THE UNIVERSITY
of EDINBURGH

Regulation of RecBCD expression in *Escherichia coli*

Irina Kalita

Thesis presented for the degree of
Doctor of Philosophy
The University of Edinburgh

December, 2021

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Irina Kalita

Acknowledgements

First of all, I would like to thank my supervisor, Prof. Meriem El Karoui, for guidance and support all the way, for sharing passion for science, for exciting discussions, for provided opportunities to learn and progress. Thanks, Meriem, for always meeting me with understanding, tolerance, and empathy, and for being such an inspiring scientist and person.

I also would like to thank Prof. Ramon Grima for introducing me to a beautiful world of stochastic modelling and Dr Sander Granneman for sharing breakthrough ideas and enthusiasm. I am grateful to Prof. Rosalind Allen, Prof. David Leach and Dr Elise Darmon for valuable discussions, critical feedback, and challenging questions.

I would love to thank my dear friend Ira Iosub for her unconditional trust and our lasting connection, for sincere advice and constructive feedback, for our fruitful collaboration, and last but not least, for introducing me to film photography. I am thankful to Alessia Lepore, Sebastian Jaramillo-Riveri, Lorna McLaren, Xavier Zaoui, James Broughton, Daniel Thédié, for making the lab an enjoyable place to work. I would like to thank James Holehouse and all members of Grima's group for taking care of me lost in math. I also would like to thank Benura Azeroglu for teaching me the art of cloning and being ready to help at any time of day and night. Thanks, Tahirah Yasmin, for the wholehearted late-evening chats in Roger Land. I would like to thank the students I had a pleasure to supervise, Rachel Jackson, Léna Le Quellec, and Achille Fraisse, for their contribution to the project as well as for challenging my teamwork skills.

I am genuinely thankful to The Darwin Trust of Edinburgh for funding my degree and supporting program extension due to Covid-19.

I am grateful to my family for doing everything possible and impossible to help me in reaching ambitious goals and satisfying my curiosity.

Lay Summary

Keeping precise genome sequence over many generations is one of the main priorities of life. Thus, all living organisms developed mechanisms to protect their DNA from various kinds of damage. One of the most dangerous types of DNA damage is when both chromosomal strands are broken simultaneously. In such cases, bacteria *Escherichia coli* employs a strategy which allows to repair DNA without any loss of information. RecBCD enzyme is one of the key players of this DNA repair pathway. In fact, RecBCD is essential for detection of a DNA lesion and its further repair. Interestingly, RecBCD can also degrade the chromosome very fast posing a potential risk of DNA loss. In line with this, having too many RecBCD molecules in a cell has been shown to reduce bacterial ability to repair their DNA. Taken together, these observations suggest that RecBCD expression needs to be regulated.

Here, RecB subunits were quantified at the protein and mRNA levels in single cells. Surprisingly, while the mRNA level is noisy, fluctuations in RecB protein levels are remarkably suppressed. The latter further suggests that RecBCD expression is regulated by a post-transcriptional mechanism. Consistently with this hypothesis, a global post-transcriptional regulator Hfq has been found to bind *recBCD* mRNAs and down-regulate RecB protein expression *in vivo*. Furthermore, when Hfq-mediated regulation was perturbed, less efficient suppression of protein noise and reduced DNA repair capacity were observed.

All in all, a molecular mechanism where Hfq fine-tunes RecB expression by inhibiting RecB translation is proposed. It is further suggested that this fine-tuning contributes to protein noise reduction and protects bacterial cells against the toxic consequences of too high RecBCD numbers.

Abstract

To preserve genome integrity, all living organisms have developed strategies to respond to chromosomal damage. One such response is the repair of double-strand breaks (DSBs), one of the most toxic forms of DNA damage. In *Escherichia coli*, DSBs are repaired via a homologous recombination pathway, which is initiated by a multifunctional enzyme, RecBCD. Primarily, RecBCD detects a double-strand end and degrades the chromosome until it encounters a short DNA motif known as a Chi site. Upon Chi site recognition, RecBCD activity is modified, resulting in the formation of a RecA-coated filament which is responsible for homology search.

Because of its important function in DSB repair, RecBCD is essential for accurate chromosome maintenance. On the other hand, overproduction of the enzyme has been observed to reduce DNA repair ability of bacterial cells. This double-natured phenomenon led to the idea that RecBCD expression needs to be maintained within an optimal range. Being mainly focused on the RecB subunit of the RecBCD complex, this work has addressed the following questions: how abundant is the enzyme within cells and how variable is its expression? Is there a regulation mechanism that maintains RecBCD expression at the optimal level? Using single-molecule microscopy techniques (smFISH and HaloTag-labelling), the low abundance of RecB subunits ($\sim$ 5 molecules per cell) was confirmed and weak *recBCD* transcription (less than one mRNAs per cell) was further established. With a simple stochastic model of gene expression, it was shown that while transcriptional outputs are noisy, cell-to-cell protein variability is lower than expected for an unregulated gene. Secondly, RecB mRNA and protein production was shown to be uncoupled in DNA-damage conditions: a decrease in *recB* transcription was compensated with more efficient translation upon DSB induction. Taken together, these observations point towards a post-transcriptional regulation mechanism. A post-transcriptional regulator, Hfq, was shown to bind *recBCD* mRNAs and the downregulation of RecB expression by Hfq was established *in vivo*. A molecular mechanism of RecB production whereby Hfq inhibits RecB translation was summarized in a stochastic model and further tested with Hfq mutants. A less effective noise reduction was detected when the Hfq-mediated regulation was perturbed. This is the first evidence of Hfq involvement in suppression of protein fluctuations in bacterial cells. Additionally, disruption of the uncovered mechanism was shown to reduce DNA repair capacity of *E. coli*, suggesting a potential phenotypical advantage of limitation of RecB numbers.

All in all, a novel Hfq-mediated regulation mechanism of RecB expression is presented in this thesis. Protein noise reduction and protection against the toxic consequences of high RecBCD numbers are proposed as potential phenotypical advantages of the uncovered fine-tuning mechanism.

Abbreviations

2D ..... two-dimensional
AIC ..... Akaike information criteria
Amp ..... ampicillin
ara ..... arabinose
BIC ..... Bayesian information criteria
bp ..... base pair
CI ..... confidence interval
Cipro ..... ciprofloxacin
CLASH ..... cross-linking, ligation and analysis of hybrids
CLIP ..... cross-linking immunoprecipitation
Cm ..... chloramphenicol
Ct ..... threshold cycle
CV (CV^2) ..... coefficient of variation (squared)
D_{KL} ..... Kullback-Leibler divergence
DNA ..... deoxyribonucleic acid
DSB(s) ..... double-stand break(s)
dsDNA ..... double-stranded DNA
FF ..... Fano factor
HR ..... homologous recombination
JF549 ..... Janelia Fluor 549
Km ..... kanamycin
LB ..... Luria broth
MIC ..... minimal inhibitory concentration
mRNA ..... messenger RNA
Nal ..... nalidixic acid
$NegBin$ (NB) ..... negative binomial
nt ..... nucleotide
OD_{600} ..... optical density at 600 nm
ODE ..... ordinary differential equation
PCR ..... polymerase chain reaction
PDF ..... probability density function
PMGR ..... plasmid-mediated gene replacement
$Pois$ ..... Poisson
RBP ..... RNA binding protein
RBS ..... ribosomal binding site
Rif ..... rifampicin

RIL-seq ..... RNA interaction by ligation and sequencing
RNA ..... ribonucleic acid
RNAP ..... RNA polymerase
RPM ..... reads per million
rRNA ..... ribosomal RNA
RT-qPCR ..... real-time quantitative PCR
SD ..... Shine-Dalgarno sequence
(sm)FISH ..... (single-molecule) fluorescent *in situ* hybridization
sRNA ..... small non-coding RNA
SSA ..... stochastic simulation algorithm
SSB(s) ..... single-strand break(s)
ssDNA ..... single-stranded DNA
TAMRA (TMR) ..... tetramethylrhodamine dye
TE ..... translational efficiency
TopoIV ..... topoisomerase IV
tRNA ..... transfer RNA
UTR ..... untranslated region
UV ..... ultraviolet light
v/v ..... volume per unit volume
WT ..... wild type
w/v ..... weight per unit volume

Contents

Chapter 1

Introduction

Noise is rather a characteristic than imperfection of the real world.

