover pathway for the global regulatory sRNA CsrB by the CsrA and CsrD proteins. 2016;p. 1–15.

11. Leng Y, Vakulskas CA, Zere TR, Pickering BS, Watnick PI, Babitzke P, et al. Regulation of CsrB/C sRNA decay by EIIA. 2016;(2015):627–639.

12. Eppler T, Postma P, Boos W. Glycerol-3-Phosphate-Induced Catabolite Repression in. 2002;184(11):3044–3052.

13. Adamson DN, Lim HN. Rapid and robust signaling in the CsrA cascade via RNA–protein interactions and feedback regulation. Proc Natl Acad Sci. 2013;110(32):13120–13125. Available from: http://www.pnas.org/content/110/32/13120.abstract.

14. Yang TY, Sung YM, Lei GS, Romeo T, Chak KF. Posttranscriptional repression of the cel gene of the ColE7 operon by the RNA-binding protein CsrA of Escherichia coli. Nucleic acids research. 2010;38(12):3936–3951.

15. Nath K, Koch AL. Protein Degradation in Escherichia coli I. MEASUREMENT OF RAPIDLY AND SLOWLY DECAYING COMPONENTS. Journal of Biological Chemistry. 1970;245(11):2889–2900.

16. Nath K, Koch A, et al. Protein degradation in Escherichia coli. II. Strain differences in the degradation of protein and nucleic acid resulting from starvation. The Journal of biological chemistry. 1971;246(22):6956.

17. Dubey AK, Baker CS, Romeo T, Babitzke P. CsrA - RNA interaction RNA sequence and secondary structure participate in high-affinity CsrA – RNA interaction. Rna. 2005;(11):1579–1587.

18. Mercante J, Edwards AN, Dubey AK, Babitzke P, Romeo T. Molecular Geometry of CsrA (RsmA) Binding to RNA and Its Implications for Regulated Expression. J Mol Biol. 2009;392(2):511–528. Available from: http://dx.doi.org/10.1016/j.jmb.2009.07.034.

19. Yang TY, Sung YM, Lei GS, Romeo T, Chak KF. Posttranscriptional repression of the cel gene of the ColE7 operon by the RNA-binding protein CsrA of Escherichia coli. Nucleic Acids Res. 2010;38(12):3936–3951.

20. Mercante J, Edwards AN, Dubey AK, Babitzke P. Implications for Regulated Expression. 2010;392(2):511–528.

21. E KH. Cell volume increase in Escherichia coli after shifts to richer media. The Journal of Bacteriology. 1990;72(1):94–101.

22. Romeo T, et al. Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB. Molecular microbiology. 1998;29(6):1321.

23. EcoCyc. EcoCyc - A member of the BioCyc database collection;. Available online at http://ecocyc.org.

24. Shimoni Y, Altuvia S, Margalit H, Biham O. Stochastic analysis of the SOS response in Escherichia coli. PloS one. 2009;4(5):e5363.

25. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. The journal of physical chemistry. 1977;81(25):2340–2361.

26. Cascales E, Buchanan SK, Duché D, Kleanthous C, Lloubes R, Postle K, et al. Colicin biology. Microbiology and Molecular Biology Reviews. 2007;71(1):158–229.

27. Bionumbers. RNA polymerase transcription rate of stable RNA - BNID 100060; 2008. Available online at http://bionumbers.hms.harvard.edu/bionumber.aspx?&id=100060&ver=23. Website.

28. Van Kampen N. Stochastic Processes in Physics and Chemistry. North-Holland; 2007.

29. Tsimring L, Volfson D, Hasty J. Stochastically driven genetic circuits. Chaos: An Interdisciplinary Journal of Nonlinear Science. 2006;16(2):026103–026103.

2.6 Publication Draft Reprint

CsrA and its regulators control the time-point of ColicinE2 release in
Escherichia coli

by

A. Götz,[1] M. Lechner,[2], A. Mader,[1] B. von Bronk,[1] E. Frey[2] and M. Opitz,[1]

[1] Faculty of Physics and Center for NanoScience, Ludwig-Maximilians-Universität München, Geschwister-Scholl-Platz 1, D-80539 Munich, Germany
[2]Department of Physics, Arnold Sommerfeld Center for Theoretical Physics and Center for NanoScience, Ludwig-Maximilians-Universität München, Theresienstraße 37, 80333 München, Germany

Title: CsrA and its regulators control the time-point of ColicinE2 release in *Escherichia coli*

Alexandra Götz[1], Matthias Lechner[2], Andreas Mader[1], Benedikt von Bronk[1], Erwin Frey[2] and Madeleine Opitz[1+]

[1] Faculty of Physics and Center for NanoScience, Ludwig-Maximilians-Universität München, Geschwister-Scholl-Platz 1, D-80539 Munich, Germany

[2] Arnold-Sommerfeld-Center for Theoretical Physics and Center for NanoScience, Faculty of Physics, Ludwig-Maximilians-Universität München, Theresienstrasse 37, D-80333 Munich, Germany

Short Title: CsrA controls the time-point of ColicinE2 release

Abstract
The bacterial SOS response is a cellular reaction to DNA damage, that, among other actions, triggers the expression of Colicins - toxic bacteriocins in *Escherichia coli* that are released to kill close relatives competing for resources. However, it is largely unkown, how the complex network regulating toxin expression controls the time-point of toxin release to prevent premature release of inefficient protein concentrations.
Here, we study how different regulatory mechanisms affect production and release of the bacteriocin ColicinE2 in *Escherichia coli*. Combining experimental and theoretical approaches, we demonstrate that the global carbon storage regulator CsrA controls the duration of the delay between toxin production and release and emphasize the importance of CsrA sequestering elements for the timing of ColicinE2 release. Here, mRNA produced from the ColicinE2 operon upon SOS response and binding to CsrA has the largest effect. Furthermore, we show that CsrA additionally binds to and is sequestrated by ssDNA originating from rolling-circle replication of the toxin-producing plasmid. Our theoretical analysis emphasizes that ssDNA is essential in ColicinE2-producing wild-type strain to enable toxin release, by reducing the amount of free CsrA molecules in the bacterial cell.
Taken together, our findings show that CsrA times ColicinE2 release and reveal a dual function for CsrA as a ssDNA and mRNA-binding protein, introducing ssDNA as a potential post-transcriptional gene regulatory element.

Author Summary

Chemical warfare by bacteriocin production is one means of pathogenic bacteria to dominate their habitat, thus increasing their potential to infect a human host. The timing of toxin release is highly relevant, as it ensures the competitive success of its producer. Here, we show that the time-point of ColicinE2 release is tightly regulated by the global carbon storage regulator CsrA. Furthermore, we demonstrate that CsrA can bind to ssDNA, an intermediate of autonomous pColE2-P9-plasmid replication. Hence, we suggest ssDNA as a new CsrA sequestering element, that in addition to known regulators such as the sRNAs CsrB and CsrC or mRNAs carrying a CsrA binding site, can affect the amount of free CsrA in the bacterial cell.

Main text

Introduction

Many pathogenic bacteria outcompete close relatives by the secretion of toxic bacteriocins (1-3), thereby increasing their own ability to dominate bacterial populations and thus increase their potential to infect a human host (3, 4). The well studied *Escherichia coli* ColicinE2 system (5-8) represents a paradigmatic model for the study of regulatory mechanisms relevant for toxin production. Here, transcriptional and post-transcriptional regulation mechanisms control ColicinE2 expression (9). However, it is not known how this regulatory network (**Fig 1a)** times ColicinE2 production and release, to ensure the production of effective toxin concentrations (*cea* gene expression) and to prevent premature toxin release (*cel* gene expression) in the wild-type strain C_{WT}.

Bacteriocins, including ColicinE2, are plasmid encoded and heterogeneously expressed (10-13) from operons under the control of an SOS promoter (9, 14) in response to external stresses. The ColicinE2 operon consists of three genes: *cea* (the colicin activity gene), *cei* (the immunity gene) and *cel* (the lysis gene) (**Fig 1a**). Upon induction of the SOS response, RecA induces autocleavage of LexA dimers, which permits the production of two mRNAs: the 'short' transcript including *cea* and *cei,* and the 'long' one comprising all three genes (9). Co-expression of the genes *cea* and *cei* is necessary, since the immunity protein ensures that the colicin remains inactive for as long as the colicin-immunity protein complex is present within the cell. This complex (15) is secreted upon *cel* gene expression (16), which leads to the death of the bacterial cell, secreting the colicin. Translation of the *cel* gene is regulated post-transcriptionally by the mRNA binding protein CsrA (17, 18). The abundance of CsrA is further regulated by the two CsrA binding sRNAs CsrB and CsrC (19).

2

Results

To study how the ColicinE2 regulatory network controls the timing of toxin expression, we investigated the dynamics of toxin production and release using a combined experimental and theoretical approach. In order to clearly distinguish toxin-expressing cells from cells that did not produce and release the toxin or only at basal levels, we introduced an additional multi-copy reporter plasmid (pMO3, **Methods**) into the wild-type strain C_{WT}, in which the genes *cea* and *cel* are replaced by sequences encoding Yellow and Cerulean fluorescence proteins (YFP and CFP), respectively (**Methods**), resulting in the reporter strain C_{REP1}. This double reporter plasmid carries all genetic sequences relevant for transcriptional and post-transcriptional regulation found in the original ColicinE2 operon (**Fig 1a,b**). This ensures that the expression of the YFP and CFP directly reflects the normal expression patterns of the *cea* (toxin production) and *cel* (toxin release) genes (13). Using single-cell time-lapse microscopy (**Methods, Fig S1**) (13), we found that C_{REP1} expresses the *cea* and *cel* genes nearly simultaneously (**Fig 1c-d**), with an insignificant delay of 4 ± 2 min.

This observation, however, is biologically implausible, as it would prevent the accumulation of effective toxin concentrations. In addition, theoretical investigations (20) of the ColicinE2 network predicted a significant delay between *cea* and *cel* gene expression. These conflicting results thus suggested that the introduction of the multi-copy reporter plasmid somehow affected the *cea-cel* delay, potentially by interfering with regulatory mechanisms derived from or otherwise linked to the ColicinE2-P9 plasmid of the wild-type strain C_{WT}, controlling the timing of *cea* and *cel* expression.

To address this question, we created a second reporter strain, S_{REP1}, which only carries the double reporter plasmid, but lacks the original pColE2-P9 plasmid (**Table S2, Fig S1**). In this strain, we find a significant delay of 75 ± 6 min between *cea* and *cel* gene expression, with a slight decrease in delay times at higher stress levels, imposed by increasing the level of the SOS response-inducing agent Mitomycin C (MitC) in the medium (**Fig 1c,d**). This indicated that the original pColE2-P9 plasmid in the C_{REP1} strain, which the S_{REP1} strain lacks, contains additional regulatory elements that are responsible for the reduction of the *cea-cel* delay in the C_{REP1} strain.

To disentangle the roles of the various regulatory elements controlling the *cea-cel* delay, we first analysed the roles of known transcriptional and post-transcriptional regulators of the ColicinE2 network. In the second step, we studied further regulatory elements present on the original ColicinE2-P9 plasmid. Finally, we integrated the observations made in these experiments into a theoretical model, to uncover the major regulatory elements controlling the *cea-cel* delay in C_{WT}.

In the first step, we asked whether and how the *cea-cel* delay observed in the S_{REP1} strain is affected by individual regulatory modules (transcriptional and post-transcriptional), starting with the role of the transcriptional repressor LexA. To assess the impact of transcriptional regulation by LexA on the duration of the delay between *cea* and *cel* gene expression, we created two S_{REP1} strain mutants in which the LexA binding sites on the reporter plasmid had been altered, such that LexA binding was expected to increase (LexA1) or decrease (LexA2) (**Methods**). As expected, the maximal fluorescence intensity (FI_{max}) and the fraction of ColicinE2 expressing cells (% ON) were decreased in the LexA1 mutant (**Fig S2**), while the delay time did not differ significantly from that of the S_{REP1} strain, indicating that stronger LexA binding does not affect the delay between *cea* and *cel* gene expression.

In contrast to our expectations, however, the values for FI_{max} and % ON were also lower in the LexA2 mutant. Moreover, the *cea-cel* delay was also markedly affected, falling to 36 ± 6 min in comparison to 75 ± 6 min for the S_{REP1} strain (**Fig S2**). This decrease is accounted for by the shift in the t_{ON} distribution for *cea* expression to later time-points (the time-point t_{ON} marks the

onset of the 'ON' state, **Methods**, **Fig S2**), indicating the absence of a post-transcriptional regulation effect. Kamensek *et al.(21)* report that an additional protein, AsnC, controls the temporal induction of the ColicinE2 operon in concert with LexA. As the AsnC protein also binds within the LexA binding sites (21, 22), changes in the LexA binding site could also alter AsnC binding, thus affecting the initiation of ColicinE2 transcription. Hence, we conclude that the delay between *cea* and *cel* gene expression can be shortened by altering the timing of the onset of *cea* expression (t_{ONcea}), but that it is difficult to determine which transcriptional protein (LexA or AsnC) is causing the shift in t_{ONcea}. However, the observed reduction of the *cea-cel* delay in the LexA2 mutant was insufficient to explain the simultaneous expression of *cea* and *cel* genes in the C_{REP1} strain.

A second regulatory element essential for ColicinE2 expression is the global carbon storage regulator CsrA, which inhibits *cel* expression by binding to the Shine-Dalgarno (S-D) sequence present in the long mRNA (**Fig 1a**). To investigate the influence of the post-transcriptional regulator CsrA on the ColicinE2 expression dynamics, we created two S_{REP1} strain mutants with altered CsrA binding sites on the pMO3 reporter plasmid (**Methods**), such that CsrA binding to the S-D sequence was either increased (CsrA1) or reduced (CsrA2) (**Fig 2**). As expected, *cea* expression was unaffected in both mutants (**Fig 2a-c**). In the CsrA1 mutant, the fraction of cells expressing the *cel* gene (% ON), and the maximal fluorescence intensity (FI_{max}) of the cells in the ON state, was reduced. In contrast, FI_{max} and % ON were increased in the CsrA2 mutant (**Fig 2a,b**). While the CsrA1 mutant showed an increased mean *cea-cel* delay with 94 ± 6 min, the CsrA2 mutant displayed a markedly shorter delay of 12 ± 2 min in comparison to the S_{REP1} strain (**Fig 2e**). This shortening of the *cea-cel* delay is due to the earlier onset of *cel* gene expression (t_{ONcel}), while the timing of *cea* gene expression (t_{ONcea}) is nearly unaffected (**Fig 2c**).

To elucidate the role of CsrA theoretically, we extended our previous mathematical model (20). This generalized model emulates the complex dynamical behaviour of the regulatory components and accounts for all regulatory interactions discussed here, including the different plasmid compositions and abundances that characterize the three strains C_{REP1}, S_{REP1} and C_{WT} (**Fig S3**, **Table S5**, please see **SI** for details of the theoretical model). As the model enables us to study various parameter values of these interactions, it gives us a controlled way to investigate the influence of the different regulatory components and mechanisms on the production and release of ColicinE2. To show that our model is indeed valid, we employed parameters motivated by experimental studies (see **SI** for details) to reproduce the delay distribution of the S_{REP1} strain (**Figs. 2e, 2f**). In agreement with the above experiment, we also find that alteration of k_M, which quantifies binding of CsrA to long mRNA resulted either in an increase ($k_M = 0,0125$) or a decrease ($k_M = 0.0018$) of the mean *cea-cel* delay in the S_{REP1} strain (**Fig 2f**). Hence, our combined experimental and theoretical analysis demonstrates that CsrA mediates the delay between toxin production and release.

While CsrA directly affects the *cea-cel* delay by deferring *cel* gene expression, regulatory elements that sequester CsrA can indirectly affect the duration of the delay by controlling the abundance of the free CsrA protein. Two known CsrA-sequestering elements are the sRNAs CsrB and CsrC (23-25). Hence, deletion of these sRNAs should lead to a strong increase in CsrA abundance and consequently extend the *cea-cel* delay. To investigate the role of these sRNAs in ColicinE2 expression, we first created a knock-out S_{REP1} strain mutant (**Methods**) lacking the sRNA CsrB, which includes 18 imperfect repeat sequences that serve as CsrA binding sites (in comparison to only 9 in CsrC)(24). We find that neither *cea* nor *cel* expression is significantly altered in the mutant (**Fig S2**). This finding was confirmed in long-term experiments, where a decrease in FI_{max} was observed only in the very late stationary phase (**Methods, Fig S4**) - which can be explained by a compensatory effect of the second sRNA CsrC(24). Accordingly, the delay was only slightly decreased (to 68 ± 6 min) in the CsrB mutant (**Fig S2**). In contrast to our expectations, a double sRNA knock-out in the S_{REP1}

strain (deletion of both CsrB and CsrC) showed increased FI_{max} (**Figs S2 and S4**) and slightly altered % ON values for both *cea* and *cel* gene expression (**Fig S2**). Interestingly, also the delay between *cea* and *cel* gene expression was significantly reduced (to 36 ± 5 min) relative to the S_{REP1} strain (**Fig S2**). In addition, the onset of expression, t_{ON}, was shifted to earlier time-points for both *cea* and *cel* (**Fig S2),** indicating that transcription of the entire ColicinE2 operon was prematurely induced, due to the increased availability of CsrA. A connection between CsrA abundance and the LexA-RecA network was previously described for ColicinE7 expression(26). The reduction of the *cea-cel* delay was primarily due to the pronounced shift of t_{ONcel} to earlier time points in comparison to the small shift in t_{ONcea}. Hence, our data imply that while deletion of a single sRNA does not affect ColicinE2 expression significantly (**Figs S2 and S4**), deletion of both sRNAs (CsrB and CsrC) leads to premature production and release of ColicinE2 (**Fig S2**). This result points to the intervention of yet unknown regulatory mechanisms.Up to now, we have considered the impact on ColicinE2 expression of regulatory factors that are present in both S_{REP1} and C_{REP1} strains. However, it was still unclear which regulatory element deriving from the original pColicinE2-P9 plasmid (**Fig S5**) reduces the *cea-cel* delay in the C_{REP1} strain. The observed differences in the *cea-cel* delay in the C_{REP1} versus S_{REP1} strain could be due to the additional 20 pColE2-P9 plasmids in the C_{REP1} strain, increasing the plasmid copy number in this strain to 75 compared to 55 in the S_{REP1} strain (**SI**). As the plasmid copy number correlates with the amount of long mRNA in the presence of an SOS response, consequently, a higher amount of long mRNA able to sequester CsrA is present in C_{REP1}. To estimate the effect of the plasmid copy number/amount of long mRNA on the *cea-cel* delay, we accounted for the exact plasmid composition for each strain in the theoretical modelling (**SI**). As described above, for the S_{REP1} strain the theoretical analysis accurately retrieved the experimentally observed *cea-cel* delay of 67 min (**Fig. 2e,f and Fig S6**). For C_{REP1} the theoretical model predicted a delay of about 24 min (**Fig S6**) that was significantly longer than the experimentally observed delay of 4 min (**Fig. 1d**). To further study the impact of the plasmid copy number on the *cea-cel* delay, we changed the origin of replication of the reporter plasmid in the way, that now only ~ 13 reporter plasmids per cell are produced, resulting in strains C_{REP2} and S_{REP2} (**Methods, Table S2**). A reduction of the plasmid copy number should extend the *cea-cel* delay, as now less long mRNA is produced and consequently more free CsrA molecules are able to bind at the S-D sequence of the *cel* gene. Indeed, the X_{REP2} strains with a decreased amount in total plasmid copy number show increased delay times compared to their corresponding X_{REP1} strain. For C_{REP2} with ~ 33 plasmid copies in total (**Methods**) we obtain a *cea-cel* delay of 25 ± 4 min. For S_{REP2} with ~13 plasmid copies we find a delay of > 101 min, as here 67% of the cells do not express the *cel* gene and consequently do not lyse during the time frame of the experiment. Hence, the higher amount of long mRNA due to an increased plasmid copy number explains a strong reduction in the *cea-cel* delay. However, it cannot explain the discrepancy in delay times between the C and S strains, with the C strains having delay times much shorter as expected with regard to their plasmid copy number.

Consequently, we investigated genetic elements on or deriving from the pColE2-P9 plasmid that might affect ColicinE2 expression. We sequenced the entire pColE2-P9 plasmid (**Methods, Fig S5a,** Genbank accession number KY348421) and performed a homology comparison of genes present on this plasmid with part of the closely related plasmid pColE3-CA38 (**Table S1**). As in pColE3-CA38, most genes on pColE2-P9 are involved in autonomous plasmid replication (**SI**), but we could not find a link between these genes and regulatory elements affecting ColicinE2 expression. At this point, we recalled that rolling-circle replication (**SI**) can lead to the accumulation of a ssDNA intermediate, as was shown for pColE3-CA38 (27). This ssDNA could interact with other regulatory elements affecting ColicinE2 expression, e.g. sequester the global regulatory protein CsrA, thereby further reducing the *cea-cel* delay in the C_{REP} strains. To address this hypothesis, we first confirmed

the accumulation of ssDNA for cells carrying the pColE2-P9 plasmid (**Fig 3**) in the absence and presence of the SOS inducing agent MitC (**Fig S5b**). Secondly, we performed gel shift analysis to investigate the binding of CsrA to both long mRNA and ssDNA (**Methods, SI**). We find that CsrA binds to a RNA oligo carrying the original nucleotide sequence of the long mRNA with a K_d of 22 ± 13 nM, which is in good accordance with values described in literature (17, 28). Furthermore, we found that the mRNA binding protein CsrA is able to bind to ssDNA with a K_d of 991 ± 164 nM (**Methods**, **Fig S5c,d**). The binding strength of CsrA to ssDNA is therefore by a factor of 45 lower as the binding strength of CsrA to sRNA (**Fig. 2d**). This finding, that CsrA can bind both sRNA as well as ssDNA was in accordance with previous studies revealing that CsrA possesses a KH domain (29), a domain that is known to enable proteins to bind to mRNA as well as ssDNA (30, 31). To investigate if CsrA binds ssDNA at the known CsrA binding sites (**Table S4**) for CsrA-RNA interaction (17, 32), we studied the binding of CsrA to ssDNA with altered CsrA binding sites. We introduced the same changes in the CsrA binding site as done before for the RNA (**Fig. 2d, Table S4**). As seen for CsrA binding to RNA, we find that CsrA binds stronger to the sequence that should allow for stronger CsrA binding (**Methods**, CsrA1 sequence), and that CsrA binds weaker to the sequence that should weaken CsrA binding due to impaired formation of the second hairpin harbouring the second CsrA binding site (**Fig S5d, Methods**, CsrA2 sequence). This indicates that CsrA uses the same binding motives on the ssDNA as on the RNA - namely the GGA motive, with the neighbouring bases enabling the establishment of a hairpin structure exposing the GGA motive to allow for accurate CsrA binding(32). However, binding of CsrA to ssDNA is by a factor of 45 less efficient than binding of CsrA to RNA. Still, ssDNA can serve as an additional CsrA sequestering element, as ssDNA is produced continuously during the bacterial cell cycle and accumulates in the cell to very high numbers (**Fig. 3**, **Fig S5b**). In contrast, long mRNA sequestering CsrA is produced only upon induction of the SOS response.

To support our hypothesis that ssDNA as an additional CsrA sequestering element could further reduce the *cea-cel* delay in the C_{REP} strains but also in the C_{WT} strain, we incorporated this additional regulatory element into the theoretical model (**SI, Figs S6 and S7**). We find that the presence of ssDNA additionally reduces CsrA abundance in the C_{REP} and C_{WT} strains. Furthermore, the presence of ssDNA in combination with the high amount of long mRNA totally suppresses the *cea-cel* delay due to the increased plasmid copy number in the C_{REP1} strain by CsrA sequestration (**Fig 4**). For the natural C_{WT} strain carrying only the 20 pColE2-P9 plasmids, our model predicts that the *cea-cel* delay lasts approximately one hour. Furthermore, the *cea-cel* delay is broadly distributed in the wild-type strain C_{WT}. Importantly, *cel* gene expression and consequently toxin release in the C_{WT} strain only occurs within the time-frame of our experimental studies if ssDNA is present (**Fig 4, Fig S8**). Hence, the delay time of a particular strain that is determined by the abundance of free CsrA is controlled by three CrsA sequestering components (**Fig. 5a**). First, the action of the sRNAs that are present in all strains studied in this work. Second, the amount of long mRNA that depends on the type and number of plasmid present in the particular strain (**Fig. 5b, SI, Table S5**) and third, the additional accumulation of ssDNA in strains carrying the pColE2-P9 plasmid (C strains, **Fig. 5b**). Notably, the effect of these three CsrA sequestering elements differs due to their occurrence (**SI**); e.g. while long mRNA is only produced upon induction via the SOS response, ssDNA is produced independently due to autonomous rolling circle replication in the presence and absence of an SOS response (**Fig S5b**).

Discussion

In this study, we investigated regulatory factors controlling ColicinE2 production and release in response to an SOS signal and demonstrate that the global carbon storage regulator CsrA

6

controls the time-point of ColicinE2 release in *Escherichia coli*. The mRNA binding protein CsrA is highly abundant in the *E. coli* cell, with 11.000-33.000 CsrA molecules in total (bound and unbound) (18). However, Taniguchi *et al*., report that only a small fraction of 474 CrsA molecules per cell are freely available (33). This indicates that CsrA sequestering elements can strongly affect the amount of free CsrA in the bacterial cell.

Two well studied CsrA sequestering components are the sRNAs CsrB and CsrC (18, 24, 25). It was shown that CsrB alone can bind up to 32% of CsrA present in the bacterial cell (18). In this study, we were able to show that besides these sRNAs, also mRNAs that carry a CsrA binding site (34) can strongly reduce the abundance of free CsrA. Furthermore, we verified that CsrA is able to bind to ssDNA originating from autonomously replicating plasmids. Our data indicate that CsrA is thereby binding to the GGA motive exposed in the second hairpin loop present in both ssDNA and long mRNA deriving from the pColE2-P9. This demonstrates the dual role of CsrA as an mRNA and ssDNA binding protein. In addition, our study shows that in *E. coli* cells carrying autonomously replicating plasmids with a CsrA binding site, ssDNA deriving from these plasmids can serve as an additional CsrA sequestering element. We speculate that ssDNA accumulating in bacterial cells could play an important regulatory role in other protein-expressing networks that rely on the expression of proteins from autonomously replicating plasmids, as is the case for many bacteriocin-producing networks.

With regard to the ColicinE2 system, our combined experimental and theoretical efforts allowed us to disentangle the different regulatory mechanisms affecting the delay between toxin production and release. We revealed that the interplay between CsrA and ssDNA, sRNAs and long mRNA, times toxin release upon induction of the SOS response (**Fig 5**). In particular, our theoretical investigations emphasized that the presence of ssDNA can enable the toxin producer to release the toxin within few hours once an SOS response has been triggered. From an evolutionary perspective, a short delay might be important for the toxin producing colony to respond quickly to changing environmental conditions and to increase its competitive success.

Material and Methods

Creation of bacterial strains used in this study

All strains used in this study are listed in **SI (Table S2)**. The strain C_{WT} represents the original wild-type strain, which carries the toxin-producing plasmid pColE2-P9. The C_{REP1} strain and the S_{REP1} strain (EMO3-C and EMO3-S, respectively) were constructed as described in Mader *et al (13)*. Both strains carry the double reporter plasmid pMO3 (13). This plasmid, pMO3, harbours the entire ColicinE2 operon, in which the genes *cea* and *cel* have been replaced by genes coding for the fluorescence proteins (FP) mVenus (YFP) and mCerulean (CFP), respectively (**Fig 1**). Hence, this plasmid retains all regulatory sequences relevant for the binding of LexA to the SOS box of the ColicinE2 operon, and of CsrA to the Shine-Dalgarno sequence on the resulting long mRNA. To investigate the role of specific regulatory elements on the duration of the delay between *cea* and *cel* expression, all mutant strains used in this study are derived from EMO3-S, the S_{REP1} strain. Construction of these mutant strains is described in the following.

To investigate the impact of the transcriptional repressor LexA on Colicin E2 expression, we altered the LexA binding site on the pMO3 reporter plasmid using site-directed mutagenesis with the Quick ChangeII kit (Agilent Technologies). According to Lewis *et al. (30)*, the strength of LexA binding to the two overlapping SOS boxes (LexA binding sites) can be estimated from the HI index. Based on these estimations, we created two LexA mutants (**SI**):

LexA1 was created using primer pair P1/P2 (**SI**). The base exchange AT-to-TA on pMO3 (**SI**) is expected to lead to tighter LexA binding, with a HI factor of 8.6 for the first SOS box, which is the more important for LexA binding (35). The resulting plasmid was named pMO4 (**SI**). LexA2 was created using the primer pair P3/P4 (**SI**). The base exchange of CTG-to-CCC in the first SOS box on pMO3 (**SI**) is expected to weaken LexA binding, with a HI index of 21.13 for the mutant sequence. The resulting plasmid was named pMO5 (**SI**).

To analyze the post-transcriptional impact of the mRNA-binding protein CsrA on ColicinE2 expression, we altered the CsrA binding site on the pMO3 reporter using site-directed mutagenesis as described above. In the first mutant strain (CsrA1), we introduced a mutation (GTC to TGT) in the second CsrA binding site (**SI**) within the Shine-Dalgarno sequence of the *cel/cfp* gene on pMO3 using the primers P5 and P6 (**SI**), generating the plasmid pMO6 (**SI**). This mutation optimizes the CsrA binding site (32) and therefore increases CsrA binding to pMO6 relative to pMO3 (**Fig 2a**). In the second mutant (CsrA2), CsrA binding was decreased (32, 36) (**Fig 2a**) by using the primer pair P7/P8 (**SI**) to alter AC to TT in the second CsrA binding site in the Shine-Dalgarno sequence of the *cel/cfp* gene on pMO3, thus inhibiting formation of the second mRNA hairpin. The resulting plasmid was named pMO7 (**SI**).

To understand the roles of the sRNAs CsrB and CsrC in CsrA sequestration and consequently in ColicinE2 expression, single and double knock-out mutants for these sRNAs were created using the Quick&Easy *E.coli* Gene Deletion Kit Nr.6 (Gene Bridges, Heidelberg, Germany). The gene coding for the sRNA CsrB was replaced in strain BZB 1011 with a kanamycin resistance cassette using the primer pair P9/P10 (**SI**). The resulting strain was named BZB 1011::CsrB, and the reporter plasmid pMO3 was transformed into this strain to produce EMO3::CsrB (CsrB) (**SI**). The single knock-out mutant of CsrC was created in a similar manner (primer P11/P12, **SI**), and was also used for the double sRNA knock-out. Here, the genomic region coding for the sRNA CsrC was replaced by a kanamycin resistance cassette. In next step, the primers P9 and P10 (**SI**) were used to replace the CsrB gene with a chloramphenicol resistance cassette. This strain was named BZB 1011::CsrB/C, and it too was transformed with the plasmid pMO3 to generate EMO3::CsrB/C (CsrB/C) (**SI**).