Inspired by Daniel Kahneman

1.1 Stochasticity in gene expression

Variability is an unavoidable component of life, particularly when considering the diversity and complexity of biological systems. For instance, the expression of the product from a single gene is a fundamental process of molecular biology, merging transcription and translation within the central concept of genetic information flow. Probabilistic nature of reactions underlying transcription and translation leads to cellular fluctuations in mRNA and protein abundance. These variations in gene expression are universal across all organisms, from viruses to higher eukaryotes and have been recognized as one of the major causes of phenotypical differences between cells [33, 79, 131, 166, 206].

In gene expression, transcription and translation events are described by coupled biochemical reactions [138, 170]. These reactions may evolve according to an anticipated scenario if the involved molecules are abundant. However, this might not be the case when there are just a few molecules of a particular species in a cell. Fluctuations in such low-copy molecules, particularly at the initial stages, may lead to significant propagation of noise within the network and prominently affect the downstream processes.

While low abundance of key enzymes is frequently seen as one of the major sources of noise, it provides only a partial explanation for the observed gene expression stochasticity within a population of cells. Another compound may arise from the essential cellular components, such as polymerases, ribosomes, proteases and amino acids, that may substantially vary among cells [145, 204]. However, because of simultaneous involvement in numerous reactions, these molecules might become a limiting factor, and thus, influence a probability of a particular reaction to occur.

Stochasticity in cell generation time has been recognized as an important contributor to biological noise [88, 173]. In fact, stochasticity in the duration of the cell cycle affects the timing of cell division and DNA replication events which inevitably leads to fluctuations in cellular content. Furthermore, in the presence of cell cycle variability, faster growing cells are over-represented in population snapshots [237]. Indeed, a fraction of cells with shorter division times increases over time in a growing cell population. Additionally, during cell division cellular components are separated between the daughter cells introducing fluctuations because of partitioning inequality or imperfect segregation. Taken together, cell cycle duration variability and molecule partitioning, which are usually considered as secondary sources of gene expression noise, can be influential compounds of non-genetic heterogeneity of a cell population [98].

Gene expression fluctuations are commonly separated into *intrinsic* and *extrinsic* noise [167, 227, 236]. These terms originate from a definition of a system versus its external environment in physics. Despite a consensus within the field of gene expression, disentangling of *intrinsic* versus *extrinsic* fluctuations within a system is highly dependent on a chosen definition of a *system* versus *environment* in each particular case.

An elegant experimentally-guided way of noise decomposition is based on a correlation analysis of two fluorescent reporters which are identically regulated but physically distanced on the chromosome of *Escherichia coli* cells [66]. In this case, the fluctuations correlated between these genes come from external sources, such as variability in RNA polymerases, ribosomes, transcriptional factors and other shared components (external noise), while the uncorrelated part arises from randomness of mRNA and protein copy numbers of the fluorescent reporter's network (internal noise). In these terms, Elowitz *et al.* set apart intrinsic and extrinsic fluctuations and demonstrated significant differences in cell phenotypes resulting from these sources of variability.

In addition to artificial circuits, stochasticity may play a crucial role in natural environments, particularly in the case of vital processes. Genetically identical cells, grown in the same environment, can have significantly various phenotypes as well as exhibiting very different behaviours. These distinctions are even more notable in the context of stress conditions such as an antibiotic exposure or a sudden environmental change. For instance, a subpopulation of bacteria persists an antibiotic treatment while the vast majority of cells do not survive [25]. Balaban *et al.* connected this observation to pre-existing heterogeneity in growth rates of single bacteria. In another example, Uphoff *et al.* demonstrated that under DNA alkylation damage, *E. coli* cells lacking the key protein, Ada, can delay the repair response for several generations. This was further shown to result in a fraction of cells with increased mutation rates [242]. Thus, phenotypic differences, caused by fluctuations among cells, can be profitable in some cases and harmful in others. These observations raised further questions of how cells control noise and what regulation mechanisms are utilized to reduce undesirable fluctuations.

Mechanisms controlling protein expression noise have been investigated in both theoretical and experimental studies over the last four decades. For instance, frequent transcription followed by inefficient translation was proposed as one of the most effective strategies to reduce protein fluctuations [111, 138, 188]. The reason behind this phenomenon can be explained by differences in time scales. Indeed, proteins typically have longer lifetimes than mRNAs, and thus, frequent transcription coupled with infrequent translation filters transcriptional stochasticity by averaging over a time interval (Fig. 1.1).

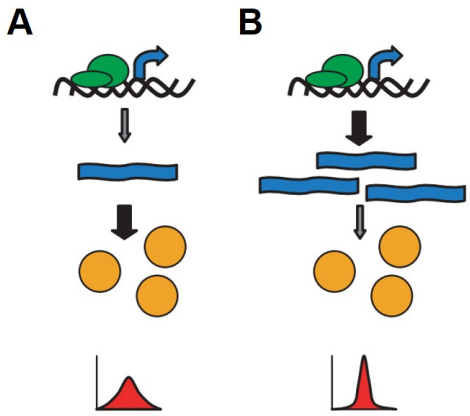

Figure 1.1: Comparison of protein noise. Infrequent transcription coupled with efficient translation leads to higher protein fluctuations (**A**) in comparison to frequent transcription and inefficient translation (**B**). The image was taken from Raser *et al.* [188].

Another powerful way of suppressing protein fluctuations is via executing a motif known as a negative feedback loop [13, 94, 236]. It was shown that autoregulation, where a protein inhibits its own transcription or translation, can reduce noise in protein numbers resulting in maintaining protein abundance within an optimal range [212, 226].

Mechanisms controlling noise are subject to evolutionary pressure. Phenotypical variability might be advantageous if, for example, environmental conditions suddenly change. Indeed, an adapted subpopulation of cells is more likely to arise from a noisy pre-existing heterogeneity of the entire population rather than a population with a deep-established trait. However, once newer characteristics have been selected, it might be preferable to decrease variability in order to assimilate an optimal trend. In other words, heterogeneity of the population was proposed to be advantageous upon stress and disadvantageous when cells are well adapted to the environment [175]. Thus, noise control may be seen as a powerful way to guide genetic assimilation during evolution.