Creation of C_{REP2} and S_{REP2}

The reporter plasmid of the X_{REP2} strains was created by a PCR of the plasmid pMO3 with the primer P25 and P26, which delete the ORI of the pMO3 plasmid. The new ORI p15A was replicated via PCR using primer P27 and P28 from the Vector pZA11MCS (EXPRESSYS). After gel purification of the vector using the Freeze 'N Squeeze™ DNA Gel Extraction Spin Columns (bio-rad), both, the vector and the ORI p15A were cut with the enzymes SalI-HF and SphI-HF (NEB) and ligated in a 1:5 ratio of vector:insert using an ElectroLigase® (NEB). The resulting plasmid pMO8 was transformed into an *E.coli* strain (XL1) for replication using an MicroPilaser Electroporation Apparatus (bio-rad) and selected on ampicillin plates. After purification of the pMO8 plasmid using a QIAprep Spin Miniprep Kit (Qiagen), the reporter strains S_{REP2} and C_{REP2} were created via transformation of pMO8 into BZB 1011 and $C_{WT,}$ respectively, with the bio-rad electroporator.

To verify the copy number of the pMO3 and pMO8 plasmids in the S_{REP1} and S_{REP2} strains, respectively, the bacteria were grown in M63 medium with antibiotic over night at 37°C and 300rpm. The cultures were then diluted to OD600 in an equal volume and the plasmids were purified using the QIAprep Spin Miniprep Kit (Qiagen). The concentration of the DNA was measured using the NANODROP 1000 instrument (ThermoScientific). This lead to a copy number of 55 ± 11 and 13 ± 4 plasmids per cell for the S_{REP1} and S_{REP2} strain, respectively.

Fluorescence microscopy
Bacteria were grown overnight at 37°C in M63 minimal medium supplemented with 0.5% glycerol as a carbon source, and with 100 µg/ml ampicillin (Carl Roth, Germany) if required. Overnight cultures were diluted to an OD_{600} of 0.05 and grown to an OD_{600} of 0.2, which represents the beginning of the exponential growth phase. Aliquots (50 µl) of these cultures were allowed to attach to poly-L-lysine (BIOCHROM, Berlin)-coated Ibidi µ-slides VI$^{0.4}$ (Ibidi GmbH, Munich) for 7.5 min and rinsed to remove unattached bacteria[13]. For time-lapse experiments, slides were then transferred to an inverse microscope, Axiovert 200M (Carl Zeiss, Germany) equipped with an Andor camera and a Zeiss EC Plan-Neofluar 100x/1.3 oil-immersion objective. A filter set with a beam splitter BS520, an excitation bandpass HC500/24 and an emission bandpass HC 542/27 was used for YFP detection. The HC filter set for CFP detection consisted of an emission filter 483/32, a beam splitter BS458 and an excitation filter 438/24. To minimize fluorescence variations deriving from day-to-day fluctuations of the excitation source, the stability of the absolute fluorescence values was verified daily using a microscope image intensity calibration kit (Invitrogen, FokalCheck™ fluorescence microscope test slide #3) and data sets were corrected accordingly. Micromanager, an open-source program (version 1.3), was used for image acquisition [37]. After the first image, the chamber was flushed with medium containing the appropriate concentration of mitomycin C (MitC, Carl Roth, Germany). Subsequently, an image was taken every 15 min over a period of 300 min. Images were analyzed using the Cell Evaluator plug-in [38] for ImageJ. Only live cells lying within the bright-field image were considered. General data analysis was performed using IgorPRO 6.22, Matlab (R2013b) and Adobe CS5 Software. FI_{max} represents the average maximal fluorescence intensity of single cells expressing the Colicin E2 operon. To quantify the numbers of cells expressing the FPs YFP and CFP (*cea* and *cel* gene expression, respectively) a threshold level was set to distinguish expressors from non-expressors, as described earlier [13]. The resulting fraction of cells expressing either *cea* or *cel* is given as the cumulative fraction. The time-point t_{ON} which marks the onset of the 'ON' state is defined as the time at which fluorescence exceeds this switching threshold. The delay time between *cea* and *cel* gene expression was then calculated as the mean of the t_{ONcel} - t_{ONcea} values for individual cells expressing both *cea* and *cel*. Upon induction with MitC, the parameters FI_{max}, % ON (**Fig S1**) and delay time (**Fig 1c**) show only little variation with MitC concentration (0.1, 0.25 and 0.4 µg/ml). Consequently, data presented in **Fig 2, and Fig S2** represent the average values of these three MitC concentrations to allow for better comparability and to improve statistics.

Long-term analysis of fluorescence development
To investigate the role of sRNA knock-outs on *cea* and *cel* gene expression on a longer time-scale, experiments were performed with the Fluostar Optima Plate Reader (BMG Labtech). A 500-µl aliquot of a starter culture at OD_{600} 0.2 was induced with the appropriate MitC concentration as described above. To prevent cultures from drying out, the plate was sealed with an O_2 permeable foil. Antibiotics were added as required, ampicillin at 100 µg/ml (Carl Roth, Germany), kanamycin at 50 µg/ml (Carl Roth, Germany) and chloramphenicol at 5 µg/ml (Carl Roth, Germany). Bacterial growth (absorbance) and YFP and CFP fluorescence development (representing *cea* and *cel* expression, respectively) was followed over a period of 16 h at 37°C, with shaking at 300 rpm.

ssDNA accumulation and purification

Bacterial strains were grown overnight at 37°C in shaken cultures. The overnight cultures were induced for approximately 75 min with 0.25 µg/ml MitC. Plasmid/ssDNA extraction was performed using the Miniprep Kit (Qiagen, Germany), and 1 µg each of C_{WT} and S_{REP1} strain extracts was cleaved with PvuI (New England Biolabs (NEB), Germany). Then 300 ng of both cut and uncut C_{WT} and S_{REP1} strain extracts were applied to an 1% agarose gel, prestained with EtBr. To validate the presence of ssDNA, single-stranded circular Phi174 and M13mp18 viral DNAs (NEB, Germany) were also applied to the gel (**Fig 3**).

Sequencing and homology analysis
Sequencing of the 6757-bp pColE2-P9 was performed by MWG Eurofins Genomics (Germany) using an ABI 3730XL sequencing instrument and the sequencing primers (P13-P24) listed in the **SI**. The sequences of specific segments such as the 2640-bp ColicinE2 operon (Genbank M29885) and the Rep protein region (Genbank D30054) were verified. In all, 1754 bp were sequenced de novo, and the resulting plasmid map of the completely sequenced pColicinE2-P9 is given in **Fig S5a**. The pColicinE2-P9 sequence has been deposited in GenBank (accession number KY348421). We also used the NCBI online tool BLAST to compare the sequences of genes in the pColE2-P9 plasmid with their homologues in pColE3-CA38, for which ssDNA accumulation was shown previously (27). The sequence homologies are given in **Table S1**.

Gel shift analysis
To determine the affinities of CsrA for three different RNA constructs representing the CsrA binding sites present in pMO3, pMO6 und pMO7 (**SI, Table S4**) we performed gel shift analysis. The N-terminal 6xHis-tagged CsrA protein used for gel shift measurements was obtained from Biozol (Germany). The folding structures of the oligos, were analyzed using the Mfold web server (39), and showed the expected double-hairpin structure that facilitates CsrA binding in the RNAs derived from pMO3, pMO6 and the ssDNA. The RNA of pMO7 however lacks this structure. The RNA was folded for 3.5 min at 85°C in 10 mM Tris-HCl, 1 mM EDTA, 200 mM KCl, 20 mM $MgCl_2$ buffer. ssDNA folding was performed under the same conditions but with 90°C.
For the gel shift analysis of RNA binding to CsrA (**Fig S5c**), serial 2-fold dilutions were made from a 6600 nM stock solution of CsrA (down to approximately 0.8nM CsrA).
To verify binding of CsrA to ssDNA we performed the same gel shift analysis as done for the sRNA oligos, with sequences of 89-bp ssDNA oligos equivalent to pMO3, pMO6 and pMO7 (**SI, Table S4**). The samples for ssDNA gel shift analysis were prepared by doing a serial dilution starting with 33.3µM CsrA stock solution (to a minimal CsrA concentration of approx. 4nM CsrA).
The binding reaction was performed in a buffer containing 15 mM Tris-HCl, 0.5 mM EDTA, 250 mM NaCl, 50mM KCl, 5 mM $MgCl_2$, 3.25 ng/µl yeast RNA, 4U RNase inhibitor (Ambion) and 10% glycerol buffer and incubated at 37°C for 30min.
Gel shift measurements were performed at room temperature with precast 4–20% Mini-PROTEAN® TGX™ Precast Protein Gels (bio-rad) in Tris/Glycine Buffer (bio-rad) and run at 85V for 1h. Pictures of the gel shift were taken with the gel chamber ChemiDoc™ MP Imaging System (bio-rad) using filters for Cy5 labels.
Using the ImageJ software the intensity decrease in the unbound band of RNA and ssDNA for increasing CsrA concentrations was analysed for all gel shifts and plotted with IgorPro 7.04. The curves were then fitted using the following equation that was adapted from (40).

$$()\left[\frac{\sqrt{()^{\square}}}{\square}\right]$$

Here FI is the measured fluorescence intensity of the RNA/ssDNA unbound band, m is the maximum FI, b the basal FI, R the used RNA/ssDNA concentration and P the CsrA concentration. The Kd for CsrA binding to ssDNA and RNA was determined on three separate days.

Theoretical modelling
For the theoretical analysis performed in this study we refer the reader to the **SI.**

Supplementary Information (SI): SI appendix (description of theoretical investigations and further experimental results and methods, eight supplementary figures and five supplementary tables) have been provided to support this article.

Author contributions M.O. designed the experimental research. E.F. designed the theoretical analysis. A. G., A. M., B. v. B. performed the experiments. M. L. and E.F. performed the theoretical analysis and simulations. All authors analyzed the data. A.G., A.M., M.L., E.F. and M.O. wrote the paper.

Author information The authors declare no competing financial interest. Correspondence and requests for materials should be addressed to M.O. (Opitz@physik.uni-muenchen.de).

References
1. Bakkal S, Robinson SM, Ordonez CL, Waltz DA, Riley MA. Role of bacteriocins in mediating interactions of bacterial isolates taken from cystic fibrosis patients. Microbiology. 2010;156(Pt 7):2058-67. Epub 2010/04/10.
2. Nedialkova LP, Denzler R, Koeppel MB, Diehl M, Ring D, Wille T, et al. Inflammation fuels colicin Ib-dependent competition of Salmonella serovar Typhimurium and E. coli in enterobacterial blooms. PLoS Pathog. 2014;10(1):e1003844. Epub 2014/01/07.
3. Kommineni S, Bretl DJ, Lam V, Chakraborty R, Hayward M, Simpson P, et al. Bacteriocin production augments niche competition by enterococci in the mammalian gastrointestinal tract. Nature. 2015;526(7575):719-22. Epub 2015/10/20.
4. Bashan A, Gibson TE, Friedman J, Carey VJ, Weiss ST, Hohmann EL, et al. Universality of human microbial dynamics. Nature. 2016;534(7606):259-62. Epub 2016/06/10.
5. Weber MF, Poxleitner G, Hebisch E, Frey E, Opitz M. Chemical warfare and survival strategies in bacterial range expansions. J R Soc Interface. 2014;11(96):20140172. Epub 2014/05/09.

6.   Reichenbach T, Mobilia M, Frey E. Mobility promotes and jeopardizes biodiversity in rock-paper-scissors games. Nature. 2007;448(7157):1046-9. Epub 2007/08/31.

7.   Kerr B, Riley MA, Feldman MW, Bohannan BJ. Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. Nature. 2002;418(6894):171-4. Epub 2002/07/12.

8.   Kirkup BC, Riley MA. Antibiotic-mediated antagonism leads to a bacterial game of rock-paper-scissors in vivo. Nature. 2004;428(6981):412-4. Epub 2004/03/26.

9.   Cascales E, Buchanan SK, Duche D, Kleanthous C, Lloubes R, Postle K, et al. Colicin biology. Microbiol Mol Biol Rev. 2007;71(1):158-229. Epub 2007/03/10.

10.   Kamensek S, Podlesek Z, Gillor O, Zgur-Bertok D. Genes regulated by the Escherichia coli SOS repressor LexA exhibit heterogeneous expression. BMC Microbiol. 2010;10:283. Epub 2010/11/13.

11.   Ozeki H, Stocker BA, De Margerie H. Production of colicine by single bacteria. Nature. 1959;184:337-9. Epub 1959/08/01.

12.   Mrak P, Podlesek Z, van Putten JP, Zgur-Bertok D. Heterogeneity in expression of the Escherichia coli colicin K activity gene cka is controlled by the SOS system and stochastic factors. Mol Genet Genomics. 2007;277(4):391-401. Epub 2007/01/12.

13.   Mader A, von Bronk B, Ewald B, Kesel S, Schnetz K, Frey E, et al. Amount of colicin release in Escherichia coli is regulated by lysis gene expression of the colicin E2 operon. PLoS One. 2015;10(3):e0119124. Epub 2015/03/10.

14.   Riley MA, Wertz JE. Bacteriocins: evolution, ecology, and application. Annu Rev Microbiol. 2002;56:117-37. Epub 2002/07/27.

15.   Wojdyla JA, Fleishman SJ, Baker D, Kleanthous C. Structure of the ultra-high-affinity colicin E2 DNase--Im2 complex. Journal of molecular biology. 2012;417(1-2):79-94. Epub 2012/02/07.

16.   Pugsley AP, Goldzahl N, Barker RM. Colicin E2 production and release by Escherichia coli K12 and other Enterobacteriaceae. J Gen Microbiol. 1985;131(10):2673-86. Epub 1985/10/01.

17.   Yang TY, Sung YM, Lei GS, Romeo T, Chak KF. Posttranscriptional repression of the cel gene of the ColE7 operon by the RNA-binding protein CsrA of Escherichia coli. Nucleic acids research. 2010;38(12):3936-51. Epub 2010/04/10.

18.   Gudapaty S, Suzuki K, Wang X, Babitzke P, Romeo T. Regulatory interactions of Csr components: the RNA binding protein CsrA activates csrB transcription in Escherichia coli. Journal of bacteriology. 2001;183(20):6017-27. Epub 2001/09/22.

19.   Suzuki K, Babitzke P, Kushner SR, Romeo T. Identification of a novel regulatory protein (CsrD) that targets the global regulatory RNAs CsrB and CsrC for degradation by RNase E. Genes Dev. 2006;20(18):2605-17. Epub 2006/09/19.

20.   Lechner M, Schwarz M, Opitz M, Frey E. Hierarchical Post-transcriptional Regulation of Colicin E2 Expression in Escherichia coli. PLoS Comput Biol. 2016;12(12):e1005243. Epub 2016/12/16.

21.   Kamensek S, Browning DF, Podlesek Z, Busby SJ, Zgur-Bertok D, Butala M. Silencing of DNase Colicin E8 Gene Expression by a Complex Nucleoprotein Assembly Ensures Timely Colicin Induction. PLoS Genet. 2015;11(6):e1005354. Epub 2015/06/27.

22.   Fornelos N, Browning DF, Butala M. The Use and Abuse of LexA by Mobile Genetic Elements. Trends Microbiol. 2016;24(5):391-401. Epub 2016/03/14.

23.   Babitzke P, Romeo T. CsrB sRNA family: sequestration of RNA-binding regulatory proteins. Current opinion in microbiology. 2007;10(2):156-63. Epub 2007/03/27.

24.   Weilbacher T, Suzuki K, Dubey AK, Wang X, Gudapaty S, Morozov I, et al. A novel sRNA component of the carbon storage regulatory system of Escherichia coli. Mol Microbiol. 2003;48(3):657-70. Epub 2003/04/16.

25.   Liu MY, Gui G, Wei B, Preston JF, 3rd, Oakford L, Yuksel U, et al. The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in Escherichia coli. The Journal of biological chemistry. 1997;272(28):17502-10. Epub 1997/07/11.

26.   Chang HW, Yang TY, Lei GS, Chak KF. A novel endogenous induction of ColE7 expression in a csrA mutant of Escherichia coli. Curr Microbiol. 2013;66(4):392-7. Epub 2012/12/19.

27.   Morales M, Attai H, Troy K, Bermudes D. Accumulation of single-stranded DNA in Escherichia coli carrying the colicin plasmid pColE3-CA38. Plasmid. 2015;77:7-16. Epub 2014/12/03.

28.   Baker CS, Eory LA, Yakhnin H, Mercante J, Romeo T, Babitzke P. CsrA inhibits translation initiation of Escherichia coli hfq by binding to a single site overlapping the Shine-Dalgarno sequence. Journal of bacteriology. 2007;189(15):5472-81. Epub 2007/05/29.

29.   Liu MY, Yang H, Romeo T. The product of the pleiotropic Escherichia coli gene csrA modulates glycogen biosynthesis via effects on mRNA stability. Journal of bacteriology. 1995;177(10):2663-72. Epub 1995/05/01.

30.   Lewis LK, Harlow GR, Gregg-Jolly LA, Mount DW. Identification of High Affinity Binding Sites for LexA which Define New DNA Damage-inducible Genes in Escherichia coli. Journal of molecular biology. 1994;241(4):507-23.

31.   Dickey TH, Altschuler SE, Wuttke DS. Single-stranded DNA-binding proteins: multiple domains for multiple functions. Structure. 2013;21(7):1074-84. Epub 2013/07/05.

32.   Dubey AK, Baker CS, Romeo T, Babitzke P. RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction. Rna. 2005;11(10):1579-87. Epub 2005/09/01.

33.   Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, et al. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science. 2010;329(5991):533-8. Epub 2010/07/31.

34.   Timmermans J, Van Melderen L. Post-transcriptional global regulation by CsrA in bacteria. Cellular and molecular life sciences : CMLS. 2010;67(17):2897-908. Epub 2010/05/07.

35.   Fernández de Henestrosa AR, Ogi T, Aoyagi S, Chafin D, Hayes JJ, Ohmori H, et al. Identification of additional genes belonging to the LexA regulon in Escherichia coli. Molecular Microbiology. 2000;35(6):1560-72.

36.   Gutierrez P, Li Y, Osborne MJ, Pomerantseva E, Liu Q, Gehring K. Solution structure of the carbon storage regulator protein CsrA from Escherichia coli. Journal of bacteriology. 2005;187(10):3496-501. Epub 2005/05/04.

37.   Edelstein A, Amodaj N, Hoover K, Vale R, Stuurman N. Computer control of microscopes using microManager. Current protocols in molecular biology / edited by Frederick M Ausubel [et al]. 2010;Chapter 14:Unit14 20. Epub 2010/10/05.

38.   Youssef S, Gude S, Radler JO. Automated tracking in live-cell time-lapse movies. Integrative biology : quantitative biosciences from nano to macro. 2011;3(11):1095-101. Epub 2011/10/01.

39.   Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic acids research. 2003;31(13):3406-15. Epub 2003/06/26.

40.   Pagano JM, Clingman CC, Ryder SP. Quantitative approaches to monitor protein-nucleic acid interactions using fluorescent probes. Rna. 2011;17(1):14-20. Epub 2010/11/26.

Figure captions
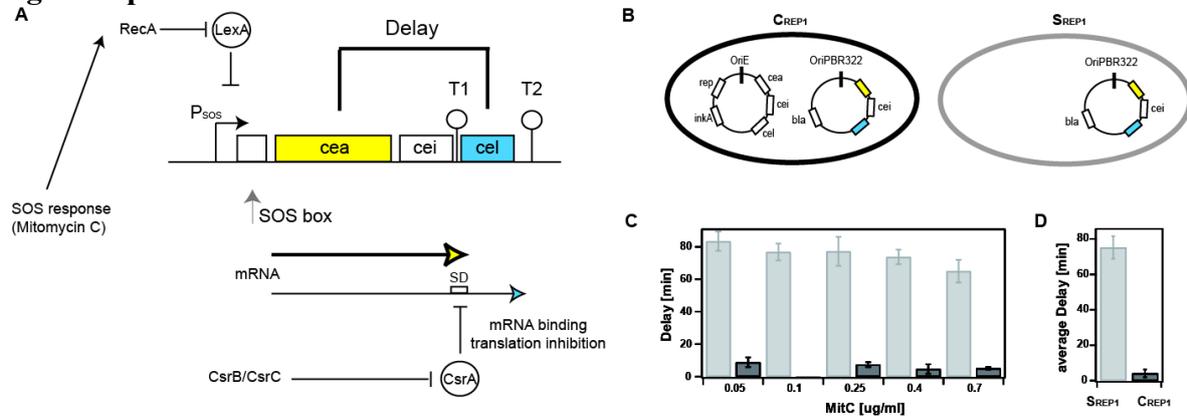
Fig 1: Time between toxin production and release. A) The regulatory network controlling the delay between ColicinE2 production (*cea*) and release (*cel*) in C_{WT}. The operon expresses the genes *cea* (ColicinE2), *cei* (immunity protein) and *cel* (protein inducing cell lysis). In the reporter plasmid, the genes *cea* and *cel* are replaced by genes encoding the fluorescent proteins YFP and CFP, respectively. The transcriptional repressor LexA inhibits expression of the operon. Cell stress causes RecA-mediated auto-cleavage of LexA dimers. Subsequently, a short *cea-cei* and a long *cea-cei-cel* mRNA are produced. Expression of *cel* is further regulated post-transcriptionally by binding of CsrA to the Shine-Dalgarno sequence (SD) within the T1 transcriptional terminator. CsrA itself is regulated by two sRNAs, CsrB and CsrC. **B**) Plasmids present in the two reporter strains C_{REP1} and S_{REP1}. The C_{REP1} strain carries both the reporter plasmid pMO3 and pColE2-P9. The S_{REP1} strain carries only the reporter plasmid. **C**) Dependence of the delay between ColicinE2 production and release by C_{REP1} (black) and S_{REP1} (grey) on the level of external stress (MitC concentration). D) Average delay between ColicinE2 production and release by C_{REP1} (black) and S_{REP1} strains (grey). Error bars depict the standard error of the mean (SEM).

14
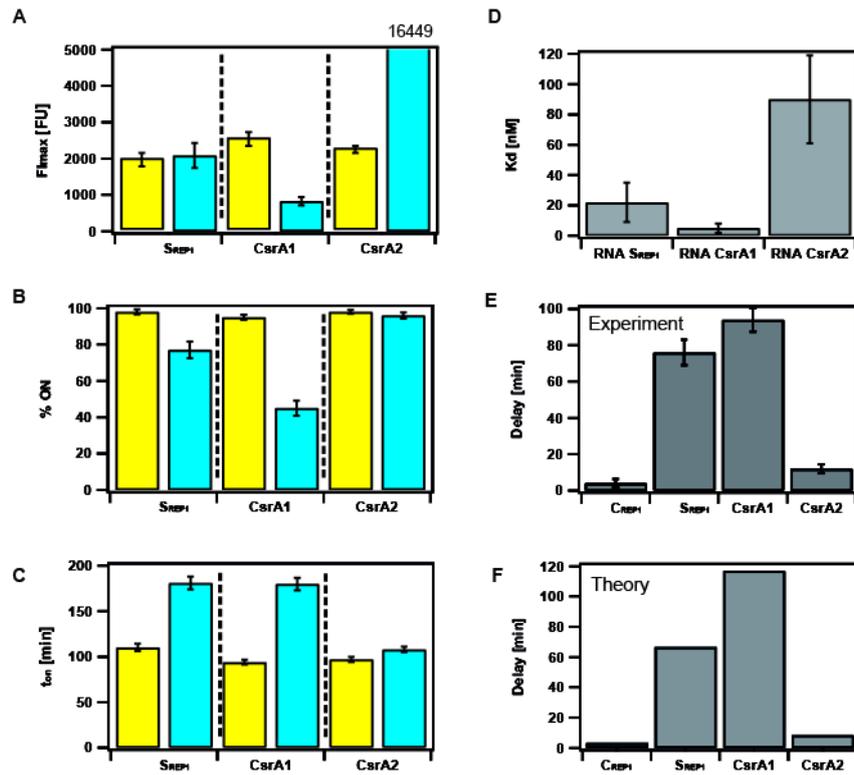
Fig 2: CsrA regulates the delay between ColicinE2 production (*cea*) and release (*cel*). A-C) Yellow: *cea* gene expression, blue *cel* gene expression of cells expressing the ColicinE2 operon in the S_{REP1} strain in comparison to mutant strains CsrA1 and CsrA2. Error bars depict the standard error of the mean (SEM). **A)** Maximal fluorescence intensity, **B)** Cumulative fraction of cells expressing the ColicinE2 operon, **C)** T_{ON} times for *cea* and *cel* gene expression. **D)** Dissociation constants (K_d) for the binding of CsrA to various RNA oligos (**SI**). **E)** Experimentally observed *cea-cel* delay. **F)** Theoretically determined mean *cea-cel* delay. C_{REP1} ($k_M = 0,007$), S_{REP1} ($k_M = 0. 0,007$), CsrA1 ($k_M = 0,0125$), CsrA2 ($k_M = 0.0018$). k_M is the theoretical binding rate constant for CsrA binding to the long mRNA.

15
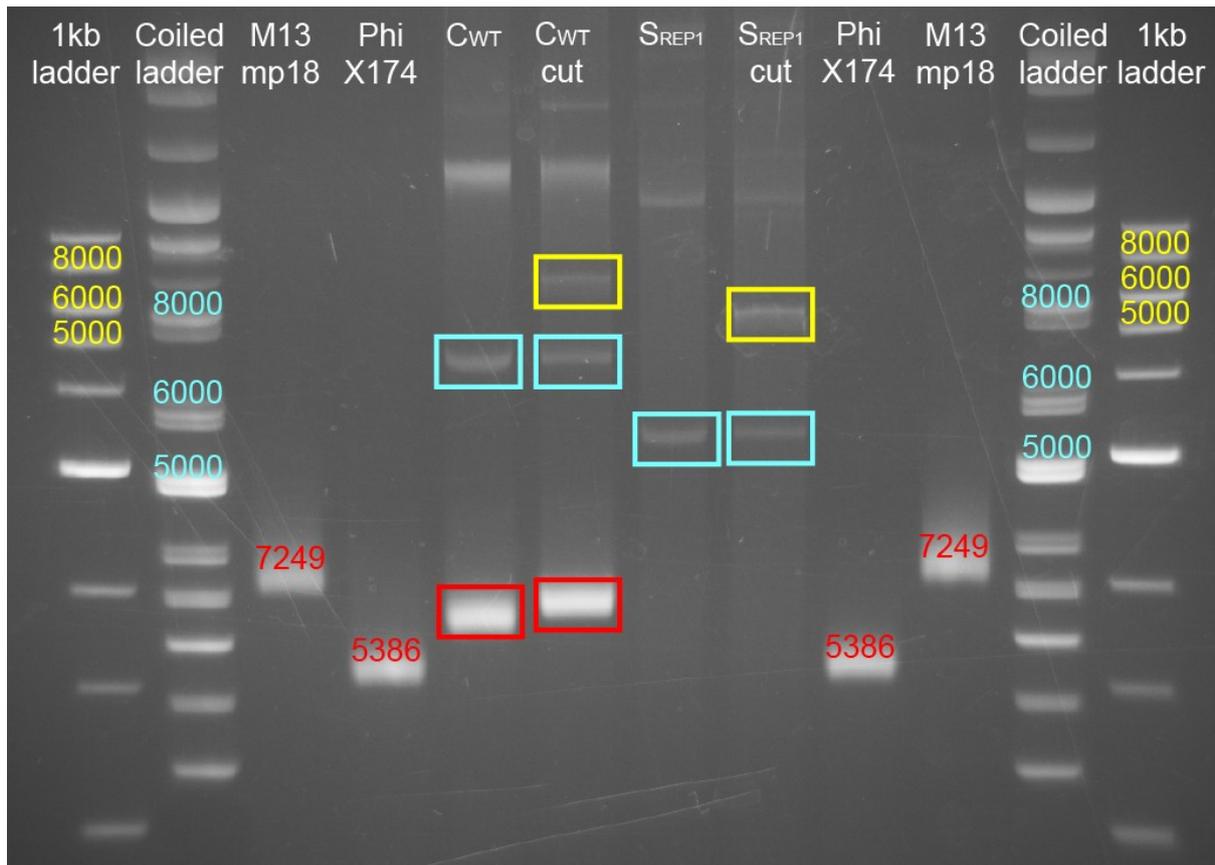
Figure 3: Accumulation of ssDNA in C_WT. Agarose gel of plasmid and ssDNAs extracted from C_{WT} and S_{REP1} strains. Lanes 1-4 and 9-12 were loaded with the indicated markers: 1-kb ladder, super-coiled ladder; 7249-bp ssDNA ring (M13mp18), and 5386-bp ssDNA ring (PhiX174). Lane 5: uncleaved C_{WT} DNA showing the 6800-bp pColE2-P9 dsDNA (blue) and ssDNA (red). Lane 6: C_{WT} DNA cleaved with PvuI, showing the linearized ds pColE2-P9 plasmid (yellow). Lane 7: uncleaved S_{REP1} strain DNA showing the 5600-bp reporter plasmid (blue). Lane 8: S_{REP1} strain DNA cut with PvuI, showing the linearized reporter plasmid (yellow).
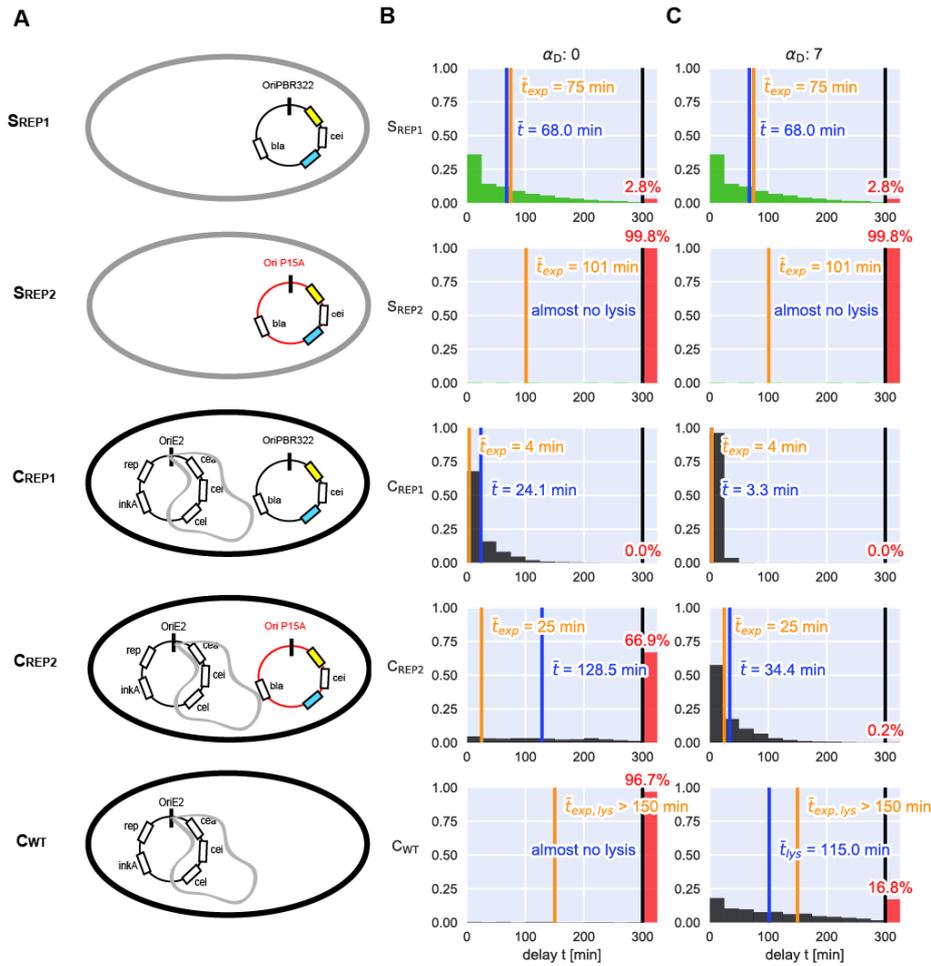
Figure 4: Theoretical analysis emphasizes the importance of the sequestering of CsrA by ssDNA for toxin release in C$_{WT}$. A) Plasmids and ssDNA (grey loops) present in the S$_{REPx}$, C$_{REPx}$ and C$_{WT}$ strains. **B,C)** Theoretical analysis of the *cea-cel* delay for all strains emphasizes the importance of ssDNA for the timing of toxin release in the colicin-producing strains (C$_{REPx}$ and C$_{WT}$). **B:** no ssDNA present. **C:** ssDNA present with $\alpha_D = 7$. The orange line indicates the mean experimental delay for the corresponding strain, the blue line the corresponding theoretical value. The red bar on the right depicts the fraction of cells not undergoing cell lysis in the theoretical model..