1.2 Mathematical modelling of gene expression in bacteria

Mathematical modelling in biology has been increasingly used as an approach to test hypotheses arising from experimental outcomes. In relation to gene expression, there are two ways to describe behaviour of a system – deterministic and probabilistic. When species of the network are present in large abundance, the first approach provides a general description of species dynamics. As the most common way to characterize a complicated system, the deterministic approach might be misleading in a low-copy regime. Taking into account that many proteins in *E. coli* are present at less than 10 molecules per cell [233], a probabilistic approach needs to be implemented in order to realistically describe a gene expression network. For instance, fluctuation analysis can be particularly informative when comparing empirical data to theoretical predictions, as such interrelations might reveal hidden mechanisms of gene expression and regulation [92].

Stochastic models of gene expression were developed and widely used to study fluctuations underlying mRNA and protein production in prokaryotic and eukaryotic cells [29, 170, 190]. A probabilistic representation, known as *two-stage* or *mRNA-protein*, is a standard landmark of stochastic gene expression [167, 236]. Four biological processes underlie a classical version of the model: transcription of a gene by an RNA polymerase; translation of an mRNA by a ribosome; mRNA degradation; and protein decay. Generally, mRNAs and proteins, are considered as the key components of the network. If the variation in gene copy numbers (for example, as a result of replication) is considered, a gene is added as a third species. A bacterial cell is assumed to be a well-stirred system of biochemical molecules, the interaction of which is described with the chemical reaction approach [244]. In these terms, transcription, translation and degradation are typically modelled as Poisson processes with constant probabilistic rates. The schematic representation of the two-stage model is shown in Figure 1.2.

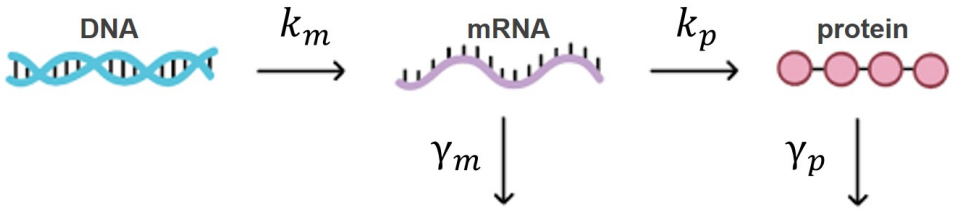

Figure 1.2: Two-stage model of gene expression. Four main processes underlie the model: (i) transcription of a gene by an RNA polymerase resulting in an mRNA molecule; (ii) translation of an mRNA by a ribosome leading to synthesis of a protein molecule; (iii) degradation of an mRNA; and (iv) protein decay. The rates of these reactions are constant and denoted as k_m, k_p, γ_m, and γ_p, respectively. The image was modified from Khan Academy.

Gene expression in bacterial cells allows further assumptions to be considered. First, based on a broadly accepted view of the exponential RNA degradation in bacteria [158, 203, 205], mRNA decay is described as a first order reaction with a constant rate γ_m. Secondly, proteins in bacteria were shown to be diluted exclusively as a result of cell growth and division [112, 233]. This process can be implicitly modelled by a first-order chemical reaction with a constant rate equal to the growth rate γ_p. Protein production is described with a chemical reaction with a constant rate k_p. Finally, transcription is usually described by one of the following ways: one- or two-state promoter model. For each particular case, the choice depends on the experimental evidence or hypothetical assumptions on the number of promoter states and rates of transitions between them.

Before delving into mathematical descriptions of the models, several general assumptions are worth discussing. First, a system requires stationary mRNA and protein distributions meaning that, although mRNA and protein numbers change in single cells, their distributions remain the same at the population level. Secondly, analysing a model in the low abundant regime eliminates fluctuations in polymerase, ribosome and protease levels. In modelling terms, it means that transcription, translation and degradation rates do not depend on concentrations of cellular enzymes, and hence, remain constant. Finally, variations (from cell cycle variability and from partitioning of mRNAs and proteins between cells as a result of cell division) are not explicitly considered. Under these assumptions, mathematical description of mRNA transcription is initially reviewed.

1.2.1 Transcriptional models: one- and two-state promoter

Previous studies showed that the promoter activity has a substantial impact on the distribution of mRNAs transcribed from the promoter [169, 195]. Therefore, stochastic models of transcription focus on the following questions: what determines the behaviour of a promoter? How does a promoter architecture influence the transcriptional output?

A good place to start answering these questions is to assume that a promoter has only one state or, in other words, is always active. It implies that mRNA molecules are produced independently with a constant probability per unit of time and (under assumption of an exponential mRNA degradation) results in a Poisson distribution for mRNA steady-state copy numbers [195]. From a biological viewpoint, this reflects the behaviour of a constitutive promoter and is often referred to as a one-state promoter model (Fig. 1.3A).

On the other hand, transcription of some genes in bacterial cells has been shown to occur in bursts [81, 219]. To describe this phenomenon, a model of transcription with bursty mRNA production has been proposed and further developed in the literature [105, 167, 173]. In the most general case, a promoter is typically assumed to have *active* (or 'on') and *inactive* (or 'off') states and mRNA dynamics is described by a telegraph process [77]. mRNA molecules are produced only when the gene is in the active state. The illustration of the two-state promoter model is shown in Figure 1.3B.

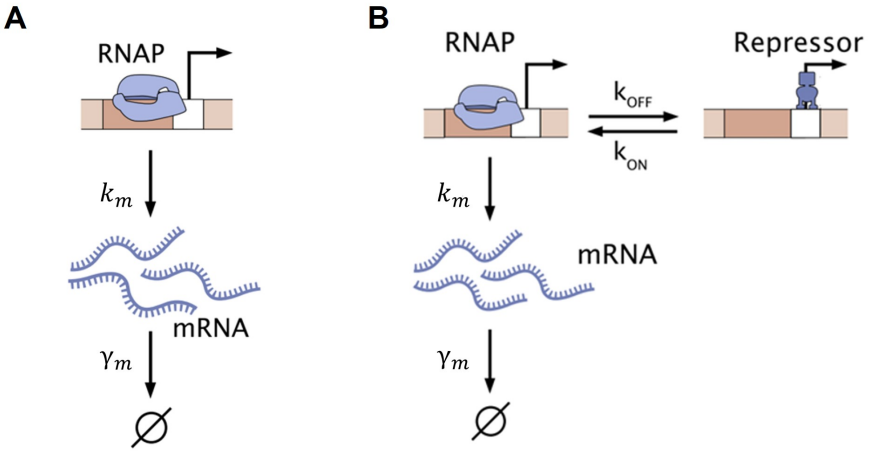

Figure 1.3: The models of transcription: one- and two-state promoter. In the one-state promoter model (**A**), a gene is always in active state and corresponding mRNAs are transcribed and degraded with constant rates k_m and γ_m, respectively. In the two-state promoter model (**B**), a gene switches between active and inactive states with rates k_{on} and k_{off} and mRNA molecules can be produced only when the gene is in the active state. The image was adapted from Sanchez *et al.* [195].