17
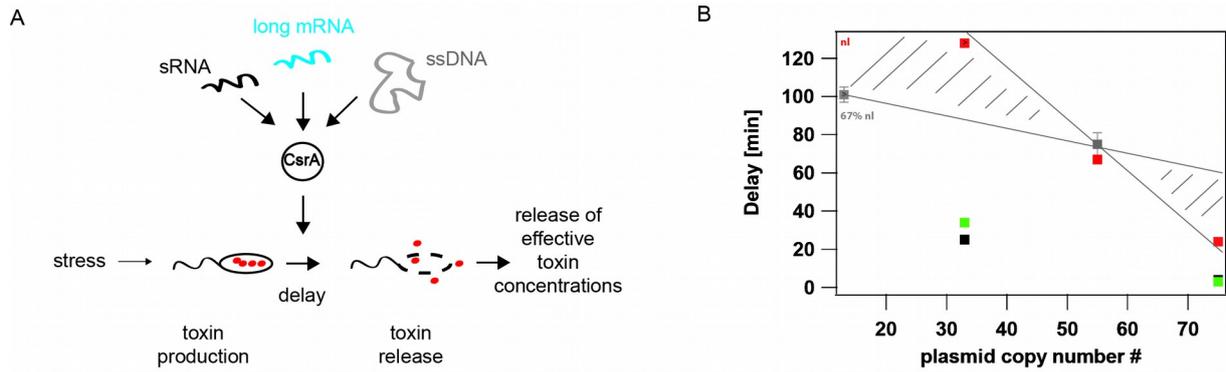
Fig. 5: Time-point of ColicinE2 release is regulated by the global carbon storage regulator CsrA. A) CsrA controls the delay between toxin production and release. This mechanism prevents premature release of ineffective toxin concentrations. CsrA abundance is regulated by several components: the long mRNA transcribed from the pColE2-P9 plasmid and the reporter plasmid pMO3 (or pMO8), the sRNAs CsrB and CsrC, and the newly discovered regulatory element ssDNA originating and accumulating from autonomous rolling circle plasmid replication. B) Our experimental and theoretical data emphasize the importance of the amount of long mRNA that correlates with the plasmid copy number, as well as the presence of ssDNA as CsrA sequestering elements affecting the *cea-cel* delay. C strains are shown in black, S strains are shown in grey. The grey scatched area depicts the area of expected delay times of all strains in dependence to the plasmid copy number only. Please note that due to the fact that at low plasmid copy numbers many cells don't lyse (nl) these delay times cannot be given as an exact value, but are estimated to lie in the depicted grey area. Red dots represent the values of the theoretical analysis in the absence of ssDNA (so plasmid copy number effect only), green markers represent the theoretical values in the presence of ssDNA, which is the case for all C strains.

18

Supplementary information for

Title: CsrA and its regulators control the time-point of ColicinE2 release in *Escherichia coli*

Alexandra Götz, Matthias Lechner, Andreas Mader, Benedikt von Bronk, Erwin Frey and Madeleine Opitz

Supplementary Information (experimental)

Accumulation of single-stranded DNA (ssDNA)

Recently, Morales *et al*.(1), detected the accumulation of ssDNA in *Escherichia coli* cells carrying any one of the plasmids pColE3-CA38, pColE9 and pColE5. This ssDNA originates from rolling-circle replication(1, 2). This type of replication enables an autonomous plasmid to replicate independently of the bacterial cell cycle and cell division(3). The replicons of the plasmids pColE2-P9 and pColE3-CA38 are closely related(4), indicating that asymmetric rolling-circle replication(5) of the pColE2-P9 plasmid (**Fig S5a**) could also lead to ssDNA accumulation, as shown in **Fig 3 and Fig S5b**. Furthermore, sequencing of pColE2-P9 and an analysis of its homology with pColE3-CA38 (**Table S1**) confirms the close relationship between the two plasmids. The Rep protein binds to the ColE2 *ori* to initiate rolling-circle replication(3), leading to ssDNA synthesis(2). The concentration of the Rep protein is held constant by the action of an anti-sense sRNA (the *inkA* gene product) that binds to the 5'-untranslated region of the mRNA encoding the Rep protein(6).

Please note that the reporter plasmid pMO3(7) used in this study is a derivative of pBAD24. Its replication mode(8) differs from that of pColE2-P9 and does not lead to the accumulation of ssDNA (**Fig 3**)(1).

Supplementary Methods

In the following, additional information relating to the construction of the specific mutant strains used in this study is given, together with a description of RNA and ssDNA oligos used for CsrA binding studies. **Table S2** provides an overview of all (mutant) strains used in this study. **Table S3** summarizes all primers used, including those employed in the construction of the mutant strains listed in **Table S4** and in sequencing of pColE2-P9. **Table S4** lists the base-pair changes introduced into the plasmid pMO3 that resulted in plasmids pMO4, pMO5, pMO6, pMO7, as well as the sequences of RNA and ssDNA oligos used for CsrA-binding studies as described in the **Methods** section of the main text.

Supplementary Tables

Table S1: Sequence homology of genes present on the plasmids pColE2-P9 and pColE3-CA38. Numbers in red indicate that the gene extends past the base-pair designated as position 1.

Gene	Sequence position on pColicinE2-P9	Sequence position on pColE3-CA38	Homology E2/E3
rep	3073-3792	3344-4258	97%
ssi	4084-4159	4376-4451	100%
mob	4182-4456	4475-4750	98%
tra	4636-6757,1-98	4929-7118,1-30	97%
incA	2943-3057	3214-3328	100%
colE2/E3 ori	3966-3997	4257-4289	94%

Table S2: Bacterial strains used in this study

Bacterial strain	Strain description	Genetic modification/information	Reference
BZB 1011			(9)
C$_{WT}$ (BZB 1011 E2C)		Carries pColE2-P9	(9)
C$_{REP1}$ (EMO3-C)	BZB 1011 E2C pMO3	Carries pColE2-P9 and the reporter plasmid pMO3	(7)
S$_{REP1}$ (EMO3-S)	BZB 1011 pMO3	Same as EMO3-C without the Colicin E2 plasmid	(7)
LexA1	BZB 1011 pMO4 Derivative of S$_{REP1}$	LexA binding sequence altered on pMO3 to achieve stronger LexA binding, resulting in pMO4	This study
LexA2	BZB 1011 pMO5 Derivative of S$_{REP1}$	LexA binding sequence altered on pMO3 to achieve weaker LexA binding, resulting in pMO5	This study
CsrA1	BZB 1011 pMO6 Derivative of S$_{REP1}$	CsrA binding sequence on pMO3 altered to achieve stronger CsrA binding, resulting in pMO6	This study
CsrA2	BZB 1011 pMO7 Derivative of S$_{REP1}$	CsrA binding sequence on pMO3 altered to achieve weaker CsrA binding, resulting in pMO7	This study
CsrB (EMO3::CsrB)	BZB 1011 CsrB::Kan pMO3 Derivative of S$_{REP1}$	CsrB::Kan, in-frame replacement of CsrB by a kanamycin resistance	This study
CsrC (EMO3::CsrC)	BZB 1011 CsrC::Kan pMO3 Derivative of S$_{REP1}$	CsrC::Kan, in-frame replacement of CsrC with a kanamycin resistance	This study
CsrBC (EMO3::CsrBC)	BZB 1011 CsrB::Cam CsrC::Kan pMO3 Derivative of S$_{REP1}$	CsrC::Kan, CsrB::Cam, in-frame replacement of CsrC by a kanamycin resistance and of CsrB by a chloramphenicol resistance cassette	This study
C$_{REP2}$	BZB 1011 E2C pMO8	Same as C$_{REP1}$, only the origin of replication on pMO3 has been changed to p15A to achieve a lower copy number of 13 copies per cell - pMO8	This study
S$_{REP2}$	BZB 1011 pMO8	Same as S$_{REP1}$, only the origin of replication on pMO3 has been changed to p15A to achieve a lower copy number of 13 copies per cell -> pMO8	This study

Table S3: Primers used in this study. Primers P1-P12 were used for construction of the mutant strains listed in **Table S2**. Primers P13-P24 were used for sequencing pColE2-P9. Primer pairs P25/P26 and P27/28 were used to create the low copy plasmid pMO8.

Name	Sequence	Purpose
P1	5´- GACGGGTACTTTTTGTACTGTACATAAAACCAGTGG - 3´	LexA1 [fwd] cloning
P2	5´- CCACTGGTTTTATGTACAGTACAAAAAGTACCCGTC- 3´	LexA1 [rev] cloning
P3	5´- GACGGGTACTTTTTGATCCCTACATAAAACCAGTGG- 3´	LexA2 [fwd] cloning
P4	5´- CCACTGGTTTTATGTAGGGATCAAAAAGTACCCGTC- 3´	LexA2 [rev] cloning
P5	5´- GGCATTCTTTCACATTAAGGAGTCGTTATG - 3´	CsrA1 [fwd] cloning
P6	5´- CATAACGACTCCTTAATGTGAAAGAATGCC- 3´	CsrA1 [rev] cloning
P7	5´- GCATTCTTTCACAACAAGGATGTGTTATGAAAAAAATAACCGG-3´	CsrA2 [fwd] cloning
P8	5´- CCGGTTATTTTTTTCATAACACATCCTTGTTGTGAAAGAATGC- 3´	CsrA2 [rev] cloning
P9	5´GTGGTCATAAAGCAACCTCAATAAGAAAAACTGCCGCGAA GGATAGCAGG AATTAACCCTCACTAAAGGGCG 3´	ΔCsrB [fwd] cloning
P10	5´TTGTCTGTAAGCGCCTTGTAAGACTTCGCGAAAAAGACGATTCTATCT TCTAATACGACTCACTATAGGGCTC 3´	ΔCsrB [rev] cloning
P11	5´ ACTGATGGCG GTTGATTGTT TGTTTAAAGCAAAGGCGTAA AGTAGCACCCAATTAACCCTCACTAAAGGGCG 3´	ΔCsrC [fwd] cloning
P12	5´GCCGTTTTATTCAGTATAGATTTGCGGCGGAATCTAACAGAAAGCAA GCATAATACGACTCACTATAGGGCTC 3´	ΔCsrC [rev] cloning
P13	5´- ACCGTATCTCCGTCATCAAC -3´	ColE2-1 [fwd] sequencing
P14	5´- CTTCCTGTGAGAACTGC -3´	ColE2-2 [fwd] sequencing
P15	5´- GTAGCGAGCGAATGAG -3´	ColE2-3 [fwd] sequencing
P16	5´- CATGATTGCCGATGTGG -3´	ColE2-4 [fwd] sequencing
P17	5´- GTGGAATACGTGGATTGC -3´	ColE2-5 [fwd] sequencing
P18	5´- GGAGAAGCTATAAACCATG -3´	ColE2-6 [fwd] sequencing
P19	5´- TCTGCTCATGTTTGACAGCTT -3´	ColE2-7 [fwd] sequencing
P20	5´- CTCTGTTCGCATGGTCAG -3´	ColE2-8 [rev] sequencing
P21	5´- CACGTTCGATGTCGTTC -3´	ColE2-9 [rev] sequencing
P22	5´- GAATACATTCTCACACGCTC -3´	ColE2-10 [rev] sequencing
P23	5´- CGTTGTTGTTGCCTGTG -3´	ColE2-11 [rev] sequencing
P24	5´- TCATCCGCCAAAACAGCC -3´	ColE2-12 [rev] sequencing
P25	5´- ATTAAGTCGACGAAGATCCTTTGATCTTTTC -3´	pMO3_noORI SalI [rev] Cloning pMO3 vector without ORI
P26	5´- ATTAAGCATGCAACGCCAGCAACGC -3´	pMO3_noORI SphI[fwd] Cloning pMO3 vector without ORI
P27	5´- ATTAAGTCGACTTGAGATCGTTTTGG -3´	p15A ORI SalI [rev] cloning
P28	5´- ATTAAGCATGCTTTCCATAGGCTCCG -3´	p15A ORI SphI [fwd] cloning

Table S4. Sequences of genetic elements. The first three rows depict sequence changes in the LexA binding site (two overlapping LexA binding SOS boxes) on the pMO3 reporter plasmid, leading to altered LexA binding (pMO4, pMO5). The following three rows show the changes made in the CsrA binding site within the second mRNA loop (which also includes the ribosome binding site of the *cel* gene and the GGA motif recognized by CsrA) that potentiate (pMO6) or weaken (pMO7) CsrA binding (**Fig 2d**). The following three rows list the sequences of RNA oligos used for CsrA binding studies (**Methods, Fig 2d**) and include the alterations in the CsrA binding site mentioned above. The last three rows give the sequences of the 89-bp ssDNA oligos used to study binding of CsrA to ssDNA by gel shift analysis (**Methods**). Bases highlighted in *green* correspond to sequence changes. Bases shown in boldface highlight the GGA motif required for CsrA binding as present within the second mRNA (plasmid), RNA oligo or ssDNA oligo loop. We confirmed the appropriate formation of secondary structures of these oligos using Mfold(10) (**Methods**).

Name	Sequence	Description
pMO3	5'-TTGATCTGTACATAAAACCAGTGGTTTTATGTACAGTATTAA-3'	LexA binding site
pMO4	5'-TTGTACTGTACATAAAACCAGTGGTTTTATGTACAGTATTAA-3'	LexA binding site
pMO5	5'-TTGATCCCTACATAAAACCAGTGGTTTTATGTACAGTATTAA-3'	LexA binding site
pMO3	5' - CACAACAA**GGA**GTCGTTATG - 3'	CsrA binding site second loop
pMO6	5' - CACAACAA**GGA**TGTGTTATG - 3'	CsrA binding site second loop
pMO7	5' - CACATTAA**GGA**GTCGTTATG - 3'	CsrA binding site second loop
RNA Oligo equivalent to sequence of pMO3	5'-Cy5- AUUUAAACAGGGCUGAAAUAUGAAUGCCGGUUGUUUAU**GGA**UGAAUGGCUGGCAUUCUUUCACAACAA**GGA**GUCGUUAUGAAAAAAUA -3'	RNA Oligo with both CsrA binding sites (GGA)
RNA Oligo equivalent to sequence of pMO6	5'-Cy5- AUUUAAACAGGGCUGAAAUAUGAAUGCCGGUUGUUUAU**GGA**UGAAUGGCUGGCAUUCUUUCACAACAA**GGA**UGUGUUAUGAAAAAAUA -3'	RNA Oligo with both CsrA binding sites (GGA)
RNA Oligo equivalent to sequence of pMO7	5'-Cy5- AUUUAAACAGGGCUGAAAUAUGAAUGCCGGUUGUUUAU**GGA**UGAAUGGCUGGCAUUCUUUCACAUUAA**GGA**GUCGUUAUGAAAAAAUA -3'	RNA Oligo with both CsrA binding sites (GGA)
ssDNA Oligo equivalent to sequence of pMO3	5'-Cy5- ATTTAAACAGGGCTGAAATATGAATGCCGGTTGTTTAT**GGA**TGAATGGCTGGCATTCTTTCACAACAA**GGA**GTCGTTATGAAAAAAATA - 3'	ssDNA Oligo with both CsrA binding sites (GGA)
ssDNA Oligo equivalent to sequence of pMO6	5'-Cy5-ATTTAAACAGGGCTGAAATATGAATGCCGGTTGTTTAT**GGA**TGAATGGCTGGCATTCTTTCACAACAA**GGA**TGTGTTATGAAAAAAATA -3'	ssDNA Oligo with both CsrA binding sites (GGA)
ssDNA Oligo equivalent to sequence of pMO7	5'-Cy5- ATTTAAACAGGGCTGAAATATGAATGCCGGTTGTTTATGGATGAATGGCTGGCATTCTTTCACATTAA**GGA**GTCGTTATGAAAAAAATA -3'	ssDNA Oligo with both CsrA binding sites (GGA)

Supplementary Figures
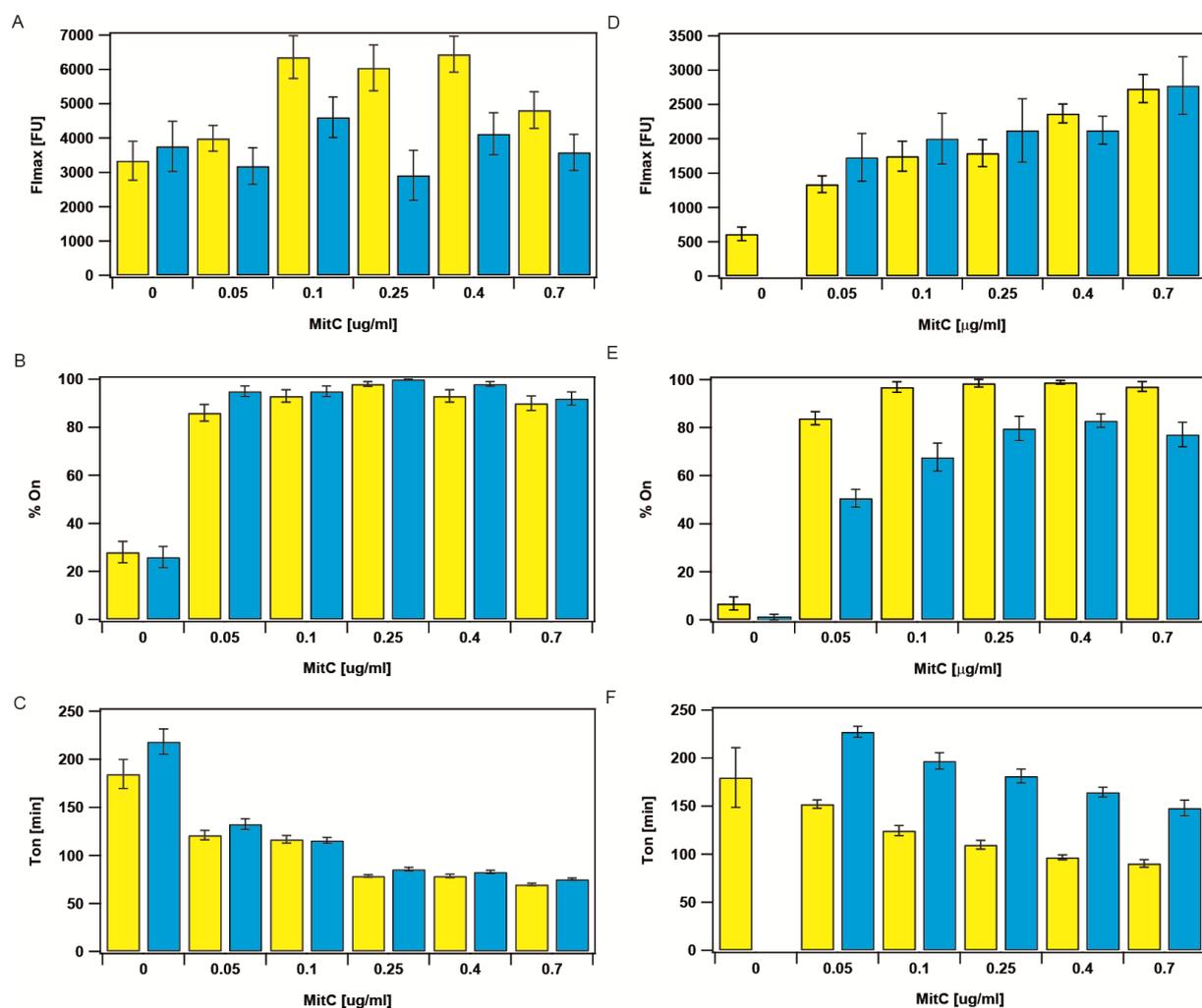
Fig S1: Dependence of the maximal fluorescence intensity, cumulative fraction and time-point of expression start of cells expressing the ColicinE2 operon in the C_{REP1} and S_{REP1} strains on the MitC concentration. Yellow: *cea* gene expression (colicin production), blue: *cel* gene expression (colicin release). Data shown here represent average values obtained from single-cell time-lapse microscopy experiments. A-C) C_{REP1}, D-F) S_{REP1}, A,D) Maximal fluorescence intensity of the cells that express the ColicinE2 operon. B,E) Cumulative fraction of cells expressing the ColicinE2 operon. C,F) Onset (t_{ON}) of *cea* and *cel* gene expression.
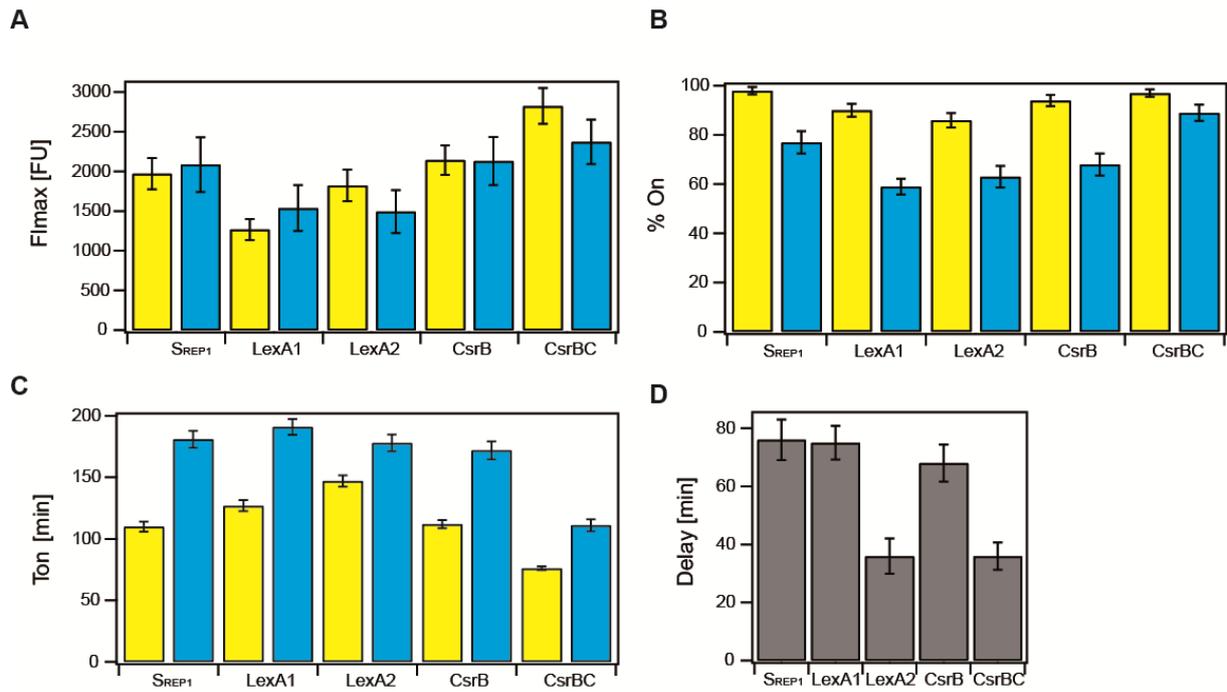
Fig S2: Impact of alteration of the LexA binding site or absence of sRNAs on ColicinE2 expression in S_{REP1} cells. A-C) Yellow: *cea* gene expression (colicin production), blue: *cel* gene expression (colicin release). **A)** Mean maximal expression of the colicin operon (relative to the S_{REP1} strain) in mutant reporter strains bearing altered LexA binding sites (LexA1 and LexA2, **SI, Methods**), or lacking the sRNA CsrB (CsrB), or missing both sRNAs CsrB and CsrC (CsrCB). **B)** Cumulative fraction of cells expressing the colicin operon. **C)** T_{ON} times for *cea* and *cel* gene expression. **D)** Delay between *cea* and *cel* gene expression.
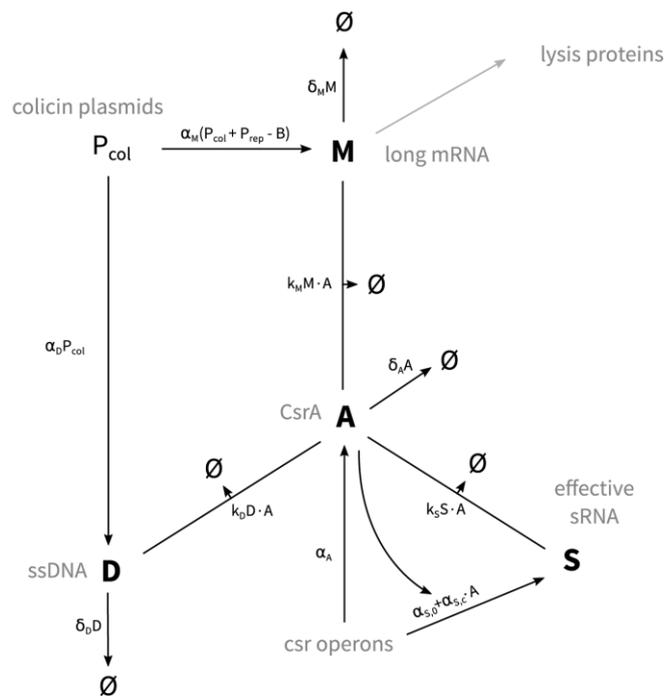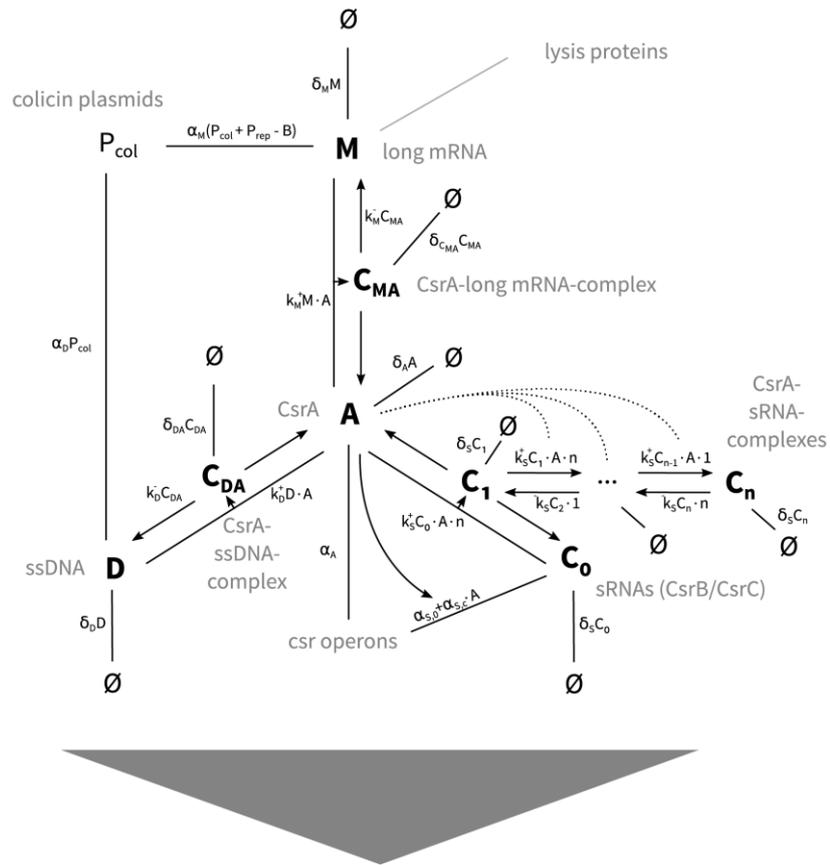
Fig S3: Biochemical network involved in the post-transcriptional regulation of ColicinE2. The top part shows the complete network, involving all interactions and components considered in this work. This complex description of the network can be reduced to the set of effective interactions shown in the lower panel. The derivation of these effective descriptions is given in section 2 of the theory part of the **SI.**
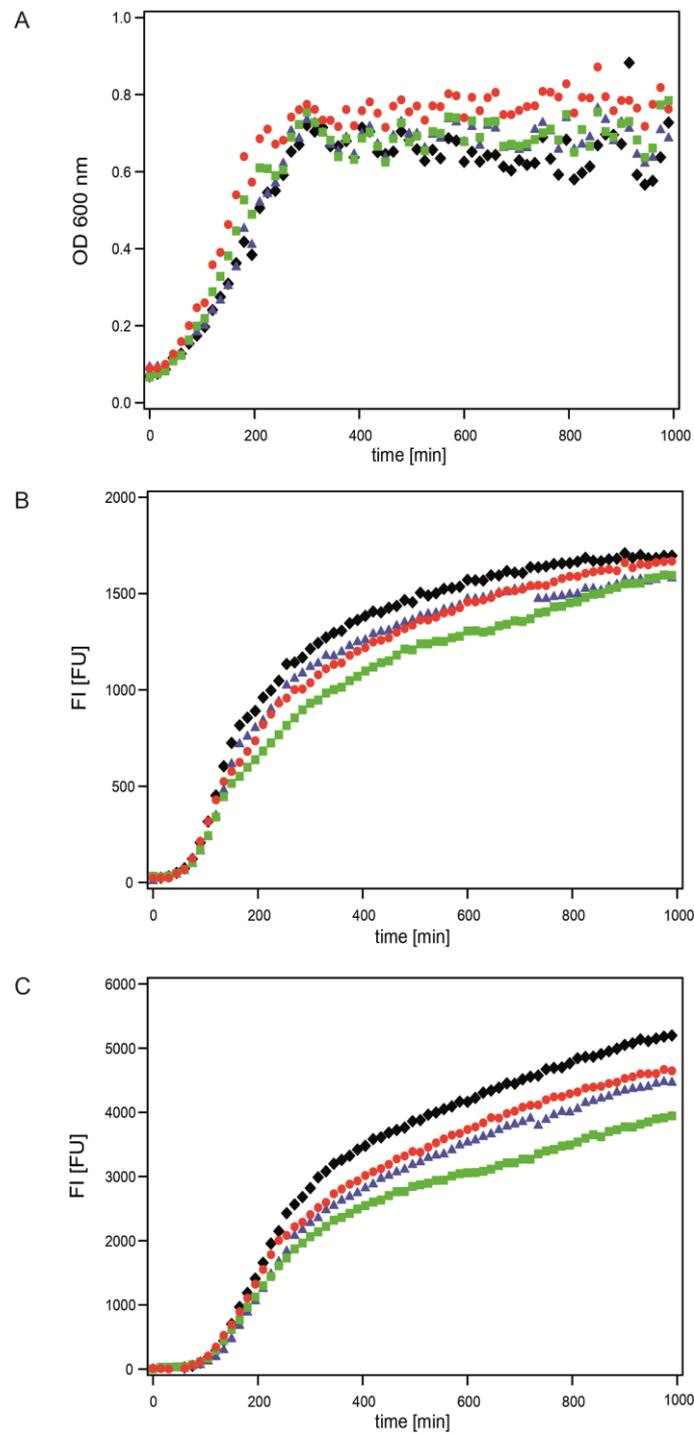
Fig S4: Effect of sRNA knock-out on Colicin E2 expression in long-term experiments (0.25 µg/ml MitC). To determine the importance of sRNA regulation for ColicinE2 expression over a longer period, plate-reader experiments were performed as described in **Methods**. **A)** Absorbance, **B)** Fluorescence intensity of cells expressing *cea* (YFP, colicin production), **C)** Fluorescence intensity of cells expressing *cel* (CFP, colicin release). Red: S_{REP1} strain, Blue: CsrC single sRNA knock-out, Green: CsrB single sRNA knock-out, Black: CsrB/C double sRNA knock-out.
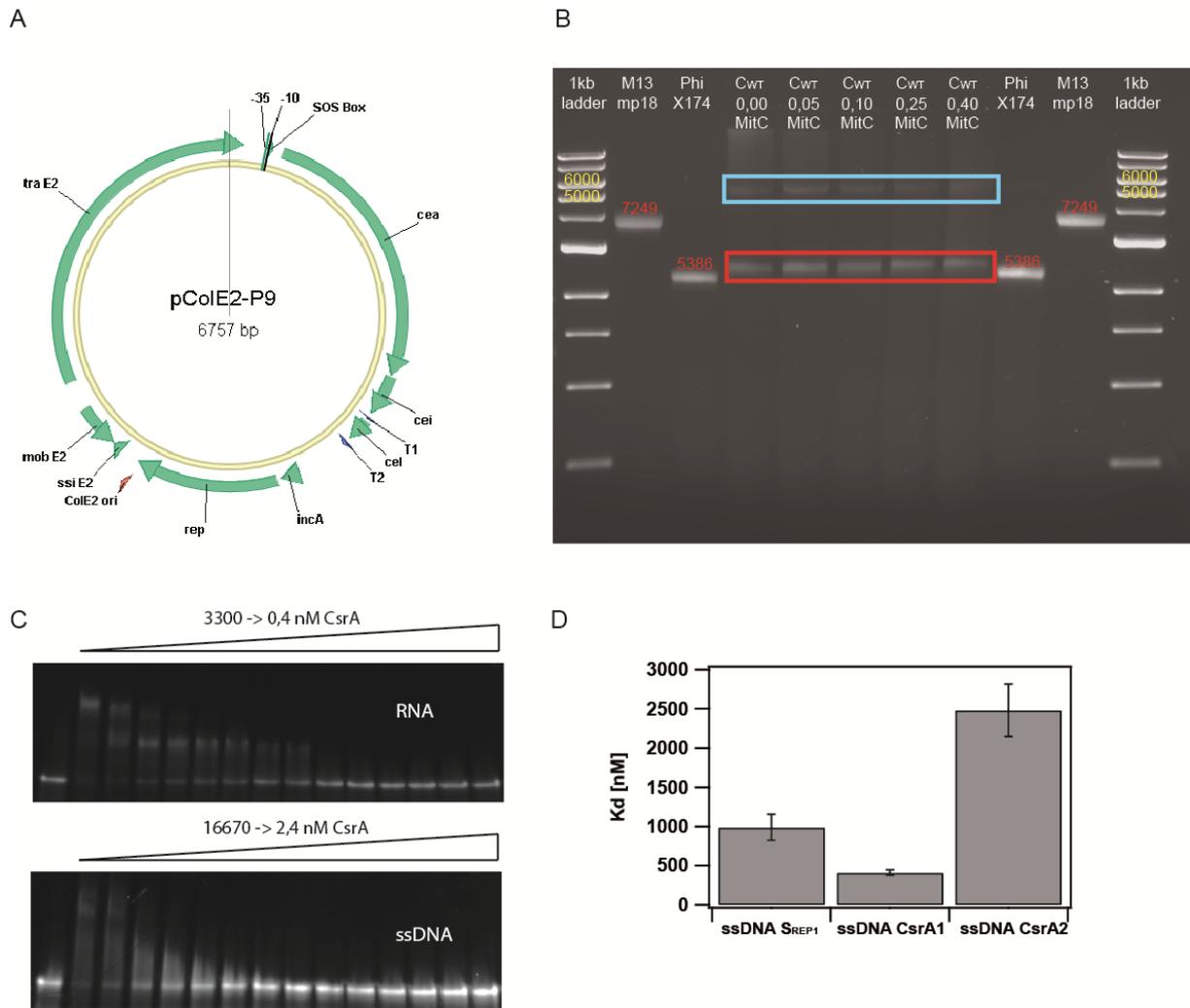
Fig S5: ssDNA accumulation in bacteria carrying pColE2-P9 and binding of CsrA to ssDNA. **A)** Map of pColE2-P9. The plasmid map was created using Vector NTI Expression Version 1.6.0. The plasmid sequence can be accessed via Genbank accession number KY348421. **B)** Accumulation of ssDNA in C_{WT} is independent of the presence of MitC. Agarose gel of plasmid and ssDNAs extracted from C_{WT}. Lanes 1-3 and 9-11 were loaded with the indicated markers: 1-kb ladder, 7249-bp ssDNA ring (M13mp18), and 5386-bp ssDNA ring (PhiX174). Lane 4-8: uncleaved C_{WT} DNA showing the 6800-bp pColE2-P9 dsDNA (blue) and ssDNA (red) at different concentrations of the SOS inducing agent MitC. The staining substance ETBR binds optimal to dsDNA and only to ssDNA if secondary structures (ds part of ssDNA) are present. Hence, binding of ETBR to ssDNA is much lower than to dsDNA. Consequently, the brighter ssDNA bands reflect the high amount of ssDNA accumulated in the bacterial cells. **C)** Gel-shift analysis of CsrA binding to RNA or ssDNA oligos equivalent to the RNA corresponding to pMO3 (**Methods**, **SI**), verifying binding of 1 or 2 CsrA molecules to single RNA and ssDNA oligos (first and second shift, respectively). In the first lane for comparison no CsrA is added. RNA or ssDNA was applied at 5 nM. **D)** Dissociation constants (K_d) for the binding of CsrA to various ssDNA oligos (**SI**). CsrA1 = stronger CsrA binding, CsrA2 = weaker CsrA binding.
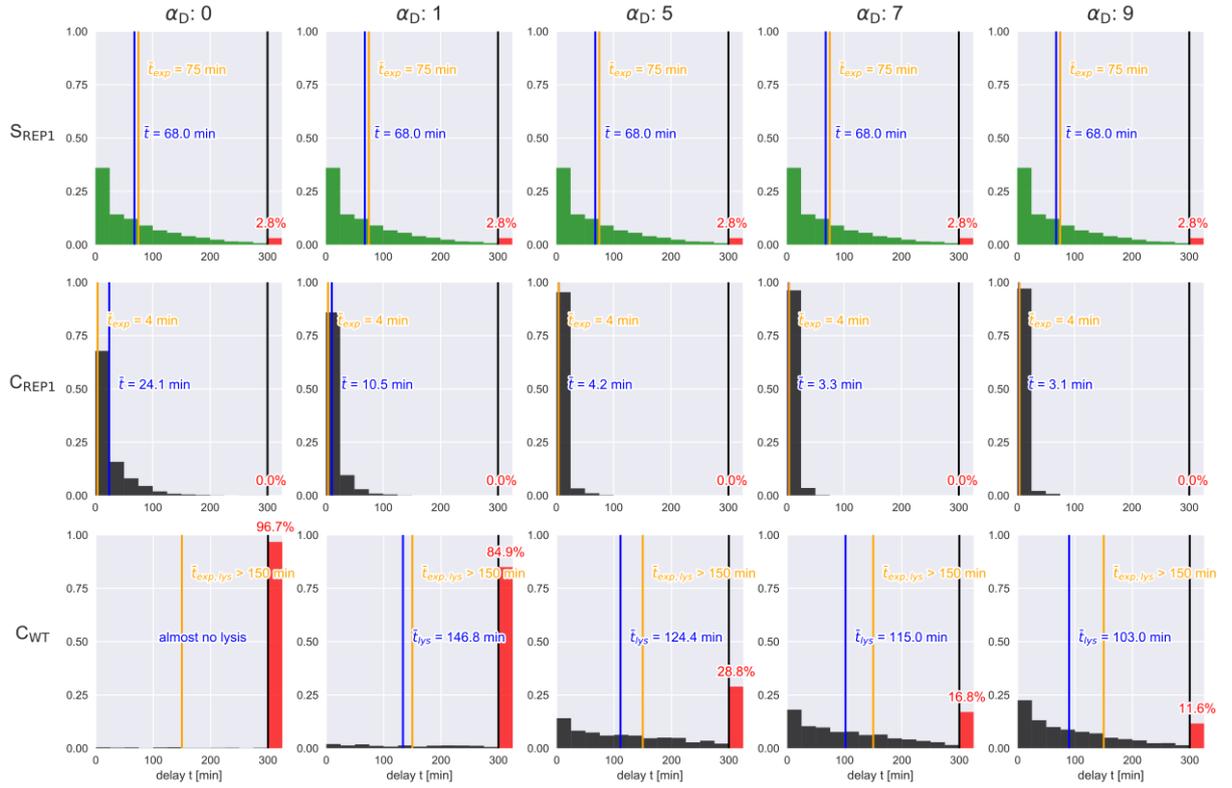
Fig S6: *cea-cel* delay-time distributions and average *cea-cel* delay-times for different ssDNA production rates and strains. The S_{REP1} strain does not produce ssDNA, and is plotted (in green) only for the purpose of direct comparison with the C_{REP1} strain. If no ssDNA is produced ($\alpha_D = 0$), we find that the C_{REP1} strain shows a broader *cea-cel* delay-time distribution, compared to the cases with ssDNA production. The wild-type strain C_{WT} does not lyse at all during the SOS signal for $\alpha_D = 0$. If we increase the ssDNA production rate, we find the experimentally observed behaviour that the C_{REP1} strain shows very short *cea-cel* delays. In the wild-type strain, a certain threshold rate of ssDNA production is required to induce a significant level of lysis, emphasizing the importance of ssDNA for toxin release. Ensemble size: 2000 realisations.
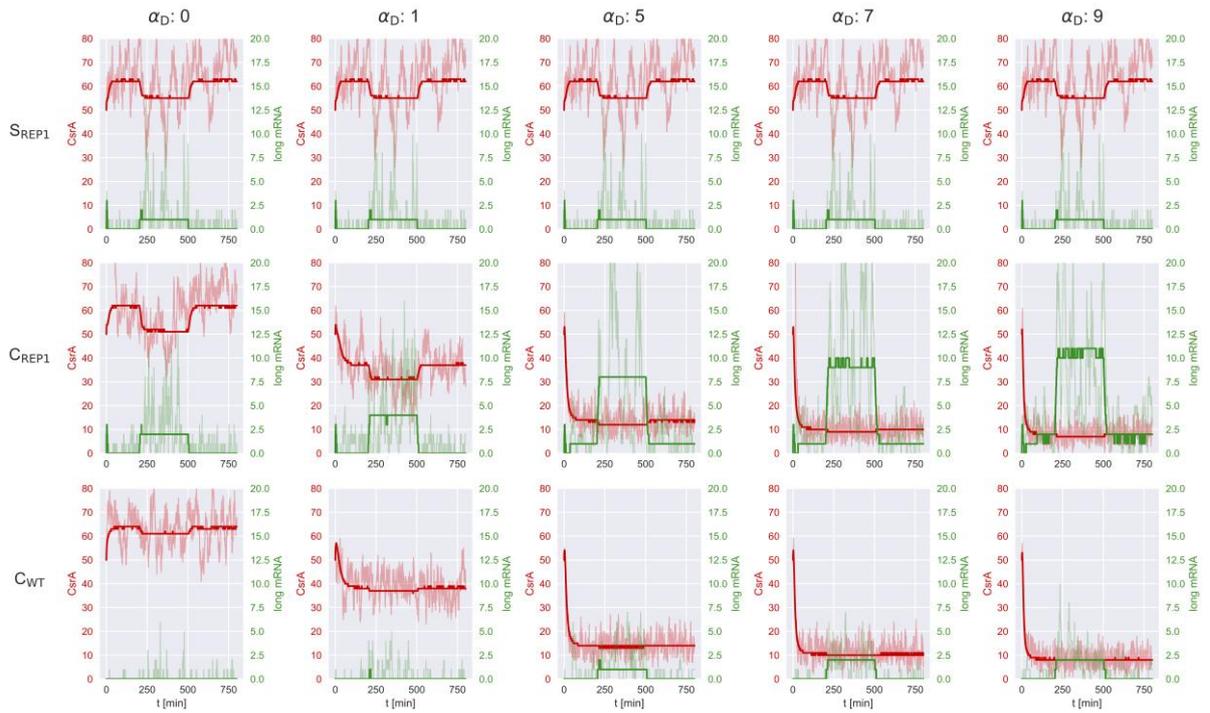
Fig S7: Average of the time evolution of the CsrA and long mRNA abundance for different ssDNA production rates and strains. Between t=200 and t=500, the system is subject to an SOS signal. In all cases, the SOS signal initiates a decrease in CsrA abundance from a previously stable level. This level is determined by the production, binding, and degradation rates of CsrA and its complex partners. As higher CsrA levels take longer and are also less likely to decrease to zero, they also directly affect the duration of the average *cea-cel* delay-time. The three plots for $\alpha_D = 0$ also show a single trajectory of long mRNA and CsrA in light green and light red, respectively. Ensemble size: 2000 realisations.
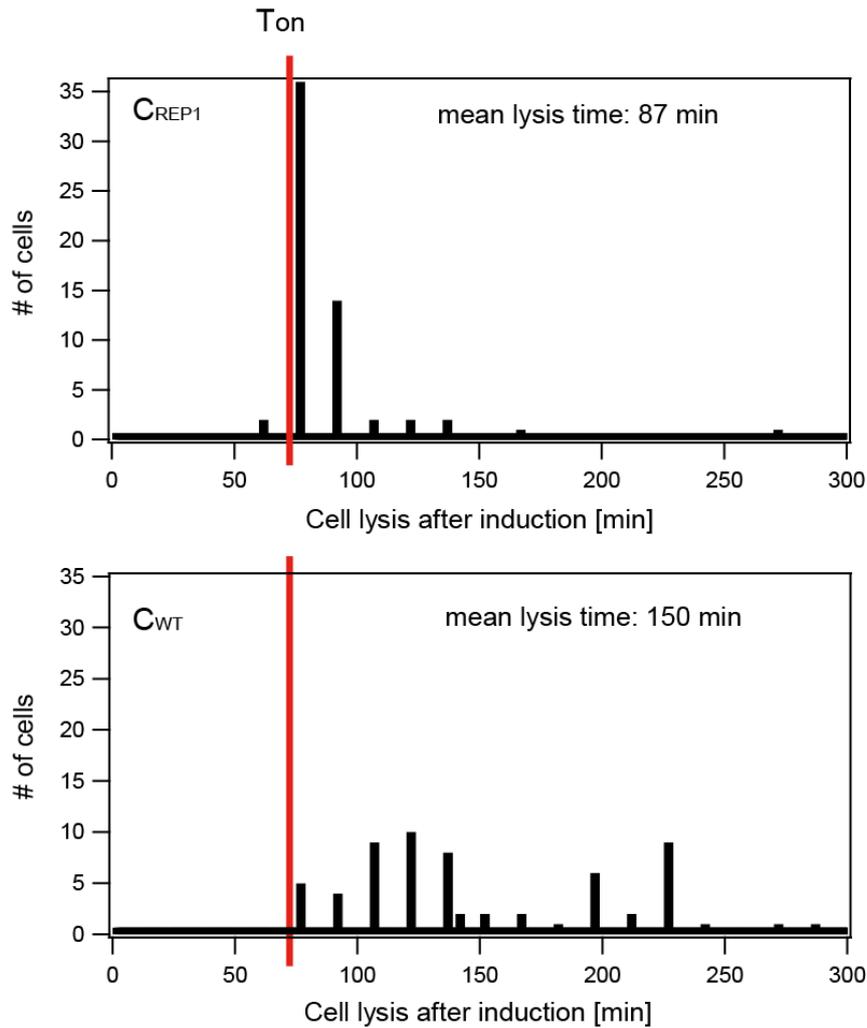
Fig S8: Cell lysis after induction with 0.25 μg/ml MitC. C_{REP1} lyses on average 87 ($\pm$ 3.9) min after induction with MitC, C_{WT} lyses considerably later at 150 ($\pm$ 6.9) min. For C_{REP1}, mean cell lysis nearly coincides with the T_{ONcea} at 69.71 $\pm$ 0.77 min(7) (red line). This interval between T_{ONcea} constitutes the delay between the SOS signal (MitC induction) and the start of *cea* gene expression. In our theoretical model, this initial delay is very short, as the system here switches directly from the pre-SOS signal 'OFF' state into the post-SOS signal 'ON' state. Hence, the red line also depicts the time-point of this switch (0 min in **Fig 4**).

Supplementary References (experimental)

1.   Morales M, Attai H, Troy K, Bermudes D. Accumulation of single-stranded DNA in Escherichia coli carrying the colicin plasmid pColE3-CA38. Plasmid. 2015;77:7-16. Epub 2014/12/03.

2.   Khan SA. Plasmid rolling-circle replication: highlights of two decades of research. Plasmid. 2005;53(2):126-36. Epub 2005/03/02.

3.   Yagura M, Nishio SY, Kurozumi H, Wang CF, Itoh T. Anatomy of the replication origin of plasmid ColE2-P9. Journal of bacteriology. 2006;188(3):999-1010. Epub 2006/01/24.

4.   Aoki K, Shinohara M, Itoh T. Distinct functions of the two specificity determinants in replication initiation of plasmids ColE2-P9 and ColE3-CA38. Journal of bacteriology. 2007;189(6):2392-400. Epub 2007/01/24.

5.   del Solar G, Kramer G, Ballester S, Espinosa M. Replication of the promiscuous plasmid pLS1: a region encompassing the minus origin of replication is associated with stable plasmid inheritance. Mol Gen Genet. 1993;241(1-2):97-105. Epub 1993/10/01.

6.   Sugiyama T, Itoh T. Control of ColE2 DNA replication: in vitro binding of the antisense RNA to the Rep mRNA. Nucleic acids research. 1993;21(25):5972-7. Epub 1993/12/25.

7.   Mader A, von Bronk B, Ewald B, Kesel S, Schnetz K, Frey E, et al. Amount of colicin release in Escherichia coli is regulated by lysis gene expression of the colicin E2 operon. PLoS One. 2015;10(3):e0119124. Epub 2015/03/10.

8.   del Solar G, Giraldo R, Ruiz-Echevarria MJ, Espinosa M, Diaz-Orejas R. Replication and control of circular bacterial plasmids. Microbiol Mol Biol Rev. 1998;62(2):434-64. Epub 1998/06/10.

9.   Kerr B, Riley MA, Feldman MW, Bohannan BJ. Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. Nature. 2002;418(6894):171-4. Epub 2002/07/12.

10.   Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic acids research. 2003;31(13):3406-15. Epub 2003/06/26.

Supplementary Information (Theory): Theoretical Model for ColicinE2 Expression Including the additional CsrA sequestering Element ssDNA

1 The biological system

The biological phenomenon we wish to describe by means of a quantitative theoretical model is the regulation of ColicinE2 release in *E. coli*. ColicinE2 is a bacterial toxin encoded by the gene *cea*, which is part of the ColicinE2 operon located on a plasmid. This operon also contains genes for an immunity protein (*cei* gene) and a lysis protein (*cel* gene). The lysis protein is part of the operon as cell lysis is the only way to release the toxin into the environment. Since lysis also means the death of the cell, the release of ColicinE2 is highly regulated. Previous studies have revealed the regulatory components controlling ColicinE2 production and release on both the transcriptional and post-transcriptional levels:

- The *transcription* of the operon is regulated by the operon's repressor *LexA*, which is part of the *E. coli* SOS response regulatory network [1, 2]. Stressful events, such as DNA damage activate a SOS response system [1], which stochastically triggers the transcription of the operon by degradation of *LexA*. Once transcription starts, two mRNA transcripts are produced: *short mRNA*, containing only the toxin and immunity protein, and *long mRNA*, which contains also the lysis protein [3].

- The *post-transcriptional* regulation (see Fig. 1 and also Fig. S3) acts on the long mRNA only. To our knowledge, its only regulator is the protein *CsrA*, which binds to the Shine-Dalgarno sequence that

is located on the long mRNA between the sequences coding for the immunity and lysis proteins [3]. When a CsrA protein binds to long mRNA and thus forms a complex with it, the gene for the lysis protein can no longer be translated, and thus the cell does not lyse [3]. By preventing lysis protein expression, CsrA regulates the release of ColicinE2. The abundance of CsrA itself is known to be regulated by the two CsrA-sequestering short RNAs (sRNAs) *CsrB* and *CsrC*, which both have several CsrA binding sites [4, 5]. Our current study suggests that, in addition, rings of single-stranded DNA (ssDNA) also sequester CsrA, and therefore represent a novel CsrA regulator. This ssDNA is created as an intermediate during the rolling circle replication of the ColicinE2 plasmids.

A particular example for the importance of these regulatory interactions is the delay between production and release of ColicinE2, which has recently been studied experimentally [6]: The translation of lysis proteins from long mRNA (and therefore lysis itself) can only start if there are free long mRNAs, that is, long mRNAs that are not bound to CsrA. From what is known about CsrA interactions, we assume in our biochemical model that a CsrA molecule can no longer regulate long mRNA when it is either sequestered, or degraded. Previous studies [7] show that CsrA is highly abundant during growth phase, mainly in form of CsrA complexes. When an SOS response is triggered, however, the production of long mRNA increases such that free CsrA abundance decreases, and eventually lysis proteins are translated from free long mRNA. This process does not happen instantaneously: Due to stochasticity in the SOS response system and the time it takes to produce, bind or degrade the regulatory components involved, we find a delay between the expression of the unregulated *cea* gene (part of the short mRNA) and the CsrA-regulated *cel* gene (part of the long mRNA). This delay is presumably not just a byproduct of regulation, but has also a biological function: It gives the cell time to accumulate ColicinE2, and thus allows for higher toxin concentrations upon the release. Moreover, it presumably also acts as a safety buffer, which prevents premature cell lysis, for instance due to fluctuations in the system [6].

In contrast to the abundances of the regulatory components, the *cea-cel-*delay is a quantity that can readily be measured experimentally by *reporter plasmids* inserted in the *E. coli* cells. These reporter plasmids carry the ColicinE2 operon, but the toxin (*cea*) and lysis (*cel*) gene are replaced by two different fluorescence protein genes (CFP and YFP). Since only these two

strain	number of reporter plasmids	number of pColE2P9 plasmids	ssDNA
S_{REP1}	≈ 55	–	–
S_{REP2}	≈ 13	–	–
C_{REP1}	≈ 55	≈ 20	accumulates
C_{REP2}	≈ 13	≈ 20	accumulates
C_{WT}	–	≈ 20	accumulates

Table S5: The five different strains and the abundance of the genetic elements that differentiate them [8].

genes are replaced, the reporter plasmids have the same promoter as the ColicinE2 plasmid, and the long mRNA transcript of the reporter plasmid has the same CsrA binding site as the original long mRNA. Consequently, a reporter plasmid behaves like the ColicinE2 plasmid, but produces two types of fluorescence proteins instead of toxin and lysis proteins. Therefore, upon introducing reporter plasmids to an *E. coli* cell, one can measure the time points of production of the corresponding fluorescence proteins; these measured time points then coincide with the time points of toxin and lysis protein production. In this study, we use two reporter plasmid types, which differ by their mean abundance in the cell: type 1 (pMO3) accumulates to about 55 plasmids per cell, whereas type 2 (pMO8) only to about 13 plasmids per cell.

With three different plasmid types, the original and the two reporter plasmids, we can construct five different strains (see also Table S5): First, the wild-type strain, C_{WT}, which carries only ColicinE2 plasmids. Inserting reporter plasmids to this strain creates, depending on the reporter plasmid type inserted, either a strain called C_{REP1} or a strain called C_{REP2}. Completely replacing the ColE2 plasmid with one of the reporter plasmid types creates another two strains, referred to as S_{REP1} and S_{REP2}. Our experiments show that the plasmid types also differ in the production of ssDNA: cells carrying the ColicinE2 plasmid do accumulate ssDNA (see Fig. 3), while this is not the case for the S_{REP1} and S_{REP2} strain cells, which only contain reporter plasmids (see Fig. 3 and Supplementary Information).

As discussed in detail in section 2, we will use mathematical modelling to infer the delay of the wild-type strain from the delay measured in the other strains containing the reporter plasmid. Before doing so, we recapitulate the main experimental findings on the delay-times in the bacterial strains containing the reporter plasmid.

In our experiments (see main text) we found that the C_{REP1} strain shows

no significant delay between *cea* and *cel* gene expression, whereas the S_{REP1} strain has a significant mean delay of 75 minutes. Moreover, we observed that the S_{REP1} strain has a broad delay-time distribution around this mean value. This raised the question as to the source of this difference. The C_{REP1} and S_{REP1} strains differ only in their plasmid composition and are both genetically identical (see Table S5). From this we conclude that the presence of ColicinE2 plasmids in the C_{REP1} strain introduces further regulatory elements (compared to the reporter plasmids), which are responsible for the shorter *cea-cel* delay times compared to the S_{REP1} strain. The reporter and the ColicinE2 plasmids contain the same regulatory sequences (see **Methods**), which means that the additional regulatory elements cannot be different mRNA transcripts specifically produced by the ColicinE2 plasmid. From what is known about the two plasmids and the regulatory network of ColicinE2 (see above), two mechanisms could in principle account for the shorter delay times in C_{REP1}, which are: First, additional production of CsrA sequestering long mRNA due to the larger plasmid copy number, and second, the accumulation of ssDNA which, as our study shows, can also sequester CsrA.

2 Mathematical model of the ColicinE2 release

In the following, we develop a mathematical model that enables us to investigate the regulation of ColicinE2 release in all five strains (C_{REP1}, C_{REP2}, C_{WT}, S_{REP1}, and S_{REP2}). The model accounts for all necessary regulatory components, including the ssDNA and the different plasmid compositions. We validate this model by reproducing the experimentally observed delay time distributions for the S_{REP1} strain. Variation of ssDNA production in the model then allows us to quantify the impact of ssDNA production and plasmid copy number on the *cea-cel* delay. Moreover, the model enables us to infer the behaviour of the C_{WT} strain, for which the *cea-cel* delay cannot be directly measured experimentally. The inferred behaviour can be validated by comparison with experimentally measured lysis times, see Fig. S8.

Several experimental studies defined and probed the regulatory networks and components involved in *E. coli* SOS responses, as well as ColicinE2 production and release [1, 3, 6, 8]. Starting from these experimental results, the regulatory interactions have also been studied using mathematical models [2, 9–11]: For the transcriptional regulation network of the *E. coli* SOS response system, a stochastic model has been presented, which is able

to reproduce the distribution of stochastic SOS activity peaks [2]. For the post-transcriptional regulation of ColicinE2 release, we recently introduced a hierarchical three-component model [9] involving long mRNA, CsrA, and an effective sRNA. This model was also combined with the stochastic SOS signal model from Ref. [2] to emulate the response of ColicinE2-producing bacteria to external stress. With this combined approach, the model shows that sRNA reduces internal fluctuations and helps controlling the level of CsrA. Moreover, the model predicts stochastically distributed delays between SOS signal and lysis, which is also seen in experiments with the S_{REP1} strain.

In this section, we extend our previous model [9], taking into account the new experimental findings presented in the main text. In particular, we incorporate the additional regulator ssDNA as well as the different plasmid copy numbers and types. For this step, it is important to know the derivation of the previous, three-component model, which is why we outline the derivation of the previous model as we develop our new model from scratch. For a detailed derivation of the three-component model, we refer the reader to Ref. [9].

2.1 Regulatory network

Our goal is to design a stochastic model that enables us to investigate the dynamics of the regulatory networks involved the SOS response and the ensuing synthesis and release of ColicinE2. To this end, we first formulate the interactions of the regulatory components as a set of (deterministic) differential equations, that is, as a mass-action model. This approach disregards any spatial effects and consider the system as *well mixed*.

Extending our previous study [9], we build a mass-action model for the SOS response, and the regulatory network for ColicinE2 production and release from the following assumptions and properties of the components (see also Fig. S3):

- The abundances of long mRNA, CsrA and effective single-binding-site sRNA (see below) are denoted by M, A and S, respectively. These abundances give the number of *free* components, that is, the number of long mRNA, CsrA and sRNA molecules that are not bound in a complex. Moreover, P_{COL} and P_{REP} denote the copy number of ColicinE2 and the reporter plasmids, respectively (there is no need to distinguish between the two reporter plasmid types for P_{REP} as they do not occur in the same cell at the same time).