One-state promoter

As mentioned above, in the one-state promoter model, transcripts are stochastically produced from a gene copy with constant probability per time unit, k_m, and mRNA degradation occurs with constant probability per transcript and time unit, γ_m. The schematic of the one-state promoter model is presented below where G denotes a gene copy and M is an mRNA molecule:

$$\textit{Transcription}: G \xrightarrow{k_m} G + M$$
$$\textit{RNA degradation}: M \xrightarrow{\gamma_m} \varnothing$$

Notations with the uppercase letters such as G and M are used to refer to a single exemplified copy of a gene, mRNA or protein molecule, while the lowercase m denotes the variable with the number of transcripts in a cell. In order to understand how the probability $\widetilde{P}(m)$ of having m transcripts in cell changes over time, all potential contributions need to be considered. First, the influx comes from the events of either (i) transcribing an mRNA molecule when there are $(m-1)$ mRNAs in a cell or (ii) degrading an mRNA when there are $(m+1)$ transcripts. Following a similar reasoning, the decrease can be caused by a transcription or degradation event

from the system state with m mRNAs. It is worth noting that the total rate of mRNA degradation, as for a first order chemical reaction, is scaled by the number of mRNAs in a cell, and thus, is equal to $\gamma_m m$. These dynamics can be summarized by the following chemical master equation [244]:

$$\frac{d\widetilde{P}(m)}{dt} = -(k_m + \gamma_m m)\widetilde{P}(m) + k_m\widetilde{P}(m-1) + \gamma_m(m+1)\widetilde{P}(m+1) \tag{1.1}$$

Because the system will be further analysed only at steady-state, the time variable in probability notation has been intentionally omitted. Mathematically, the steady-state solution is defined by the following condition: $d\widetilde{P}(m)/dt = 0$ which means that probability distribution does not change over time. As previously shown, the steady-state solution of Equation 1.1 is a Poisson distribution for mRNA numbers per cell [195, 244]:

$$\widetilde{P}(m) = \frac{(k_m/\gamma_m)^m}{m!} e^{-k_m/\gamma_m} \tag{1.2}$$

For a Poisson distribution, its mean $\langle m \rangle$ and variance σ_m^2 are equal (Eq. 1.3). Because of this property, it is a common approach to use the Fano factor, FF, defined as variance of the distribution normalized by its mean, as a measure of variability or noise. Thus, for a random process the Fano factor is expected to be equal to one (Eq. 1.4).

$$\langle m \rangle = \sigma_m^2 = k_m \gamma_m \tag{1.3}$$

$$FF = \frac{\sigma_m^2}{\langle m \rangle} = 1 \tag{1.4}$$

From the biological perspective, one-state promoter model corresponds to constitutive expression. It is worth noting that this scenario can also be achieved when the number of promoter states is more than one, but switching between the promoter states is much faster than the transcription rate k_m. In other words, if RNA polymerase binding is the rate limiting reaction, a one-state promoter model approximates system behaviour relatively well.

Two-state promoter

When the transition between gene states is not necessarily fast in comparison to the rate of transcription, a more detailed model has to be analysed. A two-state promoter model is based on the fact that a promoter has active *'on'* and inactive *'off'* states and stochastically switches between them [170, 186]. During the active state, transcripts are produced with a constant rate k_m, whereas in the inactive state, the promoter is silent and no transcripts are synthesized.

The probability (per unit time) of a transition from the 'off' to 'on' state is defined as k_{on} while the opposite transition happens with the probability (per unit time) k_{off}. The schematic for one-state promoter can be extended by explicitly including two gene states (G, G^*) and the reactions corresponding to the transitions between them:

$$\begin{aligned} \textit{Gene activation} &: G^* \xrightarrow{k_{on}} G \\ \textit{Gene deactivation} &: G \xrightarrow{k_{off}} G^* \\ \textit{Transcription} &: G \xrightarrow{k_m} G + M \\ \textit{RNA degradation} &: M \xrightarrow{\gamma_m} 0 \end{aligned}$$

The probabilities of having m number of mRNAs in the active and inactive states, denoted as $\widetilde{P}_{on}(m)$ and $\widetilde{P}_{off}(m)$, can be described by a set of two coupled chemical master equations [170]:

$$\frac{d\widetilde{P}_{on}(m)}{dt} = -(k_m + \gamma_m m + k_{off})\widetilde{P}_{on}(m) + k_m\widetilde{P}_{on}(m-1) + \gamma_m(m+1)\widetilde{P}_{on}(m+1) + \\ + k_{on}\widetilde{P}_{off}(m) \tag{1.5}$$

$$\frac{d\widetilde{P}_{off}(m)}{dt} = -(\gamma_m m + k_{on})\widetilde{P}_{off}(m) + \gamma_m(m+1)\widetilde{P}_{off}(m+1) + k_{off}\widetilde{P}_{on}(m) \tag{1.6}$$

Here, Equation 1.5 describes how the probability of being in the state $\widetilde{P}_{on}(m)$ where (i) the promoter in active and (ii) there are m mRNA molecules present in a cell varies over time. The fraction of cells being in this state decreases as a result of one of the following events: transcription of a newer mRNA $k_m\widetilde{P}_{on}(m)$; mRNA degradation $\gamma_m m\widetilde{P}_{on}(m)$; or promoter deactivation $k_{off}\widetilde{P}_{on}(m)$. In contrast, this fraction may be increased as a result of one of two reactions. The first is transcription of an mRNA molecule from the state $\widetilde{P}_{on}(m-1)$ where the promoter is active and there are $(m-1)$ molecules in a cell. Secondly, transition from $\widetilde{P}_{on}(m+1)$ to $\widetilde{P}_{on}(m)$ can be achieved by degrading a transcript while the gene is in the active state. The last option activates the gene from the inactive state when there are m mRNA molecules present in a cell. Equation 1.6 can be broken down following similar argumentation with the only difference is that mRNAs cannot be produced from the inactive gene state.

The total probability of having m number of mRNAs in a cell, $\widetilde{P}(m)$, can be calculated as a sum of $\widetilde{P}_{on}(m)$ and $\widetilde{P}_{off}(m)$: $\widetilde{P}(m) = \widetilde{P}_{on}(m) + \widetilde{P}_{off}(m)$. The exact solution for total steady-state probability has been obtained using the generating function technique [170]:

$$\widetilde{P}(m) = \left(\frac{k_m}{\gamma_m}\right)^m \frac{\Gamma\left(\frac{k_{on}}{\gamma_m}+m\right)}{\Gamma(m+1)\,\Gamma\left(\frac{k_{off}+k_{on}}{\gamma_m}+m\right)} \frac{\Gamma\left(\frac{k_{off}+k_{on}}{\gamma_m}\right)}{\Gamma\left(\frac{k_{on}}{\gamma_m}\right)} {}_1F_1\left(\frac{k_{on}}{\gamma_m}+m, \frac{k_{off}+k_{on}}{\gamma_m}, -\frac{k_m}{\gamma_m}\right) \tag{1.7}$$

where $\Gamma(x)$ is the gamma function while ${}_1F_1(a,b,c)$ denotes a confluent hypergeometric function of the first kind.