- The response to external stress ("SOS response") is regulated by the LexA/RecA system [1], which we incorporate into our model using the differential equations given in Ref. [2]. This model accounts for the production, degradation and (un)binding of the proteins LexA (L) and RecA (R), the mRNAs they are translated from (M_l and M_r, respectively), as well as the number of repressed promoters controlling the transcription of these mRNAs (B_l and B_r, respectively). In this system, LexA acts as repressor: as long as a LexA protein is bound to the promoter region of the RecA or LexA operon, no mRNA is produced. The number of repressed RecA and LexA promoters increases if LexA binds to an unrepressed promoter, and decreases as it unbinds. Therefore, the differential equations for B_l and B_r contain two terms each: a production term proportional to the abundances of LexA and unrepressed promoters, and a degradation term proportional to the number of repressed promoters. In an *E. coli* cell, there is only one promoter for each LexA and RecA, which means that B_l and B_r can take either the values 0 or 1. The differential equations then read

$$\partial_t B_r = k_r^+ (1 - B_r)L - k_r^- B_r, \tag{1}$$

$$\partial_t B_l = k_l^+ (1 - B_l)L - k_l^- B_l, \tag{2}$$

where the $k^\pm$ denote the attachment and detachment rates of LexA to/from the promoter indicated by the subscript. From unrepressed promoters the respective mRNA is transcribed, and hence, the mRNA production depends linearly on the number of unrepressed promoters. Once produced, the mRNA can spontaneously degrade. Therefore, the differential equations for the mRNAs also contain two terms each, and read:

$$\partial_t M_r = \alpha_{M_r}(1 - B_r) - \delta_{M_r} M_r, \tag{3}$$

$$\partial_t M_l = \alpha_{M_l}(1 - B_l) - \delta_{M_l} M_l, \tag{4}$$

where α and δ give the per capita production and degradation rate of the component indicated by the subscript. These mRNAs are translated to RecA and LexA proteins, respectively. Hence, the production terms of the two proteins are proportional to the respective mRNA abundance. The number of proteins decreases by spontaneous degradation. As the abundance of RecA is only affected by these two processes, its differential equation reads:

$$\partial_t R = \alpha_R M_r - \delta_R R, \tag{5}$$

6

where α_R and δ_R denote the per capita production and degradation rate of RecA. Since LexA acts as regulator in the LexA/RecA-system, its abundance is also affected by the interactions with the promoters. Consequently, the terms from eqs. (1) and (2) appear in the differential equation for LexA, but with opposite sign. Moreover, in case of an SOS signal, RecA depletes LexA, motivating an additional degradation term bilinear in L and R, with the degradation constant c_p. Apart from these interactions with the LexA/RecA-system, LexA is also the repressor of the colicin operon. Therefore, the LexA/RecA SOS response system interacts with the regulatory system of ColicinE2 production and release via B, the number of repressed promoters of the ColicinE2 operon. Its differential equation contains two terms analogous to the LexA and RecA promoters:

$$\partial_t B = k_P^+ (P_{COL} + P_{REP} - B)L - k_P^- B, \tag{6}$$

where the $k_P^\pm$ denote the attachment and detachment rates of LexA repressor to/from the ColicinE2 promoter. Unlike B_r and B_l, B can take values between 0 and $P_{COL} + P_{REP}$. The terms from eq. (6) apppear, again with opposite sign, also in the differential equation for LexA, which, altogether, reads

$$\partial_t L = \alpha_L M_l - \delta_L L - k_l^+(1 - B_l)L + k_l^- B_l - k_r^+(1 - B_r)L$$
$$+ k_r^- B_r - k_P^+(P_{COL} + P_{REP} - B)L + k_P^- B - c_p R L, \tag{7}$$

where α_L and δ_L give the per capita production and degradation rate of LexA. For a detailed discussion of these equations, we refer to Ref. [2]. Note that the SOS response system, eqs. (1)-(7), interacts with the ColicinE2 regulatory network only through the parameter B (see also next bullet point).

- The total production rate of long mRNA in the cell is proportional to the number of unrepressed ColicinE2 promoters in the cell. This number is given by the total number of plasmids in the cell, $P_{COL} + P_{REP}$, minus B, the number of promoters with the repressor LexA bound to it. Hence, the production rate of long mRNA reads

$$\alpha_M(P_{COL} + P_{REP} - B), \tag{8}$$

where α_M is the production rate per unrepressed promoter. Note that considering different plasmid types generalizes our earlier work presented in Ref. [9].

- CsrA is produced at a constant rate, α_A.

- The ColicinE2 system has two different regulatory sRNAs: CsrB and CsrC. Apart from having different numbers of CsrA binding sites and slightly different half-lifes, their mode of binding with CsrA is very similar. Hence, we assume that we can describe their regulatory impact by a single *effective sRNA* with corresponding effective parameters (see Ref. [9] for details). Using effective sRNAs in a mathematical model is indeed supported by experiments, which show that the knock-out of either CsrB or CsrC causes a compensating overproduction of the other sRNA (see the main text, and Ref. [5]). This compensation is a natural consequence of a positive regulatory effect of CsrA abundance to sRNA production (see bullet point below), and highlights the functional equivalence of CsrB and CsrC. In Ref. [9] we also showed that this effective sRNA, which contains $N \approx 10$ CsrA binding sites is equivalent to N *effective single-binding-site sRNAs*. This drastically reduces the mathematical complexity of the model.

- Several studies found that the production of CsrB and CsrC is indirectly regulated by the abundance of CsrA via the BarA/UvrY-system [3–5]. Since the details of this interaction are largely unknown, we model this positive regulation with an sRNA production rate that is a linear function of the CsrA abundance. In addition to this linear term, we also introduce a constant baseline production term, since studies show that sRNAs are also produced (at very low levels) in the absence of free CsrA [5]. Both production terms contain a factor N, as we consider effective single-binding-site sRNAs in our model (see the previous bullet point). The production term of the effective single-binding-site sRNAs thus reads

$$\alpha_{S,0} N + \alpha_{S,c} A \cdot N,$$

  with the baseline production rate $\alpha_{S,0}$, the linear coupling coefficient $\alpha_{S,c}$, and the abundance of CsrA proteins A. Note that we did not consider the positive feedback of CsrA on sRNA production in Ref. [9].

- The degradation rates of long mRNA, CsrA and the effective sRNA are each proportional to their respective abundance, and read $\delta_M M$, $\delta_A A$ and $\delta_S S$, respectively.

- CsrA can bind to both long mRNA and the effective sRNA, and thus forms CsrA-long mRNA and CsrA-sRNA complexes (C_{MA} and C_{SA},

respectively). In line with previous studies [12, 13] and our three-component model [9], we assume that the formation and disassembly of these complexes is much faster than the other processes involved in post-transcriptional regulation. Therefore, we can employ adiabatic elimination, $\partial_t C_{MA} \equiv 0$ and $\partial_t C_{SA} \equiv 0$. In Ref. [9], we show that this enables us to combine the formation, disassembly and degradation of the complexes into effective binding parameters, k_M and k_S. As a consequence, we can solve for the complex abundances, C_{MA} and C_{SA}, and eliminate them from our set of differential equations (see Ref. [9] for details).

- The precise mechanism for the degradation of CsrA-sRNA and CsrA-long mRNA complexes is not known. Here, we assume that *CsrA dimers are always degraded once their complex partner is degraded* (in other words: CsrA cannot "survive" the degradation of its partner).

- CsrA is a main regulator in growing *E. coli* cells, which is known to bind to over 700 different targets [14, 15]. In the Supplementary Information of Ref. [9] we show how one can eliminate the many targets of CsrA to obtain a reduced system, which contains only the components that are changed by the processes the model focusses on (in this case: SOS-induced production and release of ColicinE2). In the mathematical model presented in this section, we reduce the system to three CsrA targets: long mRNA, sRNA, and (see below) ssDNA. In agreement with experiments [7], the production rate of the effective sRNA is large compared to the production of long mRNA and ssDNA, such that the vast majority of sequestered CsrA proteins is bound to sRNA.

- The short mRNA is not regulated by CsrA, and hence not part of the regulatory network. However, our experiments use the translation of short mRNA (specifically, the translation of the *cea* gene) as proxy for promoter activity in the S_{REP1}, S_{REP2}, C_{REP1}, and C_{REP2} strain. To enable the experimental validation of our model, we include the production of short mRNA in our model. Due to the lack of regulation, the corresponding differential equation is decoupled from M, A and S, and reads (with $\frac{\partial M_{short}}{\partial t} \equiv \partial_t M_{short}$)

$$\partial_t M_{short} = \alpha_{M_{short}} - \delta_{M_{short}} M_{short}, \tag{9}$$

where $\alpha_{M_{short}}$ and $\delta_{M_{short}}$ are the rate constants for production and per-capita degradation, respectively.

The properties and assumptions of the SOS response and the ColicinE2 regulatory system we listed above have already been used (if not stated otherwise) in the combined model for SOS response and ColicinE2 regulation presented in Ref. [9].

ssDNA as regulatory component: In the main text, we show experimentally that single-stranded DNA (ssDNA) serves as a component of post-transcriptional regulation of ColicinE2 production and release. Since this is a novel and, so far, an undocumented role of ssDNA, we briefly discuss how it acts as a regulator for CsrA in an *E. coli* cell.

ssDNA is an intermediate in the rolling circle replication mechanism of the ColicinE2 plasmid: The plasmid consists of double-stranded DNA (dsDNA). The first step in its replication is the production of a ring-shaped ssDNA transcript. These transcripts are produced both in absence and presence of an SOS signal (Fig. S5), which means that ssDNA production is constant. It is assumed that once a ring of ssDNA is completed, it detaches from the plasmid and diffuses freely through the cell. During this time, it is converted to double-stranded DNA, which eventually results in a new plasmid. Between the detachment of the single-stranded ring and the formation of a new plasmid, the ssDNA acts as a regulator of CsrA: Since the ssDNA includes the coding sequences present in the long mRNA, CsrA can bind to the Shine-Dalgarno sequence of the *cel* gene located on the ssDNA, and thus forms an ssDNA-CsrA complex. This allows the ssDNA to regulate free CsrA levels by sequestration, similar to the CsrA regulation by sRNA.

In our mathematical model, we account for these properties of ssDNA as follows:

- The production rate of ssDNA is assumed to be proportional to the number of ColicinE2 plasmids, P_{COL}, as it is an intermediate product of the rolling circle replication mechanism of the ColicinE2 plasmid. It reads

$$\alpha_D \cdot P_{COL}, \tag{10}$$

  with the per plasmid production rate constant α_D.

- The degradation of ssDNA is proportional to the ssDNA abundance, D, and thus reads

$$\delta_D \cdot D, \tag{11}$$

  with the per capita degradation rate constant is δ_D.

- The ssDNA has two binding sites for a CsrA dimer (see **Methods** of the main text), and thus can form a complex, C_{DA}, with it. Complex formation occurs with rate k_D^+, and the complexes dissociate into ssDNA and a CsrA dimer with rate k_D^-. Apart from disassembly, we also include the possibility that a CsrA-ssDNA-complex can spontaneously degrade (meaning that both CsrA and ssDNA are degraded at the same time) by introducing the per capita rate δ_{DA}. The abundance of complexes is denoted by C_{DA}.

Taken together, we can now formulate a set of differential equations, which allows us to quantify these interactions. These interactions are also illustrated as a biochemical network in Fig. S3.

We begin with the differential equation for the time evolution of the long mRNA, M. From the properties collected above, we conclude that this equation must contain three terms: The first term describes the production of long mRNA, which is proportional to the number of unrepressed promoters. This number is calculated from the difference between the total plasmid copy number, $P_{COL} + P_{REP}$, and the number of repressed promoters, B. The abundance of long mRNA is reduced by a second and a third term: The second term describes the spontaneous degradation of long mRNA, and is proportional to its abundance, M. The third term is bi-linear (that is, it is proportional to A and M) and represents the effective coupled degradation of long mRNA in complexes with CsrA. This term combines the binding of CsrA to long mRNA, the dissociation of this complex, and its degradation in an effective binding parameter k_M. The three terms read:

$$\partial_t M = \alpha_M (P_{COL} + P_{REP} - B) - \delta_M M - k_M M \cdot A. \tag{12}$$

The derivation of the effective coupled degradation in the third term is described in detail in Ref. [9]; an analogous derivation for the ssDNA is given below. The B in the first term is determined by the LexA/RecA subsystem of the SOS response, in particular by eq. (6).

The differential equation for the time evolution of the effective single-binding-site sRNA, S, consists of terms very similar to that for long mRNA. Two terms account for spontaneous and effective coupled degradation, respectively, and are structurally the same as in eq. (12). This is due to the fact that the sRNAs regulate CsrA in the same way as CsrA regulates the long mRNA, by forming complexes. The production term is, however, different, and contains two parts: The first part, $\alpha_{S,0}$, describes a constant baseline production, which ensures the production of sRNAs in the absence of CsrA. The second part depends linearly on the abundance of free CsrA,

and thus accounts for the positive regulatory function of CsrA for the sRNAs. Taken together, these four terms give the differential equation for S:

$$\partial_t S = \alpha_{S,0} N + \alpha_{S,c} N \cdot A - \delta_S S - k_S S \cdot A. \tag{13}$$

Having described the two partners of CsrA, we now turn to the differential equation for CsrA itself. Again, this equation has a very similar structure to eqs. (12) and (13): An, in this case constant, production term, as well as a term for spontaneous degradation. Here, however, we have more than one coupled degradation term, since CsrA can bind to more than one component: long mRNA (M), sRNAs (S), and ssDNA (D). The effective coupled degradation terms for long mRNA and sRNA are exactly the same as in eqs. (12) and (13), respectively. This reflects the fact that the formation of a long-mRNA/CsrA- or sRNA/CsrA-complex has for both complex partners the same consequence, that is, it reduces the abundance of free CsrA by 1. The coupled degradation part is also responsible for the hierarchical regulation, which we discussed in [9]: The actual regulation target, long mRNA (M), exclusively binds to CsrA; the sRNAs affects the free long mRNA level only indirectly by sequestering the CsrA and thus "regulating the regulator". Moreover, we also have to account for ssDNA/CsrA-complexes. Since we have not derived an effective coupled degradation for this complex yet, we explicitly account for its formation and disassembly. This means that we have to include two terms that account for the decrease of free CsrA due to the formation of ssDNA/CsrA-complexes and the increase of free CsrA when such a complex disassembles. Altogether, the differential equation for CsrA reads

$$\partial_t A = \alpha_A - \delta_A A - k_M M \cdot A - k_S A \cdot S - \textcolor{red}{k_D^+ D \cdot A + k_D^- C_{DA}}, \tag{14}$$

where C_{DA} is the abundance of ssDNA/CsrA-complexes, and $k_D^{\pm}$ the complex binding and disassembly rate, and the terms containing the novel regulator ssDNA are highlighted in red. Note that we consider the ssDNA to have only one binding site for CsrA in our model. We account for the second binding site analogously to the many binding sites of the sRNA, that is by assuming D to be an effective, single binding site ssDNA, with an effective production rate fitted to experimental data.

The two ssDNA terms highlighted in red also appear in the differential equation for D, since the formation and disassembly of ssDNA/CsrA-complexes in- and decreases also the abundances of ssDNA. The spontaneous degradation is accounted for by a separate degradation term, already

known from the differential equations of the other components. The production term of ssDNA is proportional to P_{COL}, the number of ColicinE2 plasmids in the cell, since ssDNA is an intermediate of the ColicinE2 plasmid replication. The differential equation for ssDNA therefore reads

$$\partial_t D = \alpha_D P_{COL} - \delta_D D - k_D^+ D \cdot A + k_D^- C_{DA}. \tag{15}$$

We are still left with the dynamics of the ssDNA-CsrA-complexes, C_{DA}. The "production" term of the ssDNA/CsrA-complexes is the binding term already known from eqs. (15) and (14), but in this case with a positive sign. The number of complexes is reduced by complex disassembly, which is accounted for by the term $k_D^- C_{DA}$ that also appears in eqs. (15) and (14) with a different sign. Apart from complex disassembly, the complexes can be degraded (in the sense that the complexes and their components are destroyed) spontaneously, which is given by a spontaneous degradation term. Taken together, the differential equation for ssDNA reads

$$\partial_t C_{DA} = k_D^+ D \cdot A - k_D^- C_{DA} - \delta_{C_{DA}} C_{DA}. \tag{16}$$

Effective coupled degradation of ssDNA and CsrA: In the discussion of the ssDNA properties, we saw that ssDNA also has a Shine-Dalgarno sequence, just as the long mRNA. This suggests that we can make the same assumptions for the CsrA-ssDNA-complex as we did for the CsrA-long-mRNA-complex. In the following, we proceed analogously to the simplification of the hierarchical three component model (see Ref. [9]), and assume fast dynamics of complexes. Adiabatic elimination ($\partial_t C_{DA} \equiv 0$) yields

$$C_{DA} = \frac{k_D^+ DA}{k_D^- + \delta_{C_{DA}}} = \frac{k_D DA}{\delta_{C_{DA}}}, \tag{17}$$

with the effective binding parameter

$$k_D := \frac{k_D^+ \delta_{C_{DA}}}{k_D^- + \delta_{C_{DA}}}.$$

By inserting eq. (17) into eqs. (14)-(16), we get our final set of differential equations, which includes all four components:

$$\partial_t M = \alpha_M (P_{COL} + P_{REP} - B) - \delta_M M - k_M M \cdot A, \tag{18}$$

$$\partial_t S = \alpha_{S,0} N + \alpha_{S,c} N \cdot A - \delta_S S - k_S S \cdot A, \tag{19}$$

$$\partial_t A = \alpha_A - \delta_A A - k_M M \cdot A - k_S S \cdot A - k_D D \cdot A \tag{20}$$

$$\partial_t D = \alpha_D P_{COL} - \delta_D D - k_D D \cdot A. \tag{21}$$

13

The new regulative component ssDNA acts in the same fashion as the sRNA by binding CsrA. Compared to the original three component system, eqs. (12)-(14), the extension with ssDNA therefore resulted in a system of equations with the same types of terms (source term, spontaneous degradation, coupled degradation). We use eqs. (18)-(21) to study gene expression dynamics for all three different strains. This is done by adjusting the corresponding values for P_{COL} and P_{REP}, see section 3. Moreover, we investigate the impact of ssDNA on the regulation of ColicinE2 production and release.

3 Parameter values

For the parameters associated with long mRNA, CsrA and the effective sRNA, we adjusted the values that we determined in our previous study ([9]) according to new measurements. In particular, they were chosen such that they are in accordance with our own experimental measurements (k_M and k_S) or other studies (see below). In particular, the rates read (given per *E. coli* cell volume, and using the shorthand notation "#" for molecule numbers):

rate const.	value	unit	description
α_M	0.05	min^{-1}	production of long mRNA
$\alpha_{S,0}N$	0.1	min^{-1}	baseline production of eff. sRNA
$\alpha_{S,c}N$	0.07	$min^{-1}\cdot\#^{-1}$	production factor of eff. sRNA
α_A	4.5	min^{-1}	production of CsrA
δ_M	0.04	$min^{-1}\cdot\#^{-1}$	degradation of long mRNA
δ_S	0.023	$min^{-1}\cdot\#^{-1}$	degradation of effective sRNA
δ_A	0.00007	$min^{-1}\cdot\#^{-1}$	degradation of CsrA
k_M	0.007	$min^{-1}\cdot\#^{-2}$	eff. binding of CsrA to long mRNA
k_S	0.011	$min^{-1}\cdot\#^{-2}$	eff. binding of CsrA to sRNA

The three degradation rates ($\delta_M, \delta_S, \delta_A$) were determined in previous, experimental studies [5, 7]. The production rates ($\alpha_A, \alpha_{S,c}, \alpha_M$) were fitted such that they reproduce component abundances from experimental studies [7]. The baseline production for the sRNAs, $\alpha_{S,0}N$, is set to a low value, as only few sRNAs are produced in the absence of CsrA [5].

In Ref. [9] we showed that a Poisson-distributed plasmid copy number gives very similar results to a fixed plasmid copy number. This is due to the fact that plasmid replication happens on larger timescales than the regulatory interactions considered in our model. We retain this simplifying

assumption, and set the number of ColicinE2 plasmids constant at $P_{COL} = 20$, which is the average value [8]. For the reporter plasmids in the C_{REP1}, C_{REP2}, S_{REP1} and S_{REP2} strains, we take for type 1 the average copy number $P_{REP} = 55$ [17], and for type 2 the average copy number $P_{REP} = 13$, which we both also assume constant.

Adding ssDNA dynamics to the system introduces three new effective rates, α_D, δ_D, and k_D. We assume that the ssDNA and the mRNA are equally stable, and therefore use the same degradation rate constants for both:

$$\delta_D \equiv \delta_M = 0.04 \text{ min}^{-1} \cdot \#^{-1}.$$

In combination with the K_D-value measurements for ssDNA we could determine the coupled degradation constant to

$$k_D = 0.0001 \text{ min}^{-1} \cdot \#^{-2}.$$

Finally, we have to define the value for the production rate constant of ssDNA, α_D, which has not been explicitly measured yet. However, our experimental data suggests that ssDNA accumulates abundances about an order of magnitude larger than long mRNA. From fitting the ssDNA production to this rough abundance relation, and also to measured delay-times, we obtain

$$\alpha_D = 7 \text{ min}^{-1} \cdot \text{plasmid}^{-1}.$$

To study the influence of ssDNA on the *cea-cel* delay, we varied the value of α_D between 0 and 9, see Fig. S6 and S7. For the validation of our model, we also tested various values of $\alpha_{S,c}$ and α_M (data not shown). These tests showed that, in general, the model is robust to parameter variations, in the sense that changing a parameter value by a few percent only had minor consequences for the resulting delay times and component abundances.

4   Simulation results

The differential equations eqs. (18)-(21) give, in combination with the SOS response model, eqs. (1)-(7) (see [2]), a description of the regulatory interactions governing ColicinE2 production. They enable us to study steady states and the deterministic dynamics of gene expression for all five strains. However, the SOS response [2] shows an inherent stochasticity: The ColicinE2 promoter is not activated permanently during an SOS signal,

15

but in stochastically appearing bursts of activity. Moreover, most of the regulatory components like long mRNA occur in low abundances, such that also intrinsic demographic fluctuations in the ColicinE2 regulatory system become important. We can study stochastic effects like these by formulating the deterministic dynamics described in eqs. (1)-(7) and eqs. (18)-(21) as a stochastic process. To this end, we consider each component (M, S, A and D, as well as the components of the SOS response system) as random variables that are changed by stochastic events like production or degradation of molecules. Each of these events occurs at an average rate that equals the corresponding term in the mass action model. For instance, the effective coupled degradation of a long mRNA and CsrA happens at a rate $k_M M \cdot A$ (see eq. (18)), which decreases both the abundance of long mRNA (M) and the abundance of CsrA (A) by 1. By defining all remaining stochastic production, degradation and binding events in the system this way, we obtain a description of the SOS response and ColicinE2 regulatory system as a stochastic (Markov) process. We then use the Gillespie algorithm [18] to implement the stochastic process as a stochastic simulation. This simulation enables us to produce stochastically correct realisations of the temporal evolution of the system's random variables. The results from sufficiently large ensembles of these realisations is then the basis for the validation of the theory by experimental data.

In our simulations, we followed the scheme already developed in Ref. [9]: We initiate the system in a non-SOS state, where the parameter c_p in eq. (6) of the SOS response system (see also Ref. [2]) is set to 0, that is, RecA does not cleave LexA. Therefore, B, the number of unrepressed promoters, is low (1 for S_{REP2}, 2 for C_{WT}, 3 for C_{REP2}, 4 for S_{REP1}, and 5 for C_{REP1}). After 200 minutes, we mimic the effect of an SOS signal by increasing the parameter c_p, such that RecA catalyses the degradation of LexA, which acts as repressor for the ColicinE2 operon. This has the effect that the production of both long and short mRNA immediately increases. The SOS signal is stopped again at $t = 500$ minutes. For each set of parameters, this scheme is repeated 2000 times in order to obtain an ensemble of 2000 realisations.

To be able to compare these simulation results with experiments, we have to give an appropriate definition of the *cea-cel* delay in the simulations. As the beginning of "*cea* expression", we define the point in time at which the short mRNA level rises to two times its value before the SOS signal started. Our simulations show that the CsrA abundance decreases during the SOS signal, as more CsrA-sequestering long mRNA is produced. Once there is no free CsrA left, we find free long mRNA in the system. We

16

define the first point in time at which more than 8 free long mRNAs exist in the system as "*cel* expression". This definition accounts for the fact that in general fewer lysis proteins are produced than toxin proteins [8]. In the experiment, the expression of *cea* and *cel* are defined by the point in time the respective fluorescence intensity reaches five times its basal (i.e. pre-SOS) level. Therefore, the expression times are determined by the appearance of proteins in the experiment, but by the appearance of mRNAs in the stochastic simulations. We choose the different definition of the delay in the simulations, as the specific biochemical rates of many processes involving the mRNAs and proteins are largely unknown, and have to be fitted according to observed abundances. If we included the translation of short and long mRNA to toxin, lysis and fluorescence proteins used in the experimental study into our model as well, we would add several new parameters that require fitting to our mathematical model, without getting a more precise definition of the thresholds that determine the delay. Moreover, comparing a delay in the production of mRNAs with a delay in the production of proteins is valid in our case, since both fluorescence proteins have very similar maturation times [6], and since we are interested in the relative rather than the absolute times of protein expression.

For the C_{WT} strain, we cannot compare the *cea-cel* delay-time from our simulations with experimental results due to the lack of reporter plasmids (see main text). To still be able to validate our stochastic simulations with experimental data in this case, we use the time between the beginning of the SOS response and cell lysis (referred to as "lysis time"), which we also measured in experiments (see Fig. S8). The absolute values of the lysis time will differ between our simulations and experiments, as the simulations do not account for maturation times and other processes of equal duration in all three strains. Therefore, we do not compare lysis times themselves, but the differences of the C_{REP1} and C_{WT} strain's lysis times, which eliminates these constant factors.

In the following, we discuss the results of our stochastic simulations for the experiments presented in the main text.

4.1   The role of CsrA

As a first step to validate the mathematical model for the S_{REP1} strain (that is, with no ssDNA in the system), we test the role of CsrA for the *cea-cel* delay. The corresponding experiment varied the binding affinity of CsrA to long mRNA, see Fig. 2E in the main text. Specifically, the experiment measured the average *cea-cel* delay-time for the original S_{REP1} strain, and

for two mutant strains (CsrA1 and CsrA2) with higher and lower CsrA binding affinity, respectively. The results of these experiments (see Fig. 2E) showed that an increased CsrA binding affinity leads to longer average *cea-cel* delay-times, since the regulation of long mRNA by CsrA sequestration happens more effectively with stronger CsrA binding. Consequently, we found a much shorter average *cea-cel* delay for the mutant with lower binding affinity. In our mathematical model defined by eqs. (18)-(21), the binding affinity of CsrA for long mRNA relates to the parameter k_M^+, which appears in the effective coupled degradation parameter, $k_M = \frac{k_M^+ \delta_{C_{MA}}}{k_M^- + \delta_{C_{MA}}}$ (see Ref. [9] for details on this equation). However, the experiments do not measure the rates k_M^+ and k_M^-, bute their ratio k_M^-/k_M^+, known as K_D value (see Fig. 2D of the main text). Since the degradation rate of the complexes is much lower than the dissociation rate (this follows from the fast complex equilibration assumption), we can determine k_M from the complex degradation and the K_D values:

$$k_M = \frac{k_M^+ \delta_{C_{MA}}}{k_M^- + \delta_{C_{MA}}} \approx \frac{k_M^+ \delta_{C_{MA}}}{k_M^-} = \frac{\delta_{C_{MA}}}{K_D} \tag{22}$$

We used eq. (22) and K_D values given in Fig. 2D to determine the binding parameter constants k_M for the S_{REP1} strain and its two mutants CsrA1 and CsrA2. Using these results, we then performed a set of numerical simulations, and recorded the *cea-cel* delay-times. The mean delay times for different values of k_M are shown in main text Fig. 2F, where we also give the parameter values. For the *cea-cel* delay-times we find the same behaviour as in the experiments: higher values of k_M result in broader delay time distributions with a larger average delay time, whereas smaller values give a narrow delay time distribution with an average delay close to the start of the SOS signal. Our results show that the binding of CsrA to long mRNA has a key influence on the delay time distribution. This highlights the critical role of CsrA for the delay between toxin production and release. Therefore, we expect regulative components affecting the abundance of CsrA to have an indirect effect on the duration of the *cea-cel* delay.

4.2 *cea-cel* delay in the five different strains

The experiments discussed in the main text show that the C_{REP1}, C_{REP2}, C_{WT}, S_{REP1} and S_{REP2} all have very different *cea-cel* delay times: The C_{REP1} strain, for instance, lyses almost immediately after *cea* expression, whereas the S_{REP1} strain shows a significant *cea-cel* delay. In this section, we employ

18

stochastic simulations of the model described in eqs. (18)-(21) to study the origin of this difference and the effects of ssDNA production. Moreover, we infer the *cea-cel* delay in the C_{WT} strain from this analysis, which cannot be measured directly in experiments. To this end, we modelled the five strains in our simulations by setting P_{COL} and P_{REP} to the corresponding values (see Table S5): $P_{COL} = 20$ and $P_{REP} = 0$ for C_{WT}, $P_{COL} = 20$ and $P_{REP} = 55$ for C_{REP1}, $P_{COL} = 20$ and $P_{REP} = 13$ for C_{REP2}, $P_{COL} = 0$ and $P_{REP} = 55$ for the S_{REP1} strain, and $P_{COL} = 0$ and $P_{REP} = 13$ for the S_{REP2} strain. For each strain, we simulated 2000 realisations for different ssDNA production rates (α_D) to investigate the role of this novel regulatory component. The results of these simulations are depicted in the form of *cea-cel* delay-time histograms in Fig. S6, in which we, for a clear and concise discussion, only depict the results for the C_{REP1}, S_{REP1}, and C_{WT} strains.

As we showed in section 1, only two factors can in principle be responsible for the different *cea-cel* delays: The total plasmid copy number, and the production of ssDNA. To separately study the influence of total plasmid copy number in the strains, we first analyse the case of no ssDNA production ($\alpha_D = 0$). The delay-time histograms of the three strains for this case are depicted in the first column of Fig. S6. Comparing the histograms, we find that the total plasmid copy number has a significant effect: The *cea-cel* delay distribution of the C_{REP1} strain (75 plasmids in total) is more skewed to the right compared to the distribution of the S_{REP1} strain (55 plasmids in total), and the C_{WT} strain (with only 20 plasmids) shows almost no lysis at all. The average delay-time of the S_{REP1} strain (68 minutes) is in good agreement with experimental values (see Fig. 2E). For the other strains, however, the results do not match: The C_{REP1} strain has a mean delay time of 24.1minutes, which is significantly larger than the value we find in our experiments. Our experiments also find that the wild-type indeed does lyse after SOS responses (see Fig. S8), while the histogram of C_{WT} predicts no lysis. These results for $\alpha_D = 0$ show that the plasmid copy number does not suffice to explain the quantitative and (for the wild-type) qualitative behaviour found in our experiments. However, it already accounts for significant differences in the *cea-cel* delay-time distributions between the strains.

Before we study the additional effects of ssDNA, we discuss the origin of these differences in the *cea-cel* delay-time distributions. To this end, we consider the time evolution of the average levels of free CsrA and long mRNA, which are depicted for $\alpha_D = 0$ in the first column of Fig. S7. We find for all three strains that, after the initial equilibration, the average number of CsrA molecules remains at a constant value before an SOS

19

response ($t < 200$ min), which results from the interactions of CsrA with all its binding partners. Before an SOS signal, the three strains exhibit roughly the same average CsrA level. This changes after the response to an SOS signal (200 min $< t < 500$ min), which reduces the average CsrA level in all three strains: The C_{REP1} strain, containing 75 plasmids, has the lowest CsrA levels, whereas the wild-type strain with only 20 plasmids has a significantly higher level.

Comparing the CsrA and long mRNA levels with the corresponding *cea-cel* delay-time distributions in Fig. S6 shows that the different *cea-cel* delay time distributions are correlated with the average levels of free CsrA (and free long mRNA): The lower the average CsrA level during a SOS signal, the shorter the average *cea-cel* delay-time, and the narrower the delay-time distribution. We can explain this correlation in our mathematical model by the fact that long mRNA production increases in form of stochastic bursts during SOS responses (see section 2). The long mRNAs produced during these bursts must first sequester free CsrA, before their abundance is high enough to produce lysis proteins from it. For the C_{WT} strain, the CsrA level during the SOS response is too high to be sequestered enough by stochastic long mRNA bursts (see Fig. S7 and Fig. S6). For the S_{REP1} and the C_{REP1} strain, however, CsrA levels reach a sufficiently low average abundance during an SOS response that stochastic bursts of free long mRNA are possible, and eventually lysis protein is produced. The average level of free CsrA therefore determines the probability and hence the timing of lysis.

In Fig. S7 we can also see that the single trajectories of CsrA abundance differ qualitatively between the three strains: The trajectories of the C_{REP1} strain show large and abrupt deviations from the mean value, whereas the abundance of CsrA is closer to the mean in the C_{WT} strain. The reason for this difference is the plasmid copy number in each strain: The more plasmids with LexA-regulated promoters, the more CsrA-sequestering elements are produced during an SOS response, and thus the more susceptible the system will be to stochastic bursts in the SOS response, increasing the probability of lysis. We already discussed in the previous paragraph that the number of plasmids also strongly affects the average level of free CsrA, as more plasmids cause a larger average number of promoters to be unrepressed. This effect is due to the fact that the repressor of the ColicinE2 operon, LexA, stochastically binds to and dissociates from the promoter, and thus triggers long mRNA production for short times even in the absence of an SOS signal. The more plasmids present, the larger the number of (transiently) derepressed promoters, and hence the more CsrA-sequestering long mRNA the cell contains. These relations explain

the differences in the *cea-cel* delay between the three strains.

Finally, in order to characterise the additional effect of ssDNA, we consider the plots with ssDNA production (that is, with $\alpha_D > 0$) in Fig. S6. As the ssDNA production rate α_D increases from 0, the average delay times in the C_{REP1} and C_{WT} strain decrease, and also several cells in the C_{WT} strain lyse (see the C_{WT} histogram for $\alpha_D = 1$ in Fig. S6). We attribute this to the fact that increasing α_D results in lower average CsrA levels in the C_{REP1} and C_{WT} strain, see Fig. S7. Consequently, the average *cea-cel* delay-times in Fig. S6 decrease, and lysis of C_{WT} bacteria becomes possible. The S_{REP1} strain, which contains no ssDNA-producing ColicinE2 plasmid, but only reporter plasmids, is not affected by the increase of this parameter. The experimentally observed difference in mean delay times of the C_{REP1} and S_{REP1} strain occur when the ssDNA production rate reaches $\alpha_D = 7$ (see Fig. S6). At this rate, also the C_{WT} shows a broad *cea-cel* delay distribution. For the C_{WT} strain, we cannot compare the average *cea-cel* delay-time from our simulations with experimental results, but have to use the lysis time. For $\alpha_D = 7$, this difference is in the same order of magnitude as the experimental results (see Fig. S8). If the ssDNA production rate becomes too high, large fractions of the cell ensemble lyse even in the absence of an SOS signal, which is not seen for the C_{REP1} strain in experiments.

Taken together, these results show that the additional sequestration of CsrA by ssDNA is required for cell lysis in the C_{WT} strain, and hence necessary to produce the experimentally observed *cea-cel* delays. Therefore, ssDNA plays a key role in the regulation of ColicinE2 release.

4.3 sRNA knock-out mutants

In the main text we also discuss experiments with different sRNA knock-out mutant strains, see Fig. S4. While the two single knock-out cases (no CsrB or no CsrC) are automatically accounted for by the effective sRNA (see the bullet points on sRNA in section 2), the special case of the double sRNA knock-out mutant corresponds to setting $\alpha_{S,0} \equiv \alpha_{S,c} \equiv 0$ in eq. (19) of our model. This means that no sRNA would be produced, which is the main CsrA-sequestering element in our mathematical model. Therefore, our model predicts a large abundance of free CsrA for the double sRNA knock-out case, and hence a significantly larger *cea-cel* delay. However, we do not see this behaviour in our experiments (see Figs. S2 and S4), which in contrast show shorter average *cea-cel* delay-times in the double knock-out mutant. These experimental results indicate that, in the absence of the two sRNAs, yet unknown regulatory mechanisms become

important. In the derivation of our mathematical model (see section 2), we eliminated any subordinate targets for CsrA, and focussed on the main CsrA-sequestering elements in *E. coli*, the sRNAs CsrB and CsrC [7]. Hence, adjusting our model for the double sRNA knock-out mutant would first require to experimentally investigate the detailed interactions and components of the yet unknown regulatory mechanisms, and then to replace the S component and its interactions correspondingly in the model. As the double knock-out mutant is not part of our investigation of the *cea-cel* delay-time in the main text, we do not further extend our model for this very special case. In all other strains and mutants discussed in the main text and the Supplementary Information, sRNAs are produced and also are the main CsrA regulator. Therefore, the aforementioned differences between theoretical model and experimental observations that arise with the double sRNA knock-out mutant do not affect any statements we derive for the single knock-out or original strains using our model, eqs. (18)-(21).

References

1. Janion, C. Inducible SOS Response System of DNA Repair and Mutagenesis in Escherichia coli. *Int. J. Biol. Sci.* **4,** 338–344 (2008).

2. Shimoni, Y., Altuvia, S., Margalit, H. & Biham, O. Stochastic Analysis of the SOS Response in Escherichia coli. *PLoS One* **4** (ed Isalan, M.) doi:`10.1371/journal.pone.0005363` (2009).

3. Yang, T. Y., Sung, Y. M., Lei, G. S., Romeo, T. & Chak, K. F. Post-transcriptional repression of the cel gene of the ColE7 operon by the RNA-binding protein CsrA of Escherichia coli. *Nucleic Acids Res.* **38,** 3936–3951 (2010).

4. Timmermans, J. & Van Melderen, L. Post-transcriptional global regulation by CsrA in bacteria. *Cell. Mol. Life Sci.* **67,** 2897–2908 (2010).

5. Weilbacher, T. *et al.* A novel sRNA component of the carbon storage regulatory system of Escherichia coli. *Mol. Microbiol.* **48,** 657–670 (2003).

6. Mader, A. *et al.* Amount of Colicin Release in Escherichia coli Is Regulated by Lysis Gene Expression of the Colicin E2 Operon. *PLoS One* **10.** doi:`10.1371/journal.pone.0119124` (2015).

7. Gudapaty, S. *et al.* Regulatory Interactions of Csr Components : the RNA Binding Protein CsrA Activates csrB Transcription in Escherichia coli Regulatory Interactions of Csr Components : the RNA Binding Protein CsrA Activates csrB Transcription in Escherichia coli. *J. Bacteriol.* doi:`10.1128/JB.183.20.6017` (2001).

8. Cascales, E. *et al.* Colicin biology. *Microbiol. Mol. Biol. Rev.* **71,** 158–229 (2007).

9. Lechner, M., Schwarz, M., Opitz, M. & Frey, E. Hierarchical Post-transcriptional Regulation of Colicin E2 Expression in Escherichia Coli. *PLoS Comput. Biol.* doi:`10.1371/journal.pcbi.1005243` (2016).

10. Ronen, M., Rosenberg, R., Shraiman, B. I. & Alon, U. Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 10555–10560 (2002).

11. Krishna, S., Maslov, S. & Sneppen, K. UV-Induced Mutagenesis in Escherichia coli SOS Response: A Quantitative Model. *PLoS Comput. Biol.* **3.** doi:`10.1371/journal.pcbi.0030041`. arXiv: 0701013 [q-bio] (2007).

12. Legewie, S., Dienst, D., Wilde, A., Herzel, H. & Axmann, I. M. Small RNAs establish delays and temporal thresholds in gene expression. *Biophys. J.* **95,** 3232–3238 (2008).

13. Levine, E. & Hwa, T. Small RNAs establish gene expression thresholds. *Curr. Opin. Microbiol.* **11,** 574–579 (2008).

14. Edwards, A. N. *et al.* Circuitry Linking the Csr and Stringent Response Global Regulatory Systems. *Mol Microbiol* **80,** 1561–1580 (2011).

15. Yakhnin, H. *et al.* Complex regulation of the global regulatory gene csrA: CsrA- mediated translational repression, transcription from five promoters by Eσ70 and EσS , and indirect transcriptional activation by CsrA. *Mol. Microbiol.* **81,** 689–704 (2012).

16. Taniguchi, Y. *et al.* Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science (80-.).* **329.** doi:`10.1126/science.1188308` (2010).

17. Megerle, J. A., Fritz, G., Gerland, U., Jung, K. & Ra, J. O. Timing and Dynamics of Single Cell Gene Expression in the Arabinose Utilization System. *Biophys. J.* **95.** doi:`10.1529/biophysj.107.127191` (2008).

18. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81,** 2340–2361 (1977).

3

Optimal Time Distributions for Lysis-based Toxin Release

Project Abstract and Contributions   This follow-up project on the ColicinE2 system investigates, which toxin release distribution is optimal to fend off competitors, and how this result is determined by the parameters of the system. In chapter 2, we found that changes in parameters of our model for ColicinE2 regulation alter the resulting lysis time distribution. As the toxins are released to defend their producers against competitors, this raises the question as to which distribution is best in killing other bacteria. To identify the factors affecting the distribution, I created a conceptual model for the self-destructive toxin release of bacteria (with ERWIN FREY). Using this model in combination with a genetic algorithm (see 3.1.2), we determined optimal lysis time distributions for given parameter sets, and analysed our results in context of phenotypic heterogeneity (see 3.1.1). The optimal distributions were then successfully put to the test on a stochastic lattice-gas model. The detailed development of our model and the full presentation of the results are given in the paper draft "Optimal Time-Distributions for Lysis-based Toxin Release", which is reprinted in section 3.5.

3.1 Background

3.1.1 Phenotypic Heterogeneity

As the title of this thesis suggests, all of the projects discussed here deal with heterogeneous bacterial systems. The classic source of heterogeneity in bacterial populations are different strains, or mutants of the same strain. These two examples have in common that their differences manifest in their *genome*: A deletion mutant, for example, does not produce the deleted protein in a given situation, because it lacks the necessary genetic information. Apart from this heterogeneity in the genotype, there exists a second form: phenotypic heterogeneity [56–58]. The central difference to its genotypic counterpart is that a different behaviour in the population is not due to variations in the genome, but occurs among genetically identical individuals. The origin of these phenotypic variations can be manifold, and often stems from fluctuations in gene expression [3, 59] or other stochastic effects [60].

A prominent example are persister cells [61, 62]: In some species of bacteria, a small fraction of cells (the persisters) permanently lives in a dormant state, and as such grows much slower than other cells in their population. This behaviour seems odd from an evolutionary perspective, as it means extremely slow reproduction. In high stress environments, however, this dormant state becomes an advantage, and only the persisters survive, eventually starting a new population after the stress. This form of phenotypic heterogeneity is often referred to as bet hedging *bet hedging* [56, 63, 64], as the population maintains diversity in order to have at least some survivors in severe environmental changes.

Another common example for phenotypic heterogeneity are division of labour strategies [65, 66]: Some bacteria in a population specialise, for instance, on producing one of many necessary metabolites, which they secrete and thus provide as public good to their population. These examples show that phenotypic heterogeneity is an important factor that increases the fitness of a bacterial strain in a variety of environments.

3.1.2 Genetic Algorithms

As already pointed out in 1.1.1, natural selection comprises three principles: variation, selection, and inheritance. These principles cause the stepwise adaptation of organisms to their environment, or, in other words, their optimisation for given conditions. The fact that this optimisation mechanism works in the complex and multifactorial environments nature poses, motivated researchers to transfer the idea of natural selection to general optimisation problems, for which the variables to mutate and inherit are not genes on DNA. Currently, there exist many applications for these evolutionary or *genetic algorithms* [67], ranging from biology [68] to engineering [69]. We briefly explain the core concepts with the example of the well-known travelling salesman problem (TSP) [67, 70].

In this paradigmatic problem, we consider a salesman who needs to visit all cities in his district; the goal is to find the shortest route for his trip. For the application of a genetic algorithm it is necessary to first find a genetic representation of the possible solutions in form of an array. In our example, this array is filled with the cities in the order the salesman visits them (first entry: first city to visit, second entry: second city to visit, etc.). In other words,

the array represents the route of the salesman. Furthermore, it is necessary to define a fitness function, which allows us to evaluate different routes and compare them. For the TSP, this is the total distance of the trip. We can now generate an initial population of, in our case, random routes. For each population member, we then determine the fitness (here: total distance), according to which the most successful population subset is selected. Finally, the population of the next generation is created from the members of this subset, after genetic operations like mutations or crossovers were applied. In our example, this would correspond to, for instance, flipping the position of random cities in the route. This scheme is then repeated for a large number of generations, thereby optimising the result according to the fitness function.

It is important to note that genetic algorithms are – like their natural counterpart – heuristic methods. This means that the result of an evolutionary optimisation algorithm is only "optimal" in the sense that it is better than the variants tested during its execution. While mutations can indeed explore large parameter spaces, they can get trapped in local minima. This problem can be reduced by, for instance, repeated optimisation runs.

3.2 Motivation and Research Question

In the previous study on the regulatory network of ColicinE2 production and release, we found that a broad distribution of lysis times emerged due to the stochastic peaks of the SOS response and stochastic abundances in the ColicinE2 system. Moreover, this distribution was positively skewed (that is, skewed to the right).

Our parameter studies also showed that varying rates like, for instance, the sRNA production within reasonable limits may result in different distributions. In natural systems, such variations might occur as consequence of mutations, and are thus subject to evolution. This suggests that the observed lysis distribution might be the evolutionary optimal solution for a given condition. However, it remains unclear if such an evolutionary optimisation process would yield the observed distribution, and what parameters it should optimise for, leading to the question:

Which lysis time distribution is optimal, and what parameters shape this result?

3.3 Summary of Results

Conceptual single compartment model and key parameters   We consider a single compartment of producers at the interface to an invading non-producer colony. The producers defend themselves with toxins that require self-destruction for their release. For this setup, we derived a conceptual model comprising only toxin and non-producer abundance. To achieve this reduced form, we argued for a simultaneous switch to the SOS state and consolidated several processes into effective ones. The population dynamics are formulated as rate equations, whereas the toxin increase is modelled explicitly by stochastic lysis events. These events followed a lysis time distribution, which is given as a parameter to the system. The time evolution of the system can be solved numerically using a stepwise integration scheme. We then non-dimensionalised the system; this allowed us identify the toxin degradation and

the product of toxin binding with the toxin production as the key parameters of the model's dynamics. Furthermore, the time scale was defined by the non-producer production rate (generation time).

Optimal distributions are broad and have positive skewness   For nine exemplary parameter sets we determined the optimal distributions using a genetic optimisation algorithm in combination with the aforementioned stepwise integration scheme. We used the cumulative abundance of non-producers as the fitness function. While the resulting distributions showed a strong dependence on the parameters, their general shape was broad and positively skewed (that is, more weight at short lysis times). For low toxin degradation and high toxin effectiveness, the resulting distributions showed a pronounced peak at short lysis times. The higher the toxin degradation and the lower its effectiveness, the more weight of the distribution is shifted towards larger lysis times, at the expense of the pronounced peak. We analysed the corresponding toxin production, and showed that all distributions quickly established a significant and rather constant toxin level in the population. Moreover, we could show that reducing in the initial producer number as well as increasing the maximal time of toxin production also cause a shift towards larger lysis times at the expense of shorter ones.

Conceptual compartment model correctly predicts spatial results   We created a detailed stochastic lattice-gas model mimicking two colliding bacterial fronts to test the predictions of our simple single compartment model. To this end, we considered a lattice of coupled compartments, and explicitly accounted for the birth, death and interaction rates of all components. In the simulations, one of the two competing strains is able to release toxins self-destructively according to a given lysis distribution, whereas the other exhibits an increased growth rate. We picked two exemplary parameter sets, which represent easy and hard conditions for toxins (low degradation/high effectiveness and high degradation/low effectiveness, respectively), and determined the optimal distributions for each case. The two resulting optimal distributions were then both simulated in both conditions, and we tracked the front line positions in each of the four cases. Our results show that the single compartment model is indeed capable of producing distributions that outperform less adapted distributions in realistic scenarios.

3.4 Conclusion

In this project, we continued our work on the ColicinE2 system, and focussed on the broad lysis time distributions created by the stochastic SOS response. Both the experiments as well as our correspondingly parametrized ColicinE2 model exhibited broad lysis time distributions with positive skewness. As it is the purpose of toxins to kill other bacteria, we were interested in the question: **Which lysis time distribution is optimal for this purpose, and what parameters shape this result?**

To this end, we considered a small compartment at the front of a producer population, and developed a conceptual model for this compartment being invaded by toxin-sensitive bacteria. Non-dimensionalisation of the model revealed that the key parameters of this model are the toxin degradation and the product of toxin production multiplied by the binding to non-producers. We then employed a genetic algorithm to determine which distribution is most effective at killing the non-producers for different combinations of the key parameters. Our results indicated that **broad lysis time distributions with positive skewness are optimal**. This agrees with the general shape we found in natural systems [39], suggesting that the distributions seen in experiments are the product of natural selection. Moreover, we found for all investigated parameter regimes that the optimal distributions create and maintain a relatively constant level of toxins in the compartment. We therefore conclude that it is more efficient to attack the invaders steadily over a prolonged time, than, for instance, releasing a large amount of toxins in a concerted action.

Our parameter studies also showed that **the specific shape of the distribution is governed by the toxin degradation and the product of toxin production multiplied by the binding to non-producers.** More specifically, in conditions with high degradation and low production/binding rates, the distributions exhibit a great number of large lysis times, and consequently only a low positive skewness. This is due to the aforementioned rates, which cause the toxins to quickly disappear from the compartment and to only weakly interact with the invaders. As the toxins accumulate in the cells, it becomes favourable for producers to release late, that is, at larger numbers. Consequently, as we decrease toxin degradation and increase toxin production/binding, more and more late lysis time points get shifted to the left, and build a prominent peak at short lysis times. In another parameter study, we reduced the number of producers in the compartment, which "forces" the population to economize the lysis events. In the resulting distributions, only the number of short lysis times decreased, indicating that the cells lysing at later time points have the greatest share at killing invaders. From this result we conclude that the broad distributions can be understood as a form of "division of labour": Early lysing cells keep the invading non-producers low, such that the highly effective late-lysing cells can kill them off more easily.

In order to test the predictions of our single compartment model, we developed a detailed spatial model of two colliding bacterial colonies. Because of computational constraints, however, it was not possible to implement the evolutionary algorithm for the spatial model. To test the single compartment model nevertheless, we took optimal distributions from selected parameters in the parameter scan. Then, using the same parameters and the spatial model, we simulated the collision of a producer colony employing this distribution with a colony of non-producers. Repeating these simulations with different, non-optimised distributions

and comparing the results showed that the single compartment model indeed predicted the most successful distribution also in the spatial case. However, we expect that the single compartment model is only able to make qualitative statements for optimal distributions in the spatial case. The reason for this originates in the new role of toxin diffusion in the spatial model: Apart from just removing toxins from the compartments (which is captured by the toxin degradation in the single compartment model), the toxins spread out to other compartments, in which they either support toxin production or even create a toxin presence before SOS signals are received. These effects rely on spatial degrees of freedom, and thus cannot readily be accounted for by the single compartment model. Investigating this aspect of diffusion and how it affects the optimal distributions for the spatial model poses an interesting challenge for follow-up research in this project.

In summary, our single compartment model gave insights into basic principles and mechanisms that shape lysis time distributions. While our model was developed with the ColicinE2 in mind, only three key characteristics of this system actually enter our model: The accumulation of toxin in the cells, the broadly distributed release times, and the "fatality" of the release events. We therefore expect that the rationales we found in this project can be transferred to other cases of self-destructive substance release such as public goods [59].

3.5 Publication Draft Reprint

Optimal Time-Distributions for Lysis-based Toxin Release

by

M. Lechner,[1] and E. Frey[1]

[1]Department of Physics, Arnold Sommerfeld Center for Theoretical Physics and Center for NanoScience, Ludwig-Maximilians-Universität München, Theresienstraße 37, 80333 München, Germany

Optimal Time-Distributions for Lysis-based Toxin Release

MATTHIAS LECHNER AND ERWIN FREY

1 Introduction

The various forms of bacterial behaviour are heavily shaped by evolution in hostile environments. During their constant battle for resources like space or nutrients, bacteria developed a wide range of means to gain advantages over competitors. A well-studied and potent example are bacteriocins. These toxins are released by bacteria and diffuse into their environment, where they then kill or inhibit competing bacterial strains. The toxin producing strain is immune to the toxin. Complex regulatory networks control the toxin synthesis, and ensure a timely production, for instance, after DNA damage.

The release of toxins to the environment typically happens via suitable structures in the cell wall. However, for some toxins, such structures do not exist, and the cells can only secrete the toxins by lysis. Examples for this fatal release mechanism are pneumolysin produced in *S. pneumoniae* [1], virulence factors in *S. typhimurium* [2] and colicin E2 in *E.coli* [3]. As the self-destructive toxin secretion always means that the cell "sacrifices" its life to the "benefit" of its population, this mechanism also has been discussed in the context of "cooperation" [2]. Apart from the ensuing cell death, this toxin release mechanism differs from the non-lethal transport through cell walls in another, very significant point: A bacterium can release toxin only once, at a single time point. Toxins are thus not released continuously, but accumulate in the cell and are released only simultaneously during lysis. Therefore, the later the lysis happens, the more toxin will be freed to the environment.

From the perspective of a bacterium, this limitation to a single release time point poses a dilemma: A short lysis time results in a fast reaction against competitors, but also in lower and possibly ineffective toxin doses. Larger lysis times, on the other hand, release large amounts of toxins, but

might come too late. For a cell, the ideal reaction to a given competitive situation is in general not foreseeable, and once lysed, it cannot correct its choice any more. A possible solution to this dilemma is heterogeneous lysis in a population of toxin producing bacteria. This phenomenon indeed occurs in some of the aforementioned systems: *S. typhimurium*, for instance, exhibits noisy self-destructive toxin release, which might play a crucial role in bacterial pathogenesis [2]. Another example is the production and release of colicin E2 [3, 4, 5], which also exhibits stochastic expression in fluctuating conditions [6]. In this system, heterogeneity can also appear in the lysis itself, as the production of toxin and lysis proteins is triggered by a noisy SOS response system. The regulatory networks of the two proteins, in combination with the noise in the SOS response, then result in a broad distribution of lysis times.

In two recent studies, we experimentally measured this distribution [7] and presented a stochastic model for the regulatory components [8]. Our analysis of the stochastic model showed that variations in the production or degradation rates of the regulatory components may result in distributions of very different shape. Consequently, mutations in the regulatory system could present a means to alter the lysis time distributions of bacteria, and thus pose an evolutionary mechanism to optimize the distribution for typical competitive scenarios.

We are interested in understanding which lysis time distribution is best in killing off competitor strains, and how this distribution depends on the system parameters. Moreover, we would like to develop a simple method to determine the optimal distribution for given set of parameters. To this end, we present here a simple, conceptual model for a compartment at the boundary of a bacterial population that is invaded by competitors. We analyse the model to determine the key parameters for defence with toxins, and present an evolutionary algorithm that produces the optimal lysis time distribution for a given set of parameters. Finally, we create a detailed lattice-gas model of two bacterial fronts growing towards each other, and use it to test the results of our single compartment model.

2 Models & Results

In the following, we present a conceptual model, which allows us to study the impact of different self-destructive toxin-release time distributions in a reduced, simplified setting. To this end, we specifically focus on a situation that puts cells into an SOS state, namely the stress resulting from a collision of two bacterial colonies. Our goal is to identify which properties of the bacteria and of the environment are important, and how these properties shape the optimal distribution. Moreover, we aim to give at least qualitative predictions on the shape of the ideal distribution for a given condition.

We consider two bacterial strains, of which one (the *producer*) produces a toxin that is released via lysis. This toxin is lethal to bacteria of the other strain (called the *non-producer*, or *sensitive* strain), but not for producer individuals. The model is motivated by the ColicinE2 system [8], however, its results can readily be transferred to other self-destructive toxin release systems.

Each of the two strains forms a large bacterial colony, which grows towards the other (see Fig. 1A). When the two colonies eventually meet, the first interaction between the two strains will happen at their boundaries. To study this interaction, we do not consider the whole colony in our model, but only a compartment at the expanding producer front.

2.1   single compartment model

The size of the compartment (in terms of cell numbers) is chosen such that its extension is small enough to be considered well-mixed (i.e. all molecules and cells interact equally likely with each other), yet large enough such that fluctuations do not matter. These choices justify formulating the population dynamics in the compartment as rate equations.

As the two colonies meet, sensitive bacteria are pushed into this compartment and start to reproduce (see Fig. 1B). We assume that non-producer cells invade the compartment at a constant rate ι_N, and reproduce with rate $\beta_N N$, where β_N is their per-capita growth rate and N the non-producer abundance in the compartment. Similar to reproduction, also the spontaneous degradation of non-producers, $\delta_N N$, as well as their flux out of the compartment, $\omega_N N$, are described in form of rates proportional to N (but appear with a negative sign in the equation, as they reduce N). However, for the situation of colliding fronts of growing bacteria, these decreasing processes have significantly lower rate constants than the reproduction rate (see, for instance, [8]). To obtain a concise description of the

A

B

α_N

ι_N

k_{NT}

ϵ_T
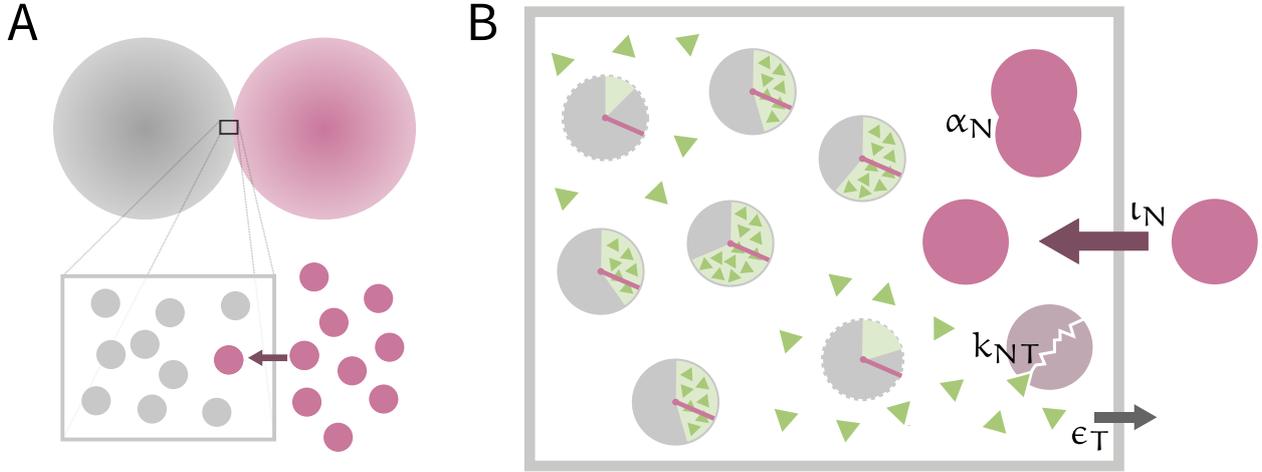
Figure 1: The single compartment model. A: We consider a compartment (grey box) at the boundary of the producer colony (grey circles), where non-producers (purple) are pushed into it. B: Our model accounts for the processes inside the compartment. Producer cells (grey circles) synthesise toxins (green triangles), and release them each at a specific time (green-shaded segment). The toxins can diffuse out of the compartment, and also kill non-producers (purple circles). In contrast to the producers, the non-producers reproduce, and are also pushed into the compartment.

system combine these processes to a effective reproduction rate constant, $\mu_N = \beta_N - \delta_N - \omega_N$.

   In natural systems, toxin production is often triggered by a SOS response, which itself is activated by external stress signals [9]. We choose reaching a critical non-producer abundance, N_{crit}, as such a stress signal, and refer to the point in time at which the cells switch to SOS state as t_{SOS}. The relation of $N_{crit} \equiv N(t_{SOS})$ to the compartment size $P(t_{SOS})$ quantifies the stress: the smaller $N(t_{SOS})$, the larger the competitive stress created by the invaders. Because we consider a well-mixed system, we assume that *all* producer cells in the compartment immediately and simultaneously switch to the SOS state: The producers stop growing and start producing toxins within their cells [4, 10], while the non-producers continue growing and invading the compartment. The toxins accumulate within the cells, and are only released to the compartment when a producer cell lyses (dies) [3].

   The specific time at which a producer lyses is drawn from a common lysis time distribution, which is our major study subject. A lysis event increases the number of (released) toxins in the compartment, T, by an amount $\mu_T(t - t_{SOS})$. This function μ_T is monotonically increasing with time due to the accumulation of toxin in cells, and depends on the time

since the cell entered the SOS state, $t - t_{SOS}$, as t_{SOS} marks the (common) starting point of toxin production.

The toxin, however, cannot be produced arbitrarily long: Cells only switch to the SOS state when they are severely damaged, and can only survive this damage for a limited amount of time [7]. To account for this, the lysis time distribution is bounded by a maximal time of toxin production, t_{max}. Its specific value has to be chosen in relation to other parameter values: From experiments we know, that cells in the SOS state maximally survive about 10 generations [7].

As the compartment is the only part of the colony in contact with non-producers, we assume that no toxins flow into it from the outside, and hence that producer cell lysis is the only process increasing toxin abundance. However, toxins can diffuse out of the compartment at rate $\omega_T T$, and also spontaneously degrade at rate $\delta_T T$. For a compact description, we proceed analogously to the non-producer abundance and join the rates for these two processes into an effective per-capita degradation rate constant $\epsilon_T = \omega_T + \delta_T$, resulting in the effective degradation rate, $\epsilon_T T$. Apart from being released into the compartment or being degraded, the toxin molecules can also hit and kill non-producer cells. We assume that this process does not change the toxin abundance, but decreases N with a rate $k_{NT} \cdot N \cdot T$, where k_{NT} is the toxin binding constant.