As a result of the complex form of the exact solution (Eq. 1.7), it might be computationally challenging to apply for data analysis and inference. Therefore, a convenient and biologically reasonable approximation is often considered where the rate of deactivation is much larger than the activation rate and also larger than the mRNA degradation rate ($k_{off} \gg k_{on}$ and $k_{off} > \gamma_m$). This limiting case corresponds to short and infrequent bursts of mRNA production and its solution converges to a gamma distribution which is characterized by two parameters, k_{off}/k_m and k_{on}/γ, [186]:

$$\widetilde{P}(m) = \frac{k_{off}}{k_m} \frac{1}{\Gamma\left(\frac{k_{on}}{\gamma}\right)} \left(\frac{k_{off} m}{k_m}\right)^{\frac{k_{on}}{\gamma}-1} e^{-\left(\frac{k_{off} m}{k_m}\right)} \tag{1.8}$$

Because the gamma distribution is continuous, its discrete analogue, the negative binomial distribution [68], is frequently used for comparison with experimental data [217]. For convenience, an additional term for mRNA bursts has been introduced and defined as $b = k_{on}/(k_{on}+k_{off})$.

While the entire distribution of mRNA molecules per cell provides valuable insights into potential regimes of the system, experimental data can be often inferred with the first two moments of the distribution, mean and variance. These moments have been computed elsewhere by either the generating function approach or by integrating the master equation [109, 186, 231]:

$$\langle m \rangle = \frac{k_m}{\gamma_m} \frac{k_{on}}{k_{on}+k_{off}} \tag{1.9}$$

$$FF = \frac{\sigma_m^2}{\langle m \rangle} = 1 + \frac{k_{off}}{k_{on}+k_{off}} \frac{k_m}{\gamma_m + k_{on} + k_{off}} \tag{1.10}$$

As seen from Equation 1.10, the existence of several promoter states increases the Fano factor. In other words, the output of transcription becomes noisier than expected for a constitutive promoter (in comparison to Eq.1.4). Moreover, within the limit of $k_m \ll k_{on} + k_{off}$, the Fano factor is approaching one. This approximation is known as a fast promoter limit and corresponds to quick switching between the promoter states. Intuitively, this scenario mimics constitutive transcription. Indeed, if gene activation-deactivation transition is fast, the rate of transcription k_m becomes the main one to define mRNA synthesis.

Biological interpretation of the transcription models is worth discussing. Generally, a promoter in the inactive state does not necessarily need to be physically inhibited by a repressor molecule. However, if the behaviour of the gene displays strong bursty transcription, this might be a hint of a possible mechanism which keeps the promoter in an inactive state.

To describe more complicated promoter architectures, multiple-state promoter models have been proposed [266]. The general idea is based on the transitions between more than two promoter states. Additionally, more advanced models accounting for cell cycle and DNA replication, molecule partitioning as a result of cell division [28, 173], RNA polymerase pausing [41], chromosome supercoiling effect [37], distinction between nascent and mature mRNAs [73, 260] have been described in the literature over the last decade. While representing the process of transcription more realistically, detailed models are usually complicated for a stochastic analysis and frequently investigated with an approach based on simulations.

1.2.2 mRNA-protein model

Transcriptional models have been further extended by including protein production. Specifically, the probability of translation per time unit and per transcript is defined as k_p which leads to $k_p m$ as the total rate of protein production. Considering that proteins in bacterial cells are mostly stable and do not have an active degradation mechanism, protein decay can be modelled as dilution as a result of cell division as $\gamma_p p$ [167]. The most common description of the entire protein synthesis model is known as *mRNA-protein* or *two-stage* model. Here, its version with bursty transcription is reviewed.

Following the notations above, the uppercase letter P is used to refer to a single protein molecule in the reaction scheme and the lowercase p denotes the variable with the number of protein copies in a cell. An exemplified mRNA burst size is represented as n and sampled from a geometric distribution with a mean size b [170, 186, 195]. Thus, the entire reaction network for the two-stage model with burst transcription can be schematically presented as following:

$$\textit{Transcription}: G \xrightarrow{k_m} G + nM$$
$$\textit{RNA degradation}: M \xrightarrow{\gamma_m} \varnothing$$
$$\textit{Translation}: M \xrightarrow{k_p} P + M$$
$$\textit{Protein dilution}: P \xrightarrow{\gamma_p} \varnothing$$

The dynamics of the mRNA and protein average levels ($\langle m \rangle$ and $\langle p \rangle$) can be described by a set of the following rate equations (as elsewhere, for example [167]):

$$\frac{d\langle m \rangle}{dt} = b\,k_m - \gamma_m \langle m \rangle \tag{1.11}$$

$$\frac{d\langle p \rangle}{dt} = k_p \langle m \rangle - \gamma_p \langle p \rangle \tag{1.12}$$

The steady-state solutions for mean mRNA and protein numbers can be found by imposing the conditions $d\langle m \rangle / dt = 0$ and $d\langle p \rangle / dt = 0$ to the rate equations (Eq.1.11 and Eq.1.12):

$$\langle m \rangle = b\frac{k_m}{\gamma_m} \qquad \langle p \rangle = b\frac{k_m k_p}{\gamma_m \gamma_p} \tag{1.13}$$

As a result of the linearity of the system, the exact mRNA and protein variances can be found from the chemical master equation using, for example, van Kampen's Ω-expansion method [64, 244]:

$$\sigma_m^2 = (1+b)\langle m \rangle \qquad \sigma_p^2 = \langle p \rangle + \frac{k_p^2(1+b)\langle m \rangle}{\gamma_p(\gamma_m + \gamma_p)} \tag{1.14}$$

Noise can be estimated in several ways. Two most common measures of noise are Fano factor (FF) and coefficient of variation (CV or CV^2). The Fano factor, defined as the variance over average, has been mentioned in relation to transcriptional models and is equal to one for a random process. It is a convenient measure of fluctuations when comparing to the univariate Poissonian process as, in this case, the variance is typically expected to be proportional to the mean. For a multivariate process, the variance over squared average is considered as a more reasonable measure of noise as it allows for more transparent separation of sources of noise coming from different variables [167]. Nevertheless, the mathematical expressions for both ways of mRNA and protein noise estimation are given below [167, 171, 207].

$$FF_m = 1 + b \tag{1.15}$$

$$FF_p = 1 + \frac{(1+b)k_p}{(\gamma_m + \gamma_p)} \tag{1.16}$$

$$CV_m^2 = \frac{\sigma_m^2}{\langle m \rangle^2} = \frac{1+b}{\langle m \rangle} \tag{1.17}$$

$$CV_p^2 = \frac{\sigma_p^2}{\langle p \rangle^2} = \frac{1}{\langle p \rangle} + \frac{1+b}{\langle m \rangle} \frac{\gamma_p}{\gamma_m + \gamma_p} \tag{1.18}$$

Finally, the analytical probability distribution of protein numbers under assumption of long-lived proteins relative to mRNA lifetimes was derived by Shahrezaei and Swain [207]. More sophisticated models have been built on this ground representation by including specific characteristics of a pathway, effects of replication and gene dosage, such sources of fluctuations as variation in cell cycle, growth rates, partitioning of molecules as a result of cell division and other [76, 94, 116, 207, 218].

1.3 Stress responses in *Escherichia coli*

Escherichia coli face various types of stress and environmental changes during the lifetime. To combat these, cells have developed numerous stress response programs which typically activate particular pathways controlled by specific regulators [191]. The most common stress response systems are reviewed below.

Oxidative response

Oxidative damage arises as an inevitable consequence of oxygen metabolism through formation of chemical compounds known as reactive oxygen species (ROS). The most common forms of ROS, superoxide radicals and hydrogen peroxide, are highly toxic by-products of oxygen utilization that have been shown to damage cellular processes via the oxidation of DNA nucleobases, formation of single-strand DNA breaks, interference with RNA transcription, inhibition of translation, and post-translational modifications [71]. In order to respond to the oxidative damage, bacteria have developed strategies to neutralize ROS. For instance, the transcription response to the accumulation of hydrogen peroxide in cells is regulated by a transcription factor OxyR and a small RNA OxyS. The oxidative stress caused by superoxide radicals is regulated through a two-stage SoxRS pathway: activation of SoxR leads to the induction of expression of SoxS which further promotes the expression of oxidative stress specific proteins [46].

Acid stress response

While *E. coli* cells preferentially grow in neutral environments with pH = 7, they have been demonstrated to survive extremely low pH (pH = 2.5) [82]. The ability of bacterial cells to resist to the extremely acidic environments is granted by activation of several acid resistance mechanisms: oxidative system AR1, glutamate-dependent system AR2, arginine-dependent system AR3, and lysine-dependent system AR4. While being regulated differently among each other, the amino-acid dependent systems mainly rely on the activity of inducible decarboxylases which are specific for each amino-acid. The most efficient acid resistance pathway AR2 is regulated by decarboxylases GadA and GadB, and antiporter GadC which activity is triggered by acidic pH [53].

Osmotic stress response

Another physiological challenge for bacterial cells is posed by significant changes in osmotic pressure of external environments. While bacterial membrane is permeable for water molecules, most solutes are not able to freely diffuse through. This results in rapid changes of solute concentrations within a bacterial cell when there is a sudden increase or decrease in the external osmolarity [189]. A passive response of *E. coli* cells is to lose or gain water and osmotic pressure on the cell wall. This first physiological change in turns triggers an active osmoregulatory response that depends on the upshift or downshift in the osmolarity.

After a shift of the external medium to a higher osmolarity, potassium uptake is upregulated that leads to accumulation of such solutes as potassium, proline, glycine betaine. As amount of these osmolytes increases in the cytoplasm, water influx resumes followed by cell volume increase and osmotic pressure recovery [176, 177]. In contrast, a hypoosmotic shock causes fast water influx into a cell leading to cell volume expansion. As inner membrane tension increases, mechanosensitive channels are activated to export nonspecific solutes and to avoid cell lysis [36, 40]. As the result of this active response, cell volume and osmotic pressure recover to the initial values. Osmotic pressure within a cell significantly influences interactions between cellular components and their dynamics thus it is important to understand how cells response to such stress [258].

Envelop stress response

Various environmental changes can lead to the perturbations of bacterial envelop, a compartment that constitutes a permeability barrier between the cell and its surroundings. In order to maintain and protect the envelop, *E. coli* cells execute two main signal transduction pathways: via Cpx and σ_E [185]. The two-component system CpxAR encodes the membrane localized sensor histidine kinase CpxA and the response regulator CpxR. CpxA is induced by accumulation of misfolded proteins as the result of envelope alterations [54]. This further stimulates phosphorylation of CpxR which controls the transcription of target genes such as protein folding and degrading factors. The σ_E response is specific to the alterations in the outer membrane protein biogenesis and mediated by an inner membrane antisigma factor RseA, a periplasmic protein RseB and proteases DegS and YaeL [10].

Heat- and cold-shock responses

Heat-shock and cold-shock allow cells to adapt to changes in external temperature. σ^{32} is the heat shock promoter-specific subunit of RNA polymerase which is positively controlled at the level of transcription. The σ^{32} response activates chaperons, proteases and other heat-shock proteins including DnaK, DnaJ, GrpE, GroEL and GroES [20]. Induction of these proteins mainly contributes to the heat shock response by removing unfolded cytoplasmic proteins which accumulate upon the increase in the temperature [146].

A downshift to the low temperature leads to such structural and functional changes such as rigidification of the membrane, protein folding, DNA conformation, inhibition ribosome assembly and activity [255, 267]. Once a cold shock occurs, bacteria slow down the growth and decrease regular protein synthesis production in order to activate synthesis of cold shock proteins (Csp) which are responsible for a short-term response as well as long-term adaptation to low temperatures [255].

Stringent and general responses

Stringent response, also known as amino acid starvation response, is triggered upon depletion of nutrients in the medium. Bacteria respond to the nutritional stress by quick reduction of cell growth through the inhibition of RNA synthesis and translation [225]. It is regulated by the derivatives of GTP and GDP, guanosine pentaphosphate (pppGpp) and tetraphosphate (ppGpp), which sense the change in amino acid availability. The synthesis of (p)ppGpp is performed by the enzymes RelA and SpoT [181]. (p)ppGpp further upregulates expression of the genes encoding amino acid uptake and biosynthesis resulting in the adjustments of gene expression and physiology [181]. As one of the major regulators in bacteria, (p)ppGpp takes part in the control of many cellular processes such as metabolism, adaptation, persistence, survival, cell division, motility, biofilm formation, and virulence.

Another stress response is known as general response and regulated via an alternative sigma factor RpoS [27]. While RpoS translation is inhibited during exponential growth, the level of RpoS is rapidly accumulated as cells enter stationary phase. RpoS has been shown to control the expression of 10% of *E. coli* genes [254]. When RpoS levels are elevated, many RpoS-dependent genes are activated and bacteria become resistant to a wide range of stresses.

Crosstalks between stress responses

In fact, many stress response pathways are closely connected. The major overlap is between the (p)ppGpp-induced stringent response and the general response. (p)ppGpp has been shown to be positively correlated with RpoS suggesting that these two pathways are parts of a more global response [27]. Additionally, RpoS is induced at low temperatures and high osmolarity what suggests the similarity of adaptation to cold shock and osmolarity upshift. There is also the evidence of RpoS involvement in the acid resistance of cells in the stationary phase. Interestingly, both heat-shock and cold-shock responses are regulated via sensing of RNA folding. After heat shock, unfolding of *rpoH* mRNAs facilitates translation of σ^{32} through a feed-forward mechanism [161]. Similarly, upon cold shock, the RNA chaperones Csps underlie the global regulation of mRNA metabolism and translation. Such synchronized alignment of response programs to various perturbations suggests common strategies to cope with stresses and the intertwined nature of stress responses.

1.4 DNA damage and repair

During cell lifespan, the stability of genome is constantly threatened by external DNA-damaging agents as well as due to internal cellular processes such as metabolism and replication. To preserve genome integrity, cells have evolved mechanisms to detect different lesions such as DNA cross-links, insertions, deletions, mismatches, single-strand gaps and double-strand breaks. While a particular repair pathway is undertaken for each type, the first step of the DNA damage response is general and known as SOS response.

The SOS response is an inducible system which is activated upon DNA damage and controlled by two key proteins: a transcriptional repressor LexA and an inducer RecA [103]. When cells grow normally, LexA dimers repress transcription of the genes belonging to the SOS regulon by binding to a specific sequence within their promoter region. Once DNA damage is detected, RecA proteins bind to single-stranded DNA (ssDNA), forming a nucleoprotein filament (RecA-ssDNA). This filament promotes self-cleavage of LexA and further derepression of SOS-response genes [135]. These genes include error-prone polymerases, excision repair protein, translesion DNA replication enzymes, cell division inhibitor. Importantly, *recA* and *lexA* belong to the SOS regulon allowing for fast activation and shutting down of the induction signal. Furthermore, not all SOS genes are induced at the same time and to the same extent. Such partial activation of the response allow for precise synchronization between the repair resources and damage level. High-fidelity repair pathways are induced first while less accurate mechanisms involving error-prone polymerases PolII (*polB*), PolIV (*dinB*), and PolV (*umuC*, *umuD*) are activated only upon persistent DNA damage [135, 178].