Finally, we take the aforementioned terms together to formulate a system of rate equations. As the initial time point of our model, $t_0 \equiv 0$, we choose the simultaneous switching to the SOS state, $t_0 := t_{SOS}$, which marks the beginning of toxin production. This choice has the advantage that the dynamics of the producer abundance, P, must not be explicitly accounted for in our model: From t_{SOS} on, the producers do not reproduce, and thus only the $P(t_{SOS}) := P_0$ cells already present can add toxin to the compartment. We are left with only two dynamic variables describing the system: N and T. Collecting all terms that affect the non-producer abundance, we get a rate equation for N,

$$\frac{dN}{dt} = \iota_N + \mu_N \cdot N - k_{NT} \cdot N \cdot T. \tag{1}$$

For the toxin abundance, we have to proceed differently, as we can formulate only the toxin degradation as a rate equation:

$$\frac{dT}{dt} = -\epsilon_T T. \tag{2}$$

The toxin release by the P_0 producer cells happens according to a stochastic (point) process Δ_t of exactly P_0 lysis time points τ_l, which are drawn from
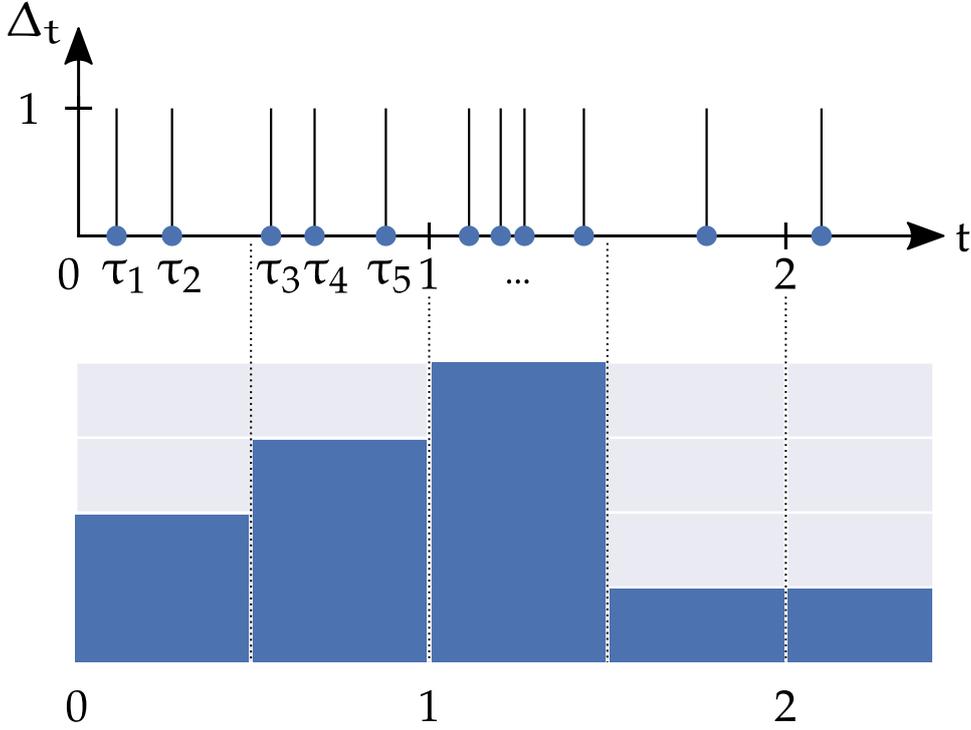
Figure 2: Illustration of the stochastic process and the underlying distribution. The point process Δ_t distributes P_0 lysis time points τ_l on the interval $[0, t_{max}]$. We visualise such distributions using histograms (lower part).

a specific lysis time distribution on the interval $[0, t_{max}]$. At each lysis time point τ_l, the time-dependent increment reads $\mu_T(\tau_l)$. Since it is the lysis time distribution that we want to study here, we consider Δ_t explicitly.

Before we discuss how to analyse this set of equations, we first want to identify the key factors that influence the dynamics of the system. To this end, we non-dimensionalise the equations by rescaling the variables:

$$t' = \mu_N\, t, \qquad\qquad N' = \frac{\iota_N}{\mu_N}\, N, \qquad\qquad T' = \frac{\mu_N}{k_{NT}}\, T, \qquad (3)$$

and also $\alpha'_T(t') = \frac{\mu_T(t')}{\mu_N}$ and $\epsilon'_T = \frac{\epsilon_T}{\mu_N}$. We can then set the timescale to the reproduction of non-producers, $1 \equiv \frac{1}{\mu_N}$. This simplifies eqs. (1) and (2) to the following form:

$$\frac{dN'}{dt'} = 1 + N' - N' \cdot T', \qquad (4)$$

$$\frac{dT'}{dt'} = -\epsilon'_T T'. \qquad (5)$$

The increments $\mu_T(t)$ of the stochastic process Δ_t have to be rescaled in the

same way as eq. (2), and thus reads

$$k_{NT}\mu_T'(t'). \tag{6}$$

For a clear and concise description, we will only use rescaled quantities in the following, and drop the prime in the notation.

With these non-dimensionalised equations (and increment) at hand, we can now investigate the effects of different lysis time distributions. To this end, we have to define a criterion by which we can compare different distributions with each other. In the situation discussed here, the producers are threatened by non-producers invading their habitat and possibly outgrowing them. In our model, both reproduction and degradation of the non-producers are proportional to the non-producer abundance. For producers, it is hence favourable to reduce this abundance of non-producers as much and as soon as possible, to minimize the penetration of non-producers into the producer colony. Consequently, a distribution which results in a lower N over time is better than one which causes larger values. We therefore choose the cumulative non-producer abundance,

$$\int_{t_0}^{t_{max}} N \, dt, \tag{7}$$

as comparison criterion, where $t_{max} < \infty$ is again the latest time point at which a cell can release toxin (see above) . We consider a distribution as "better" than another one if it results in a lower cumulative non-producer abundance. Hence, to compare two distributions with each other, we have to calculate the cumulative non-producer abundance for each distribution, and then compare the result.

However, as the toxin abundance increments as a stochastic process, $k_{NT}\mu_T(t) \cdot \Delta_t$, we cannot simply solve the coupled system of equations (4) and (5) numerically, which would be necessary to obtain the cumulative non-producer abundance, eq. (7). To find a solution nevertheless, we consider the properties of the specific stochastic process (see also Fig. 2): Δ_t is a point process of exactly P_0 discrete, independently and identically distributed lysis time points τ_l at which the toxin abundance increases. If we now want to compare arbitrary distributions for the τ_l, it is hard to find a general solution of eqs. (4) and (5), as these distributions could in principle not even be described in a closed functional form like, for instance Gamma distributions. For a given empirical distribution (that is, $\{\tau_l\}$, a given set of lysis time points), however, we can calculate the cumulated non-producer abundance with the following scheme: First, we numerically solve the system of rate equations (4) and (5) (without the stochastic term)

between t_0 and the first sample in the point process, τ_1. Then, the number of toxins is increased by the increment $k_{NT}\mu_T(\tau_1)$, and we continue by numerically solving the system up to τ_2. This procedure is repeated for all the P_0 lysis times τ_l up to t_{max}. From the resulting solutions for N we can then obtain the cumulated non-producer abundance.

The piecewise integration procedure not only enables us to compare different distributions, but is also the basis to determine the optimal distribution for a given set of parameters. To this end, we employ an evolutionary algorithm, in which an ensemble of compartments (each with a different lysis time distribution) undergoes repeated mutation, evolution, and selection. More specifically, the algorithm comprises the following steps:

1. Initialize each compartment of the ensemble with an initial distribution (here: all lyse at $t = t_{max}/2$)

2. For all ensemble compartments (but one), mutate the distribution according to a mutation rate m: From the P_0 lysis times in the distribution, select m random ones, and add a random shift value, drawn from a normal distribution, to them.

3. Calculate the cumulative non-producer abundance for each compartment in the ensemble with the procedure described above.

4. Take the results of all compartments in the ensemble, and select the two distributions with the lowest cumulative non-producer abundance.

5. For one half of the ensemble, replace each distribution with the best distribution, for the other half the second best. This reduces the chance to get stuck in a local optimum.

6. Go back to step 1. The best and second best distributions are now the new initial distributions.

A more detailed description of the algorithm and its concrete implementation is given in the section "Detailed Evolutionary Algorithm" of the Supporting Information.

2.2 results and discussion of single compartment model

Before we employ the evolutionary algorithm to determine the optimal distribution, we first discuss the model parameters. For the increment function $\mu_T(t)$, we assume a simple linear function, $\mu_T(t) = \alpha_T \cdot t$, where

α_T is a proportionality constant. With this definition, we immediately see from eqs. (4), (5) and (6) that ϵ_T and $k_{NT}\alpha_T$ are the key parameters of the system, which we will investigate in the following. Moreover, we set t_{max}, as discussed above, to 10 generations of N, a value motivated from experimental studies [7]. To meet the conditions on the compartment size explained in 2.1, we set $P_0 = 500$; the non-producers are assumed to create a high stress level, and trigger the SOS signal already with $N_0 = 100$ individuals in the compartment. The optimal distributions for nine exemplary parameter sets are depicted as (blue) histograms in Fig. 3. These results also enable us to analyse the dependence of the optimal distribution on ϵ_T and $k_{NT}\alpha_T$.

For all nine cases, the optimal distribution is rather broad, meaning that the lysis times are not limited to a narrow regime. Moreover, all distributions show a positive skewness. In other words, optimal distributions span a wide range of lysis times, but earlier ones have a larger weight than later ones. This general shape of the distributions has interesting consequences for the toxin abundance (see orange lines in Fig. 3): it builds up early and then maintains a relatively constant level. In particular, we find no pronounced peaks in toxin abundance; the noisiness of the toxin levels in Fig. 3 is due to the toxin not being released continuously, but at discrete lysis events.

The form of the optimal distributions therefore shows that to efficiently reduce the non-producer population, it is necessary to have a significant level of toxins in the compartment *over an extended time span* (rather than, for instance, a concerted burst of toxins at a specific point in time). Consequently, it is necessary to build up the significant toxin level fast, since the non-producer death rate is proportional to the toxin abundance. However, the faster a cell lyses, the fewer toxin it can release, as toxins need to accumulate in the cells. To reach the same level of toxins, thus more fast-lysing cells are needed than cells with larger lysis times. A distribution that produces a constant significant toxin level must therefore have a positive skewness (in other words: more weight at short lysis times). To ensure that this level is maintained while the toxins are constantly removed from the compartment, lysis events must also occur over a wide span of lysis times, which is facilitated by the broadness of the optimal distributions.

While we observe the positive skewness for all investigated parameter combinations, we also find that the specific shape of the distribution depends on the toxin degradation ϵ_T and toxin effectiveness $k_{NT}\alpha_T$. The larger ϵ_T, the more weight of the distribution is at higher lysis times (see Fig. 3). In these cases of large ϵ_T, the quickly released low toxin doses

hardly build up a significant level, as the toxins quickly leave the compartment. Therefore, delaying the creation of a specific toxin level to later time points at which more toxin has been produced in the cells is beneficial for producers. For small values of ϵ_T, the situation is different: already low levels of toxin pay off for the producers, as the toxins then remain longer in the compartment.

For increasing $k_{NT}\alpha_T$, we find that the weight of the distributions shifts to shorter lysis times. The larger toxin binding and production rates increase the chance of toxins killing non-producers, rendering an earlier lysis beneficial. In particular, very high values of $k_{NT}\alpha_T$ (see last row in Fig. 3) result in an extremely pronounced peak at very low lysis times. In these cases, the toxin is so effective that even low toxin levels almost completely eradicate the non-producers.

In the discussion above, we argued that optimal distributions are "broad", in the sense that indeed the tails of the distributions stretch out to t_{max} (see Fig. 3). This raises the question on the influence of the parameter t_{max}. Its value is experimentally motivated, but may vary between species and environments. Hence, we compared histograms with different t_{max} but fixed combinations of ϵ_T and $k_{NT}\alpha_T$, see Fig. 4. We find that the distribution shapes are qualitatively similar, but increasing t_{max} results in longer tails at the expense of shorter lysis times. This becomes particularly evident when the parameters are unfavourable for the producers (large ϵ_T and low $k_{NT}\alpha_T$), where a gap emerges at short lysis times . We find similar changes to the resulting optimal distributions also when the number of lysing producers, P_0, varies: the smaller P_0, the less weight distributions have at short release times (see Fig. 5). To understand this similarity, we recall that the parameter P_0 determines how many cells can lyse, and consequently, how many lysis events can be distributed. Lower values of P_0 then mean fewer events to distribute over the investigated timespan. However, the same situation occurs when increasing the timespan, while P_0 remains at its original value: The difference between the old and new (larger) timespan is filled up with lysis events at the expense of shorter lysis times of the distribution. The length of the timespan is defined by t_{max}, which explains the connection between P_0 and t_{max}.

Our results on varying t_{max} also give insight to the role of the different lysis time regimes: For large t_{max} (or low P_0), only few lysis events are distributed over the timespan, and thus have to be economized. The resulting distributions then show lysis events only in the large lysis time regime. Due to the low number of producer cells, it is in this case optimal to produce rather more toxin than to react fast. Interestingly, if more

lysis events can be distributed, it becomes more efficient if some cells lyse early. While the "early lysers" cannot kill the sensitive strain as effective as with later toxin release, they inhibit the growth of its abundance in the population. This makes the lysis events that occur later more efficient. Therefore, the purpose of the early lysis regime is to keep the non-producer abundance at such low levels that late lysis events have higher impact.

Taken together, we find that broad distributions with positive skewness are the most efficient in competition with sensitive bacteria. This shape of the distribution creates a constant, significant level of toxins in the population, which is necessary to kill sensitive bacteria as fast as possible. Further analysis revealed the role of different lysis regimes: Fast lysing cells inhibit the growth of the non-producer, and thus make the more massive toxin release by late lysing cells more effective. We were able to determine the optimal distributions for given toxin degradation ϵ_T and toxin effectiveness $k_{NT}\alpha_T$, using our single compartment model and a evolutionary algorithm. However, it remains unclear if the results of this conceptional model indeed hold in a more realistic scenario, as many processes including diffusion are only treated effectively. To answer this question, we develop a detailed spatial model of two colliding bacterial colonies in the following, and, for corresponding parameter sets, put our optimal distributions to the test.
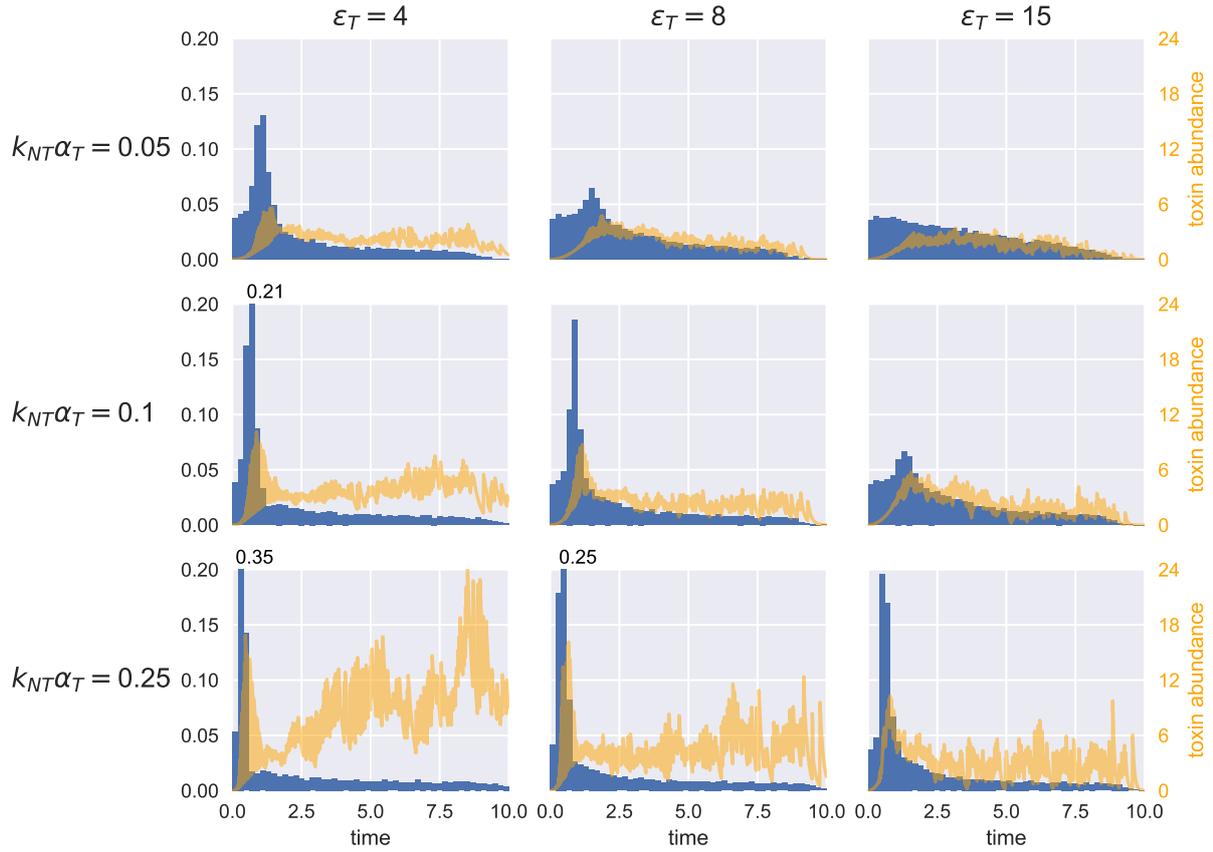
Figure 3: Optimal lysis time histograms for different values of the two model parameters. These histograms are the result of the evolutionary optimisation algorithm applied to the rescaled eqs. (4) and (5) after 20,000 generations, averaged over 20 distributions. The (rescaled) toxin concentration resulting from a single, exemplary distribution is plotted in orange. The smaller ε_T and the larger $k_{NT}\alpha_T$, the more favourable are the parameter conditions for toxin production.
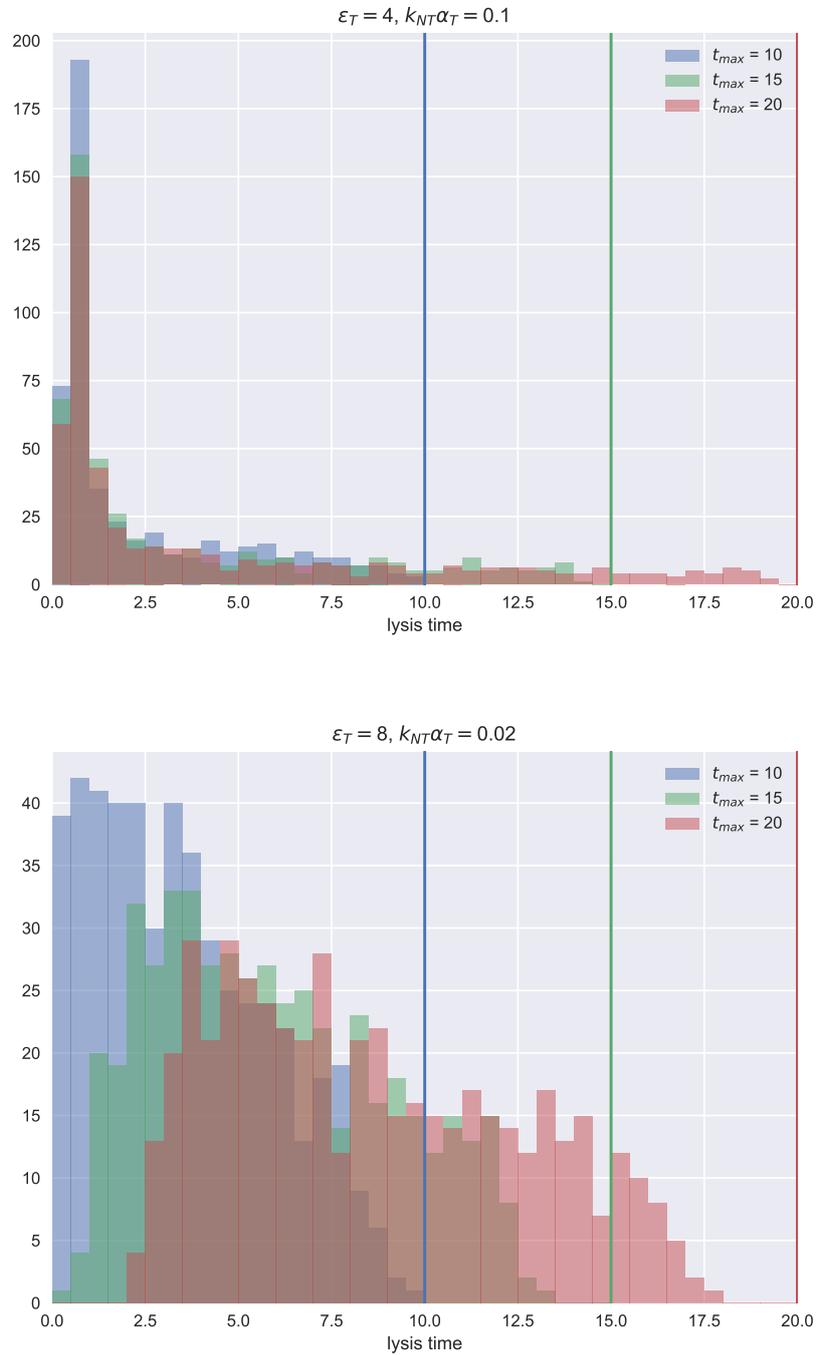
Figure 4: Two example parameter sets, that show how for different values of t_{max} (indicted by colored vertical lines) influence the optimal distribution. In both cases, the distribution is stretched out to t_{max} at the expense of shorter lysis times. If not specified otherwise, the same parameters as in Fig. 3 are used.
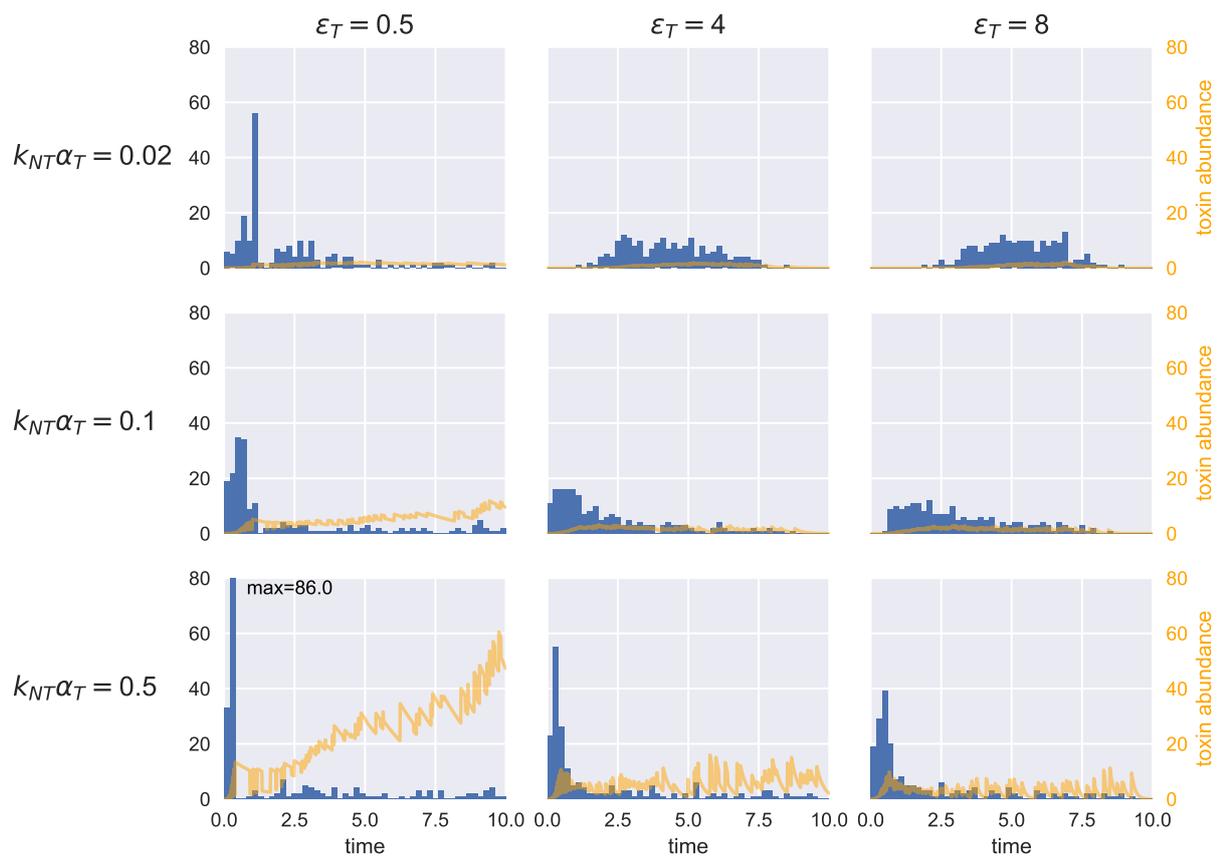
Figure 5: Lysis time histograms for the same parameters as in Fig. 3, but with $P_0 = 200$. Compared to results depicted in Fig. 3 ($P_0 = 500$), the lysis time distributions are qualitatively similar, but have less weight at short lysis times. These histograms are the result of the evolutionary optimisation algorithm after 20,000 generations. The (rescaled) toxin concentration is plotted in orange.
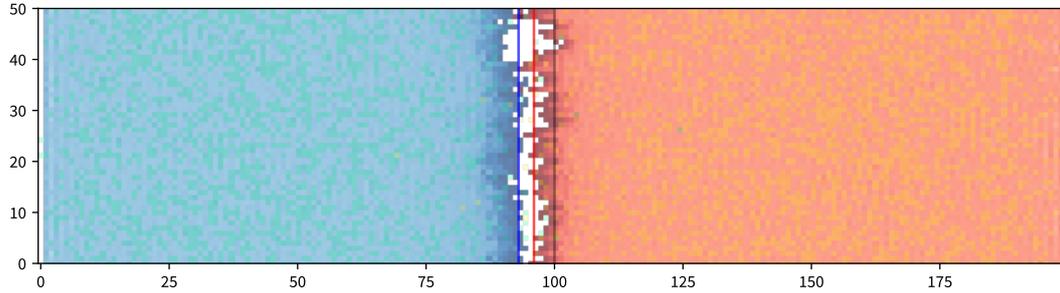
Figure 6: The setup of the spatial model, just before the colonies collide. The producer strain is depicted in blue, the non-producer in red. The front lines are depicted in corresponding colours. The green and yellow dots represent toxins and metabolic excretes, respectively. Parameters: $\beta_P = 1, \beta_N = 2, \delta_{P,N} = 0.05$, $D_{P,N} = 1 \, D_{T,M} = 15, \delta_{T,M} = 0.1, \alpha_T = 0.5, k_{NT} = 0.1$

2.3 spatial model

For the spatial model, we consider two bacterial colonies growing on a solid medium (that is, the bacteria are attached to the medium). As with the single compartment model, we are interested in the case of colliding fronts. However, we no do not consider just a single compartment, but a large set of compartments in the collision region of the two colonies. The compartments are arranged on a square lattice and each compartment is connected to its four nearest neighbours. We analyse a rectangular stripe, which is 200 compartments long and 50 compartments wide (see Fig. 6). As initial condition, the colonies populate opposite sites of the stripe, and are initially separated by an unpopulated gap in between. Moreover, we assume periodic boundary conditions for the long side of the stripe (that is, the stripe is the surface of an open cylinder).

We again consider two bacterial species, labelled P and N, of which only P produces toxins, T. However, the N strain produces metabolic excretes, M, which allow the produce strain to sense the presence of its competitor. The production of toxins is thus not triggered by a simultaneous switch to the SOS state within the compartment, but individually by a producer having contact to a molecule of the excrete . The time between start and release of toxins is again determined by a specific distribution.

In order to obtain a more realistic description for the dynamics inside the compartments we will lift many of the simplifying assumptions made for the single compartment model. In particular, since we now consider a stochastic process, all rates are now considered separately: We explicitly account for birth, death and hopping rates of the two strains, $\beta_{P,N}$, $\delta_{P,N}$ and $D_{P,N}$, respectively. Analogously, we now also consider the diffusion

rate of the toxins and excretes, $D_{T,M}$, separately from their spontaneous degradation, $\delta_{T,M}$. Moreover, we assume that a low percentage of spontaneously degrading producer cells also releases a small amount of toxins. This process is observed in experiments. The metabolic excretes are released exclusively via this mechanism.

As stated above, we aim to model bacteria that stick to the surface of a solid medium. In this situation, both the cells and the toxins can move on the lattice, but in different ways. The bacteria are attached to the surface of the medium. Therefore, they do not diffuse, but are pushed towards the outside of the colony by growing bacteria in the inside. We implement this pushing by selecting bacteria at a per capita rate $D_{P,N}$, and letting them hop according to abundance gradients along the growth direction of the colony. More specifically, we compare the abundance of the compartment behind and in front of the selected compartment (relative to the growth direction of its strain): If the front compartment has a lower abundance, the cell hops into it, whereas, if the front compartment has equal or larger abundance, the cell hops into the compartments either right or left of the front compartment. If these compartments also have a larger abundance, the bacterium hops backwards.

As the toxins and metabolic excretes are significantly smaller than the bacteria, it is reasonable to assume that they diffuse freely in any direction. This process is implemented as a per-capita diffusion rate $D_{T,M}$, by which toxins and excretes hop from their current compartment to one of the four neighbouring compartments.

2.4 Results and Discussion of Spatial Model Test

Because of computational limitations, we cannot employ the genetic algorithm for the spatial model. We therefore do not determine the optimal distribution for the spatial model, but compare the optimal distribution (determined using the single compartment model) with other distribution for given parameter sets. We consider two distinct parameter sets: one, which poses hard conditions for toxin efficiency (high toxin degradation and low toxin binding: $\epsilon_T = 15, k_{NT} = 0.1$), and a second one with low degradation and high binding ($\epsilon_T = 1, k_{NT} = 0.5$), which are favourable for toxin producers. The toxin production is in both cases set to $\alpha_T = 0.5$, which enables us to use the corresponding optimal distributions with the values $k_{NT}\alpha_T = 0.05$ and $k_{NT}\alpha_T = 0.25$ determined from the single compartment model in Fig. 3[1]. In the favourable conditions, the optimal

---

[1]All remaining parameters are the same as given in Fig. 6

distribution shows a large peak at short lysis times, whereas it is much broader with more weight at large lysis times in hard conditions.