One of the most common types of DNA damage are single-stranded DNA (ssDNA) gaps which can happen during DNA replication. In *E. coli* such ssDNA gaps are repaired via RecFOR and homologous recombination pathways. The RecF, RecO and RecR mediate the loading of RecA protein onto ssDNA which further promotes the homology search and associated DNA strand exchange [156].

Mismatch repair (MMR) is another accurate pathway which is primary specialized on the correction of DNA base-base mismatches and small nucleotide insertion/deletion mispairs generated during replication and recombination events. In order to repair a mismatch, two functions have to be performed: recognition of the mismatch and identification of the newly synthesised DNA strand with the mistake. MutS, MutL and MutH are the key players in the MMR mechanism. MutS and MutL are responsible for mismatch recognition and recruitment of MutH [150]. Shortly after DNA replication, the daughter strand is temporarily unmethylated what allows for identification of the newly synthesized strand from the parental one by MutH. Upon its recruitment and activation by MutS and MutL, MutH incises the unmethylated daughter strand providing the initiation nick for mismatch excision [121]. Thus, MMR pathway enables to correct DNA preventing propagation of mutations in further generations.

While MMR can detect and repair only mismatched base pairs, another pathway known as nucleotide excision repair (NER) is responsible for removing and correcting bulky lesions in DNA such as, for example, thymine dimers. NER is regulated by the UvrABC endonuclease enzyme complex. A UvrA-UvrB complex recognizes DNA distortions caused by thymine dimers. Then, the UvrA loads UvrB onto a damaged region, then UvrC binds to UvrB, resulting in the UvrBC-DNA incision complex. UvrD, DNA polymerase I, removes the excised segment and fills in the resulting gap. The remaining nick is finally connected by DNA ligase [154].

1.4.1 Double-strand breaks

Double-stranded DNA (dsDNA) breaks (DSBs), as simultaneous interruption of both chromosomal strands, are considered as one of the most toxic types of DNA damage [50]. If unrepaired, a DSB results in a lethal outcome and, if repaired incorrectly, might lead to accumulation of deleterious mutations and genomic rearrangements [259].

Double-strand breaks can be caused by various exogenous sources, such as ionizing radiation, mechanical stresses and DNA-damaging agents [140]. An example of the latter is a clinically-relevant antibiotics class, quinolones (in particular, ciprofloxacin). Quinolones target bacterial topoisomerases TopoIV by trapping them onto DNA resulting in formation of irreversible topoisomerase-DNA cleavage complexes [179]. Because topoisomerases are primarily responsible for removing DNA supercoiling, the functionality of these enzymes is essential for DNA replication and cell division. In Gram-negative bacteria, such as *Escherichia coli*, ciprofloxacin specifically acts on gyrase and TopoIV enzymes by trapping them on DNA and preventing DNA strand ligation [113, 115]. This scenario results in a topoisomerase-DNA cleavage complex that blocks the DNA replication machinery and eventually leads to the formation of a double-strand break.

In addition to external stimuli, DSBs can also occur accidentally as a result of endogenous metabolic reactions and replication events during normal cell cycle [140]. For instance, unrepaired single-strand breaks (SSBs) lead to DSBs of one of the chromosomes during DNA replication. Indeed, the progression of replication machinery over a nick stops the replication fork and leaves one broken chromosome [57]. Another internal source of DSBs is known as *fork reversal* and may emerge during DNA replication. In this case, the two newly replicated chromosomal strands anneal together what results in a double-stranded DNA end [147]. Additionally, replication fork breakage in the chromosome terminus region has been recently shown to result in spontaneous DSBs in 18% of wild-type *Escherichia coli* cells per each generation [215]. Therefore, in order to preserve genetic integrity, double-stranded DNA breaks have a supreme priority to be promptly detected and faithfully corrected.

Across various organisms, two major DSB repair pathways have been developed: homologous recombination (HR) and non-homologous end joining (NHEJ). The first one is based on the presence and utilization of an intact copy of the chromosome as a template for repair, while NHEJ involves modification of the broken DNA strands and its direct ligation. In *E. coli*, DSBs are repaired exclusively via the homologous recombination pathway [230].

Although DSB repair via homologous recombination varies among the organisms, the main steps of the pathway are conserved. Once a dsDNA break occurs, the DNA ends are detected by MRN or MRX protein complex in eukaryotes or yeasts, respectively. Further resection of 5' ends is performed in two stages. First, Sae2 gets recruited by MRN/X and these two proteins degrade the 5' ends to form 3' single-strand DNA overhangs. Secondly, the resection is taken over by the helicase Sgs1 and nucleases Exo1 and Dna2 [17, 187]. Once the 3' single-strand DNA ends are created, the RPA proteins bind to the ssDNA regions. Next, Rad51 forms nucleoprotein filament on the ssDNA which further searches for a similar intact DNA sequence in a cell. Once such is found, the Rad51-ssDNA nucleoprotein invades into the DNA region and uses it a template for the repair. DSB repair via homologous recombination in *E. coli* is described below.

1.4.2 Functions of RecBCD enzyme

The multi-subunit nuclease/helicase enzyme, RecBCD (also known as ExoV), enzyme plays a key role in initiating homologous recombination pathway. RecBCD detects a double-stranded DNA (dsDNA) end and degrades both chromosomal strands. Upon recognition of a short DNA sequence, known as *chi* (χ) site (5'-GCTGGTGG-3'), the RecBCD complex stops cleaving the 3'-end strand and initiates loading of the RecA protein onto the 3' single-stranded (ssDNA) overhang [222, 257]. This leads to the formation of a RecA-ssDNA nucleoprotein filament [19], which catalyses homology search, strand exchange and branch-migration [147]. The initial stages of the homologous recombination pathway is illustrated in Figure 1.4.

The crystal structure of RecBCD (bound to DNA) has significantly contributed to the understanding of its functions [214]. The enzyme consists of three subunits: RecB, RecC and RecD (Fig. 1.5). RecB has been shown to have 3'-to-5' helicase and nuclease activity while RecD is reported to be a 5'-to-3' helicase [58]. The RecC subunit is responsible for interaction with χ sites [87]. Moreover, multiple mutational studies have shown that *recB* and *recC* genes are essential for DNA repair whereas *recD* mutants are recombination proficient [117, 256]. Taken together, these observations highlight the importance of RecB and RecC subunits for faithful DNA repair.