To compare the effect of these two distributions, we first set up two simulations with favourable conditions, that is, $\epsilon_T = 1$ and $k_{NT} = 0.5$. In the first simulation, the producer releases toxins according to the distribution optimised for hard environments, in the second one, toxin release events are drawn from the distribution for easy conditions During the simulation runs, we recorded for both distributions each strain's front line positions over time (Fig. 7A). We find that for both distributions, the producing strain A is able to repel the sensitive one.   Apart from this general advantage due to toxin production, our results confirm the conclusions from the single compartment model: The distribution optimized for favourable conditions indeed outperforms its counterpart (Fig. 7A). This is because the large peak at short lysis times allows the efficient toxins to act early on. We repeated the aforementioned testing scheme also for hard conditions (Fig. 7B). With the change of conditions, also the roles reverse: now the distribution optimised for low toxin/non-producer interaction and high degradation performs best, as now larger toxin concentrations (and thus later lysis times) are necessary to effectively reduce the sensitive strain.
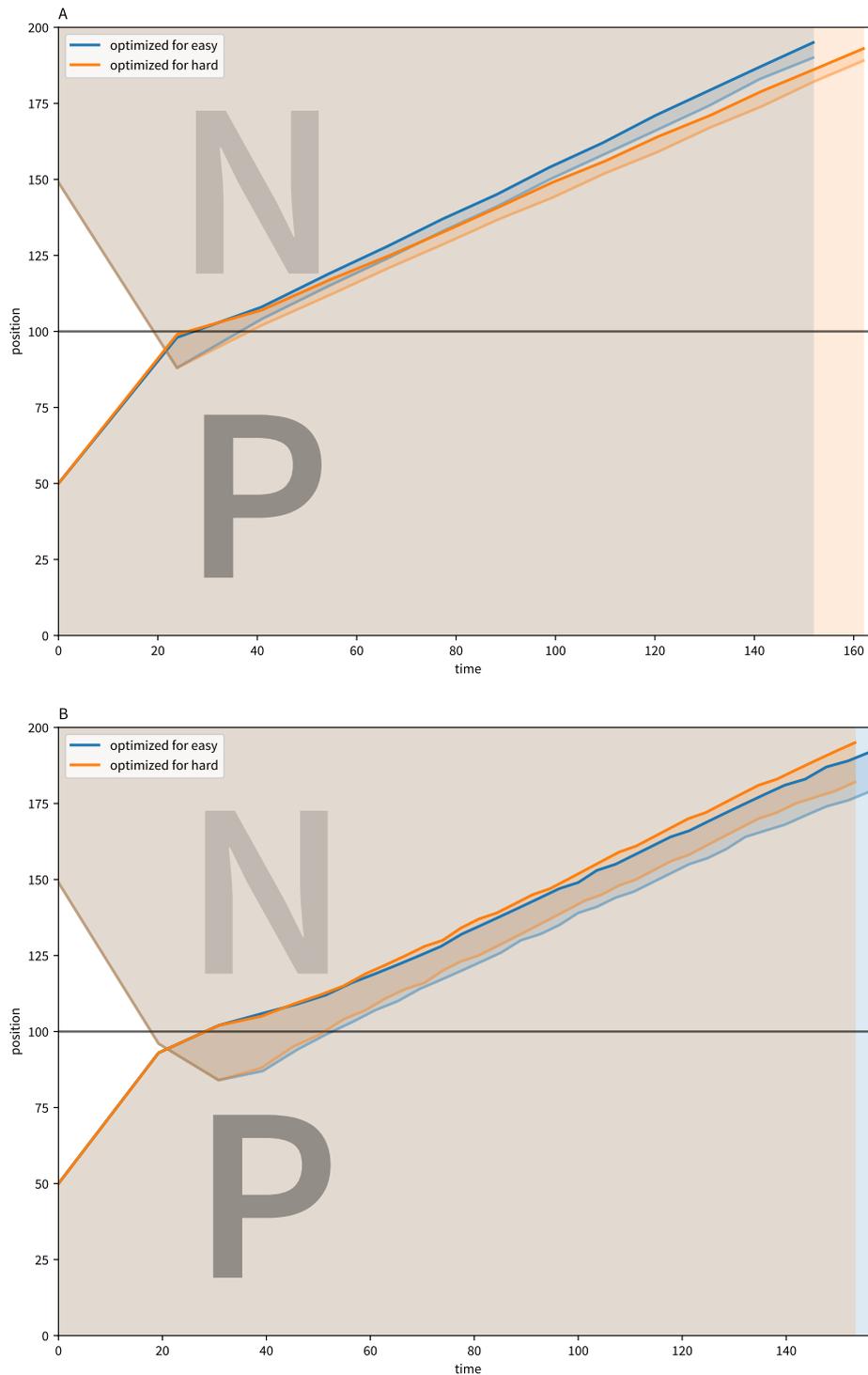
Figure 7: Testing the results of the well-mixed model. Both plots show the front lines, which are defined as the column where at least half of the compartments contain at least one cell of the corresponding species. We consider two exemplary parameter sets. The first (A) is favourable for toxins, meaning that is has low toxin degradation ($\epsilon_T = 1$) and high toxin-interaction ($k_T = 0.5$), whereas the second (B) is the opposite ($\epsilon_T = 15, k_T = 0.1$). We use the corresponding optimal distributions from Fig. 3 for this parameter sets, and let both compete against a non-producing strain in both conditions. The results support our findings from the single compartment model: Each optimized distribution outperforms its counterpart in its own optimal conditions.

3 Summary and Outlook

In this manuscript, we have studied the role of distributions underlying the stochastic release of bacterial toxins. We considered a boundary compartment of a toxin producing colony during a collision with a toxin-sensitive non-producer colony. For this scenario, we developed a simple model using only the toxin and non-producer abundance as dynamic variables, with the toxin being released according to a specific lysis time distribution. Non-dimensionalisation of the resulting equations revealed that the effective toxin degradation ϵ_T, as well as the product of toxin synthesis rate and interaction with the sensitive strain, $k_{NT}\alpha_T$, are the key parameters affecting the system. Using an evolutionary algorithm, we determined the optimal lysis time distributions for different sets of ϵ_T and $k_{NT}\alpha_T$. For two exemplary parameter sets, we tested the optimal distributions with a lattice-gas model of two colliding colonies, and showed that the results of our conceptual model also hold qualitatively for a more realistic system.

The distributions resulting from the evolutionary optimisation process show a clear dependency on the toxin production, ϵ_T, and the product of toxin binding and production rate, $k_{NT}\alpha_T$. If $k_{NT}\alpha_T$, is high, the optimal lysis time distributions have a large, pronounced peak at short lysis times, as the toxins then can effectively kill the sensitive strain already quite early. A remaining small amount of lysis events distributed over the timespan up to t_{max} then ensure that further incoming or surviving non-producers are kept at low levels. In the case of a low binding and production product $k_{NT}\alpha_T$, however, this peak disappears, and more weight is put to later lysis times, where it is more broadly distributed. This is because now more toxin needs to accumulate in the cells to counteract the non-producers. A large value of toxin diffusion, ϵ_T, increases this effect, as it determines how long (on average) a toxin can act against sensitive bacteria.

Our results present another example of how heterogeneous phenotypes provide benefits for a bacterial population. For the special case of heterogeneous lysis, previous work focused on heterogeneity in the sense that only a fraction of the population produces toxin [2, 6]. Here, however, we have investigated heterogeneity in the *timing* of self-destructive toxin release, and studied its consequences to bacteria in competitive situations. To keep the model simple and focus on the relevant factors, we make the assumption that all cells subject to stress will produce toxins and lyse eventually. This also agrees with previous experiments [7], which show that above a certain threshold stress level the SOS response is irreversible and all affected cells lyse.

One of the general results of this study is that the optimal distributions are always broad, meaning that they span almost the whole range of possible lysis times. A naive interpretation of this property might have been bet hedging, since broad distributions let at least some bacteria lyse at the respective optimal time point for a given condition. Bet hedging has indeed been discussed in the context of phenotypic heterogeneity of bacteria before [11, 6]. However, our study shows that the distributions can rather be understood as a form of division of labour: the fast-lysing bacteria in the population react quickly to stress imposed by competitors, and minimise their abundance and dispersal. This gives bacteria that lyse later the time necessary to produce effective amounts of toxins, which then can kill off and thus repel the non-producing strain.

In this manuscript, we discussed the consequences of heterogeneity in the release of toxins. Our research was motivated from experimental studies on ColicinE2 prosuction and release, which is triggered by an SOS response system. The model in this study, however, relies only on few, and very general conditions. Therefore, our results are independent from the underlying SOS system, and can also be transferred to other self-destructive bacterial toxin release systems. Taken together, we presented a general model to analyse the effects and effectiveness of toxin release distributions, which enables the study of a new aspect of phenotypic heterogeneity.

References

[1] James C. Paton. The contribution of pneumolysin to the pathogenicity of Streptococcus pneumoniae, mar 1996.

[2] M Ackermann, B Stecher, N E Freed, P Songhet, W.-D. Hardt, and M Doebeli. Self-destructive cooperation mediated by phenotypic noise. *Nature*, 454(7207):987–990, 2008.

[3] Eric Cascales, Susan K Buchanan, Denis Duché, Colin Kleanthous, Roland Lloubès, Kathleen Postle, Margaret Riley, Stephen Slatin, and Danièle Cavard. Colicin biology. *Microbiol. Mol. Biol. Rev.*, 71(1):158–229, 2007.

[4] Peter Mrak, Zdravko Podlesek, Jos P. M. van Putten, and Darja Žgur-Bertok. Heterogeneity in expression of the Escherichia coli colicin K activity gene cka is controlled by the SOS system and stochastic factors. *Mol. Genet. Genomics*, 277(4):391–401, apr 2007.

[5] David Nellinger Adamson and Han N Lim. Rapid and robust signaling in the CsrA cascade via RNA–protein interactions and feedback regulation. *Proc. Natl. Acad. Sci.*, 110(32):13120–13125, 2013.

[6] Bihter Bayramoglu, David Toubiana, Simon van Vliet, R. Fredrik Inglis, Nadav Shnerb, and Osnat Gillor. Bet-hedging in bacteriocin producing Escherichia coli populations: the single cell perspective. *Sci. Rep.*, 7:42068, 2017.

[7] Andreas Mader, Benedikt von Bronk, Benedikt Ewald, Sara Kesel, Karin Schnetz, Erwin Frey, and Madeleine Opitz. Amount of Colicin Release in Escherichia coli Is Regulated by Lysis Gene Expression of the Colicin E2 Operon. *PLoS One*, 10(3), 2015.

[8] Matthias Lechner, Mathias Schwarz, Madeleine Opitz, and Erwin Frey. Hierarchical Post-transcriptional Regulation of Colicin E2 Expression in Escherichia coli. *PLOS Comput. Biol.*, 12(12):e1005243, dec 2016.

[9] Yishai Shimoni, Shoshy Altuvia, Hanah Margalit, and Ofer Biham. Stochastic Analysis of the SOS Response in Escherichia coli. *PLoS One*, 4(5), 2009.

[10] Chris Jones and I Barry Holland. Role of the SulB (FtsZ) protein in division inhibition during the SOS response in Escherichia coli: FtsZ

stabilizes the inhibitor SulA in maxicells. *Proc. Natl. Acad. Sci. U. S. A.*, 82(18):6045–9, 1985.

[11] Martin Ackermann. A functional perspective on phenotypic hetero-geneity in microorganisms. *Nat. Rev. Microbiol.*, 13(8):497–508, 2015.

Supporting Information

Detailed Evolutionary Algorithm

Here, we describe in detail the evolutionary algorithm we implemented to obtain the optimal distribution for a given set of parameters. The algorithm runs on an ensemble of 40 compartments, each with a distribution of P_0 lysis time points. On this ensemble, it performs the following steps:

1. Initialize each compartment of the ensemble with an initial distribution. We chose all producer cells to lyse at half the maximal time (that is, a delta peak at $t = 5$) as the first initial distribution.

2. For all ensemble compartments (but one), mutate the distribution according to a mutation rate m: From the P_0 lysis times in the distribution, select m random ones. For each of the selected waiting times, determine a normal distributed random number, and add it to the waiting time, resulting in a new waiting time. If this waiting time is below 0 or above t_{max}, reflective boundaries apply. We leave one compartment unmutated, to make sure that the mutated ones are better than the current state.

3. Calculate the cumulative non-producer abundance for each compartment in the ensemble with the procedure described in the main text.

4. Take the results of all compartments in the ensemble, and select the two distributions with the lowest cumulative non-producer abundance. Taking two instead of just the first one reduces the likelihood to get trapped in local minima.

5. For one half of the ensemble, replace each distribution with the best distribution, for the other half the second best.

6. Go back to step 2 to start the next generation. The best and second best distributions are now the new initial distributions.

For the optimal distributions presented in the results section, we simulated 20,000 generations for 20 times, and determined the optimal histogram from the sum of the resulting optimal distributions. In each simulation run, we started with a high mutation rate of 100 mutations (of 500 producers in total) to explore a broad "landscape" of distributions. After 2000 generations, we reduced this rate to 10 mutations.

Bibliography

[1] J. L. Sonnenburg, L. T. Angenent, and J. I. Gordon. *Getting a grip on things: how do communities of bacterial symbionts become established in our intestine?* In: **Nat. Immunol.** 5.6 (June 2004), pp. 569–573. DOI: 10.1038/ni1079 (cit. on p. 1).

[2] S. G. Wright. *Evolution in Mendelian populations.* In: **Genetics** 16.2 (1931), pp. 97–159 (cit. on pp. 1, 6, 7).

[3] M. Thattai and a. van Oudenaarden. *Intrinsic noise in gene regulatory networks.* In: **Proc. Natl. Acad. Sci. U. S. A.** 98.15 (July 2001), pp. 8614–9. DOI: 10.1073/pnas. 151588598 (cit. on pp. 1, 42, 170).

[4] M. Kimura. *Solution of a process of random genetic drift with a continuous model.* In: **Proc. Natl. Acad. Sci. U. S. A.** 2.2 (1955), pp. 144–150 (cit. on pp. 1, 6).

[5] C. R. Darwin. **On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life**. 1st ed. London: John Murray, 1859 (cit. on p. 6).

[6] M. Kimura. **The Neutral Theory of Molecular Evolution**. Cambridge University Press, 1983 (cit. on p. 6).

[7] T. Ohta and J. H. Gillespie. *Development of neutral and nearly neutral theories.* In: **Theor. Popul. Biol.** 49.2 (1996), pp. 128–142. DOI: 10.1006/tpbi.1996.0007 (cit. on p. 6).

[8] D. L. Hartl and A. G. Clark. **Principles of Population Genetics.** 4th ed. Sinauer Associates, 1989 (cit. on p. 6).

[9] R. A. Blythe and A. J. McKane. *Stochastic models of evolution in genetics, ecology and linguistics.* In: **J. Stat. Mech. Theory Exp.** 2007.7 (July 2007), P07018–P07018. DOI: 10.1088/1742-5468/2007/07/P07018. arXiv: 0703478 [cond-mat] (cit. on p. 6).

[10] F. Eggenberger and G. Pólya. *Über die Statistik verketteter Vorgänge.* In: **ZAMM - J. Appl. Math. Mech. / Zeitschrift für Angew. Math. und Mech.** 3.4 (1923), pp. 279–289. DOI: 10.1002/zamm.19230030407 (cit. on p. 7).

[11] R. Pemantle. *A survey of random processes with reinforcement.* In: **Probab. Surv.** 4 (2007), pp. 1–79. DOI: 10.1214/07-PS094. arXiv: 0610076 [math] (cit. on p. 7).

[12] W. Brian Arthur, Y. M Ermoliev, and Y. M Kaniovski. *Path-dependent processes and the emergence of macro-structure.* In: **Eur. J. Oper. Res.** 30.3 (1987), pp. 294–303. DOI: 10.1016/0377-2217(87)90074-9 (cit. on p. 7).

[13] A. Melbinger, J. Cremer, and E. Frey. *Evolutionary game theory in growing populations.* In: **Phys. Rev. Lett.** 105.17 (Oct. 2010), p. 178101. DOI: 10.1103/PhysRevLett.105. 178101 (cit. on pp. 7, 9).

[14] J. Cremer, A. Melbinger, and E. Frey. *Growth dynamics and the evolution of cooperation in microbial populations.* In: **Sci. Rep.** 2.1 (Dec. 2012), p. 281. DOI: 10.1038/srep00281. arXiv: 1203.5863 (cit. on pp. 7, 9).

[15] J. Cremer, A. Melbinger, and E. Frey. *Evolutionary and population dynamics: A coupled approach.* In: **Phys. Rev. E** 84.5 (Nov. 2011). DOI: 10.1103/PhysRevE.84.051921 (cit. on pp. 7, 9).

[16] A. R. Templeton. *The Theory of Speciation Via the Founder Principle.* In: **Genetics** 94.4 (1980), pp. 1011–1038. DOI: 10.2307/2408012. arXiv: 0002114v2 [arXiv:hep-ph] (cit. on p. 7).

[17] J. H. Gillespie. *Is the population size of a species relevant to its evolution?* In: **Evolution** 55.11 (2001), pp. 2161–2169. DOI: 10.1111/j.0014-3820.2001.tb00732.x (cit. on p. 7).

[18] G. Price. *Selection and covariance.* In: **Nature** 227 (1970), pp. 520–521 (cit. on p. 9).

[19] S. Okasha. *Multi-level selection, covariance and contextual analysis.* In: **Br. J. Philos. Sci.** 55.3 (Sept. 2004), pp. 481–504. DOI: 10.1093/bjps/55.3.481 (cit. on p. 9).

[20] B. Alberts et al. **Molecular Biology of the Cell**. 4th ed. Garland Science, 2002 (cit. on pp. 38, 39).

[21] Y. Shimoni et al. *Stochastic Analysis of the SOS Response in Escherichia coli.* In: **PLoS One** 4.5 (2009). Ed. by M. Isalan. DOI: 10.1371/journal.pone.0005363 (cit. on pp. 39, 40, 42).

[22] M. Radman. *SOS repair hypothesis: phenomenology of an inducible DNA repair which is accompanied by mutagenesis.* In: **Basic Life Sci** 5A (1975), pp. 355–67. DOI: 10.1007/978-1-4684-2895-7_48 (cit. on p. 39).

[23] J. W. Little and D. W. Mount. *The SOS regulatory system of Escherichia coli.* In: **Cell** 29.1 (1982), pp. 11–22. DOI: 10.1016/0092-8674(82)90085-X (cit. on p. 39).

[24] M. Ronen et al. *Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics.* In: **Proc. Natl. Acad. Sci. U. S. A.** 99.16 (2002), pp. 10555–10560. DOI: 10.1073/pnas.152046799 (cit. on p. 39).

[25] N. Friedman et al. *Precise temporal modulation in the response of the SOS DNA repair network in individual bacteria.* In: **PLoS Biol.** 3.7 (2005), pp. 1261–1268. DOI: 10.1371/journal.pbio.0030238 (cit. on p. 40).

[26] M. Nishie, J. Nagao, and K. Sonomoto. *Antibacterial peptides "bacteriocins": an overview of their diverse characteristics and applications.* In: **Biocontrol Sci** 17.1 (2012), pp. 1–16. DOI: 10.4265/bio.17.1 (cit. on p. 40).

[27] E. Cascales et al. *Colicin biology.* In: **Microbiol. Mol. Biol. Rev.** 71.1 (2007), pp. 158–229. DOI: 10.1128/MMBR.00036-06 (cit. on p. 40).

[28] H. Masaki and T. Ohta. *Colicin E3 and its immunity genes.* In: **J. Mol. Biol.** 182.2 (1985), pp. 217–227. DOI: 10.1016/0022-2836(85)90340-7 (cit. on p. 40).

[29] A. P Pugsley and M. Schwartz. *Expression of a gene in a 400-base-pair fragment of colicin plasmid ColE2-P9 is sufficient to cause host cell lysis.* In: **J. Bacteriol.** 156.1 (1983), pp. 109–114 (cit. on p. 40).

[30] T. Y. Yang et al. *Posttranscriptional repression of the cel gene of the ColE7 operon by the RNA-binding protein CsrA of Escherichia coli.* In: **Nucleic Acids Res.** 38.12 (2010), pp. 3936–3951. DOI: 10.1093/nar/gkq177 (cit. on p. 41).

[31] T. Romeo. *Global regulation by the small RNA-binding protein CsrA and the non- coding RNA molecule CsrB.* In: **Mol. Microbiol.** 29.6 (1998), pp. 1321–1330. DOI: 10.1046/j.1365-2958.1998.01021.x (cit. on p. 41).

[32] J. Timmermans and L. Van Melderen. *Post-transcriptional global regulation by CsrA in bacteria.* In: **Cell. Mol. Life Sci.** 67.17 (2010), pp. 2897–2908. DOI: 10.1007/s00018-010-0381-z (cit. on p. 41).

[33] M. Y. Liu et al. *The RNA molecule CsrB binds to global regulatory protein CsrA and antagonize its activity in Escherichia coli.* In: **J. Biol. Chem.** 272.28 (1997), pp. 17502–17510. DOI: 10.1074/jbc.272.28.17502 (cit. on p. 41).

[34] T. Weilbacher et al. *A novel sRNA component of the carbon storage regulatory system of Escherichia coli.* In: **Mol. Microbiol.** 48.3 (2003), pp. 657–670. DOI: 10.1046/j.1365-2958.2003.03459.x (cit. on p. 41).

[35] P. Babitzke and T. Romeo. *CsrB sRNA family: sequestration of RNA-binding regulatory proteins.* In: **Curr. Opin. Microbiol.** 10.2 (2007), pp. 156–163. DOI: 10.1016/j.mib.2007.03.007 (cit. on p. 41).

[36] L. S. Waters and G. Storz. *Regulatory RNAs in Bacteria.* In: **Cell** 136.4 (2009), pp. 615–628. DOI: 10.1016/j.cell.2009.01.043. arXiv: NIHMS150003 (cit. on p. 41).

[37] D. N. Adamson and H. N. Lim. *Rapid and robust signaling in the CsrA cascade via RNA–protein interactions and feedback regulation.* In: **Proc. Natl. Acad. Sci.** 110.32 (2013), pp. 13120–13125. DOI: 10.1073/pnas.1308476110 (cit. on p. 41).

[38] M. F. Weber et al. *Chemical warfare and survival strategies in bacterial range expansions.* In: **J. R. Soc. Interface** 11.96 (2014), p. 20140172. DOI: 10.1098/rsif.2014.0172 (cit. on p. 41).

[39] A. Mader et al. *Amount of Colicin Release in Escherichia coli Is Regulated by Lysis Gene Expression of the Colicin E2 Operon.* In: **PLoS One** 10.3 (2015). DOI: 10.1371/journal.pone.0119124 (cit. on pp. 41, 173).

[40] M. Morales et al. *Accumulation of Single-Stranded DNA in Escherichia coli Carrying the Colicin Plasmid pColE3-CA38.* In: **Plasmid** (2015) (cit. on p. 41).

[41] S. Gudapaty et al. *Regulatory Interactions of Csr Components : the RNA Binding Protein CsrA Activates csrB Transcription in Escherichia coli Regulatory Interactions of Csr Components : the RNA Binding Protein CsrA Activates csrB Transcription in Escherichia coli.* In: **J. Bacteriol.** (2001). DOI: 10.1128/JB.183.20.6017 (cit. on p. 42).

[42] E. Levine et al. *Quantitative Characteristics of Gene Regulation by Small RNA.* In: **PLoS Biol.** 5.9 (2007), e229. DOI: 10.1371/journal.pbio.0050229 (cit. on p. 42).

[43] E. Levine and T. Hwa. *Small RNAs establish gene expression thresholds*. In: **Curr. Opin. Microbiol.** 11.6 (Dec. 2008), pp. 574–579. DOI: 10.1016/j.mib.2008.09.016 (cit. on pp. 42, 44).

[44] N. G. van Kampen. **Stochastic Processes in Physics and Chemistry**. 3rd ed. Amsterdam: Elsevier, 2007 (cit. on p. 42).

[45] C. W. Gardiner. **Handbook of Stochastic Methods**. Ed. by H. Haken. 3rd ed. Springer Series in Synergetics. Berlin: Springer, 2009 (cit. on p. 42).

[46] P. Mehta, S. Goyal, and N. S. Wingreen. *A quantitative comparison of sRNA-based and protein-based gene regulation*. In: **Mol. Syst. Biol.** 4.221 (2008). DOI: 10.1038/msb.2008.58. arXiv: 0809.0527 (cit. on pp. 42, 44).

[47] J. Elf and M. Ehrenberg. *Fast Evaluation of Fluctuations in Biochemical Networks With the Linear Noise Approximation*. In: **Genome Res.** 13.11 (Nov. 2003), pp. 2475–2484. DOI: 10.1101/gr.1196503 (cit. on pp. 42, 44).

[48] D. T. Gillespie. *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions*. In: **J. Comput. Phys.** 22.4 (Dec. 1976), pp. 403–434. DOI: 10.1016/0021-9991(76)90041-3 (cit. on p. 42).

[49] D. T. Gillespie. *Exact stochastic simulation of coupled chemical reactions*. In: **J. Phys. Chem.** 81.25 (1977), pp. 2340–2361. DOI: 10.1021/j100540a008 (cit. on p. 42).

[50] S. Legewie et al. *Small RNAs establish delays and temporal thresholds in gene expression*. In: **Biophys. J.** 95.7 (2008), pp. 3232–3238. DOI: 10.1529/biophysj.108.133819. arXiv: 0812.0025 (cit. on p. 44).

[51] S. Kamenšek et al. *Genes regulated by the Escherichia coli SOS repressor LexA exhibit heterogenous expression*. In: **BMC Microbiol.** 10.1 (2010), p. 283. DOI: 10.1186/1471-2180-10-283 (cit. on p. 44).

[52] P. Mrak et al. *Heterogeneity in expression of the Escherichia coli colicin K activity gene cka is controlled by the SOS system and stochastic factors*. In: **Mol. Genet. Genomics** 277.4 (Apr. 2007), pp. 391–401. DOI: 10.1007/s00438-006-0185-x (cit. on p. 44).

[53] S. A. Khan. *Rolling-circle replication of bacterial plasmids*. In: **Microbiol. Mol. Biol. Rev.** 61.4 (1997), pp. 442–55 (cit. on p. 45).

[54] S. A. Khan. *Plasmid rolling-circle replication: Recent developments*. In: **Mol. Microbiol.** 37.3 (2000), pp. 477–484. DOI: 10.1046/j.1365-2958.2000.02001.x (cit. on p. 45).

[55] J. Paulsson and M. Ehrenberg. *Noise in a minimal regulatory network: plasmid copy number control*. In: **Q. Rev. Biophys.** 34.1 (Feb. 2001), pp. 1–59 (cit. on p. 45).

[56] J.-W. Veening, W. K. Smits, and O. P Kuipers. *Bistability, Epigenetics, and Bet-Hedging in Bacteria*. In: **Annu. Rev. Microbiol.** 62.1 (Oct. 2008), pp. 193–210. DOI: 10.1146/annurev.micro.62.081307.163002 (cit. on p. 170).

[57] M. Ackermann. *A functional perspective on phenotypic heterogeneity in microorganisms*. In: **Nat. Rev. Microbiol.** 13.8 (2015), pp. 497–508. DOI: 10.1038/nrmicro3491 (cit. on p. 170).

[58] J. van Gestel and M. A. Nowak. *Phenotypic Heterogeneity and the Evolution of Bacterial Life Cycles*. In: **PLoS Comput. Biol.** 12.2 (Feb. 2016). Ed. by C. O. Wilke, e1004764. DOI: 10.1371/journal.pcbi.1004764 (cit. on p. 170).

[59] M. Ackermann et al. *Self-destructive cooperation mediated by phenotypic noise*. In: **Nature** 454.7207 (2008), pp. 987–990. DOI: 10.1038/nature07067 (cit. on pp. 170, 174).

[60] M. Bauer et al. *Ecological feedback in quorum-sensing microbial populations can induce heterogeneous production of autoinducers*. In: **Elife** 6 (2017), pp. 1–38. DOI: 10.7554/eLife.25773. arXiv: arXiv:1707.07959v1 (cit. on p. 170).

[61] N. Q. Balaban. *Bacterial Persistence as a Phenotypic Switch*. In: **Science (80-.).** 305.5690 (Sept. 2004), pp. 1622–1625. DOI: 10.1126/science.1099390 (cit. on p. 170).

[62] O. Gefen and N. Q. Balaban. *The importance of being persistent: heterogeneity of bacterial populations under antibiotic stress*. In: **FEMS Microbiol. Rev.** 33.4 (2009). Ed. by D. Gutnick, pp. 704–717. DOI: 10.1111/j.1574-6976.2008.00156.x (cit. on p. 170).

[63] H. J. E. Beaumont et al. *Experimental evolution of bet hedging*. In: **Nature** 462.7269 (Nov. 2009), pp. 90–93. DOI: 10.1038/nature08504 (cit. on p. 170).

[64] B. Bayramoglu et al. *Bet-hedging in bacteriocin producing Escherichia coli populations: the single cell perspective*. In: **Sci. Rep.** 7 (2017), p. 42068. DOI: 10.1038/srep42068 (cit. on p. 170).

[65] J. V. a. N. Gestel, H. Vlamakis, and R. Kolter. *Division of Labor in Biofilms : the Ecology of Cell Differentiation*. In: **Microbiol. Spectr.** 3.2 (Apr. 2015), pp. 1–24. DOI: 10.1128/microbiolspec.MB-0002-2014.Correspondence (cit. on p. 170).

[66] Z. Zhang, D. Claessen, and D. E. Rozen. *Understanding microbial divisions of labor*. 2016. DOI: 10.3389/fmicb.2016.02070. arXiv: arXiv:1011.1669v3 (cit. on p. 170).

[67] J.-Y. Potvin. *Genetic algorithms for the traveling salesman problem*. In: **Ann. Oper. Res.** 63.3 (June 1996), pp. 337–370. DOI: 10.1007/BF02125403 (cit. on p. 170).

[68] F. Van Batenburg, A. Gultyaev, and C. Pleij. *An APL-programmed genetic algorithm for the prediction of RNA secondary structure*. In: **J. Theor. Biol.** 174.3 (June 1995), pp. 269–280. DOI: 10.1006/jtbi.1995.0098 (cit. on p. 170).

[69] J. Lohn, G. Hornby, and D. Linden. *An evolved antenna for deployment on NASA's space technology 5 mission*. In: **J. Chem. Inf. Model.** 53.9 (2013), pp. 1689–1699. DOI: 10.1017/CBO9781107415324.004. arXiv: arXiv:1011.1669v3 (cit. on p. 170).

[70] J. Grefenstette et al. „Genetic algorithms for the traveling salesman problem". In: *Proc. first Int. Conf. Genet. Algorithms their Appl.* Lawrence Erlbaum, New Jersey (160-168), 1985, pp. 160–168 (cit. on p. 170).

Acknowledgements

First, I would like to express my gratitude to my supervisor Erwin Frey. Thank you for giving me the opportunity to work in your group and making it possible for me to participate in the scientific life. I particularly appreciate your constant trust and support throughout our projects, especially in difficult situations.

I am also grateful to my collaborators (in order of appearance) Karl Wienand, Felix Becker, Heinrich Jung, Mathias Schwarz, Madeleine Opitz, and Alexandra Götz, for all their work and help in our joint projects. I very much enjoyed the creative scientific (and also non-scientific) discussions with you! Especially, I want to thank my fellow admin Karl for the many many rides to and from Antholz and Großhadern, talking about science, music, nerd culture, and organising weddings. Special thanks also go to Madeleine. Thank you for giving me the opportunity to join the ColicinE2 project, for ad-hoc microbiology courses, and for your patience and support!

I would like to thank Stefano Duca and Matthias Bauer, who I had the pleasure to co-supervise with Hannes Knebel during their master theses. Special thanks go to Hannes for joining me in these projects, and the countless scientific and non-scientific discussions in your office.

During my doctoral studies, I shared my office (and my whiteboard) with Matthias Rank, who in many ways turned out to be the best man for this job. Thank you for the great atmosphere in our office, your friendship, and all the insights to the mysterious world of Scottish country dancing. I would also like to thank our office-neighbour Marianne Bauer for her short-term proofreading of my thesis.

I am grateful to all the present and previous members of Erwin's chair for the joyful and creative atmosphere. Apart from the people already mentioned, I especially want to mention my former office-mate Markus F. Weber, our follow-up admins Patrick Wilke and Laeschkir Hassan, and the Mensa-co-debaters Cornelius Weig, Isabella Graf, Felix Kempf, Silke Bergeler, Emanuel Reithmann, Jonas Denk, Louis Reese and Cristina Pop.

Many thanks must also go to my friends and my family, especially my parents, for their constant love and support over all these years.

Last, and probably the most, I want to thank Sarah, for keeping my life heterogeneous in spite of all its fluctuations.

❦