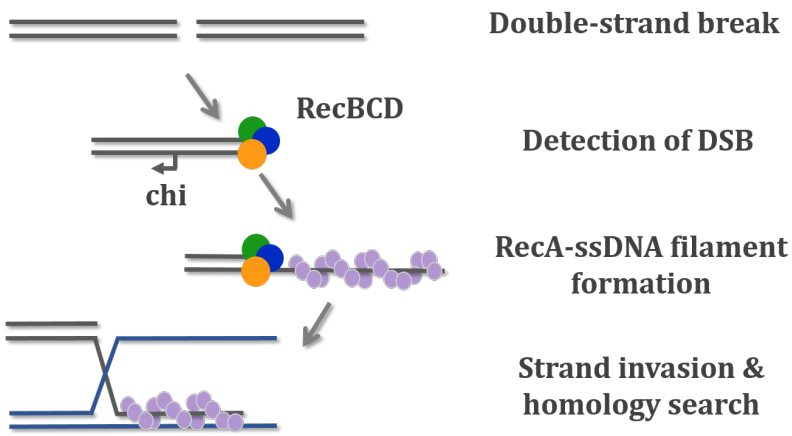

Figure 1.4: The initial stages of the homologous recombination pathway in *Escherichia coli*. RecBCD enzyme detects a double-stranded DNA end and starts processing both chromosomal strands till an eight base pair sequence, chi site. After chi recognition, the RecBCD complex continues cleaving 5'-end only and initiates loading of RecA onto 3' overhang. This RecA-ssDNA filament is further responsible for strand invasion and homology search.

Apart from the key role in DSB repair, RecBCD performs a protective function by degrading linear foreign DNA fragments that might have entered a bacterial cell. Extraordinarily fast and processive DNA degradation by RecBCD was demonstrated by both bulk assays and single-molecule studies *in vitro* [31, 129, 192]. Consistent with those measurements, RecBCD was shown to process the chromosome up to $>\sim$100 kb with the speed of $\sim$1.6 kb/s in live bacteria [257]. Thus, by being a powerful DNA exonuclease and an efficient recombinase, RecBCD enzyme combines different functions, such as DNA degradation and DNA repair.

Interestingly, RecBCD overproduction was demonstrated to result in the impairment of DNA repair and decrease of cell viability upon UV damage [55, 57]. This observation was proposed to be a result of increased chromosomal degradation and suggested that too many molecules of RecBCD enzyme impose a threat for cells upon DNA damage exposure. Along this line, recombination-deficient *recA* mutants have been demonstrated to display pronounced DNA degradation by RecBCD and have a high fraction of anucleate cells [216, 263]. Taken together with the necessity for DNA DSB repair [67, 147, 256], it leads to the hypothesis that there may be a regulation mechanism controlling RecBCD expression in bacterial cells.

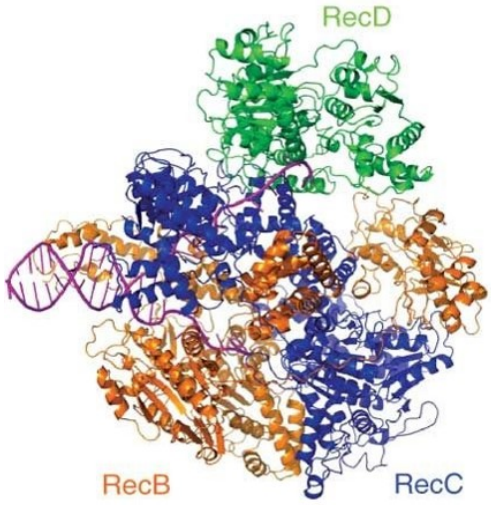

Figure 1.5: Cristal structure of RecBCD enzyme bound to DNA. The complex consists of three subunits: RecB, RecC, and RecD, which are shown in orange, blue and green, respectively. DNA is colored in magenta. The figure was taken from Singleton *et al.* [214].

1.4.3 RecBCD expression

While the cellular functions, enzymatic activities and crystal structure of RecBCD have been extensively studied, the details of its expression remain unclear. In *E. coli*, the RecBCD subunit genes are encoded by two operons located in the same locus of the chromosome (Fig. 1.6). One operon controls the expression of RecC whilst the other is polycistronic and encodes PtrA, RecB and RecD [48, 89]. One study has reported an additional internal promoter controlling *recD* expression [16]. However, the interpretations of *recD* polypeptide synthesis is troublesome as the results of protein synthesis gel where different versions of *recD* gene were expressed from plasmids (i) show no visible difference in RecD expression between expression of the entire gene (with a potential promoter region) and the truncated version (without the promoter region) ([16], Fig.4); (ii) the gel electrophoresis result with RecD expression was not performed on one gel but instead was run on two separate gels and aligned at the position of RecD afterwards. That could assidentially led to misalignment of RecD with a nearby protein ArgA ([16], Fig.4). Altogether, this makes the conclusion on a separate promoter for the *recD* gene is not entirely convincing.

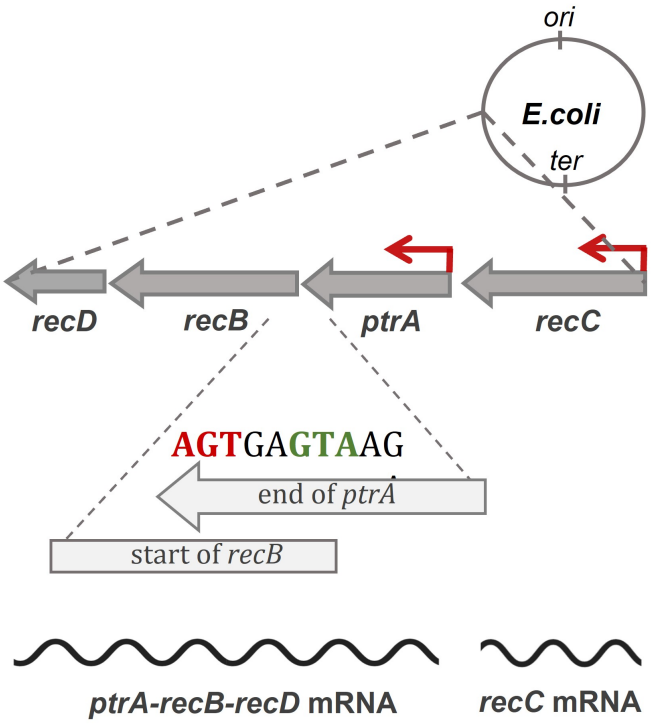

Figure 1.6: *recBCD* locus and its position in *Escherichia coli* chromosome. *recBCD* genes are encoded by two operons: *recC* is regulated by one promoter while polycistronic operon *ptrA-recB-recD* is controlled by another promoter. The overlap between the reading frame sequences of *ptrA* and *recB* genes are shown. The start codon of *recB* gene (highlighted in green) is located within the open reading frame of *ptrA* (*ptrA* stop codon 'TGA' is highlighted in red).

The *ptrA* gene, encoding a periplasmic protease (protease III), has been inserted between *recC* and *recB* [49], and thus, interrupted the usual order, *recC-recB-recD*. The event of *ptrA* insertion happened in the phylogenetic tree upon separation of the *Buchnera* specie. Similar to *Escherichia coli*, closely related species *Shigella dysenteriae*, *Salmonella enterica*, *Klebsiella pneumoniae*, *Yersinia enterocolitica*, *Photorhabdus asymbiotica*, and *Edwardsiella tarda* have *ptrA* insertion between *recC* and *recB-recD*. In contrast, *Buchnera aphidicola*, *Vibrio cholerae* and further species such as *Psychromonas arctica* and *Shewanella algae* have the usual order of *recC-recB-recD* locus (Fig. 1.7). Presumably, these local genetic rearrangements have led to the separation into two operons, *recC* and *ptrA-recB-recD*, that are controlled by independent promoters [61, 196].

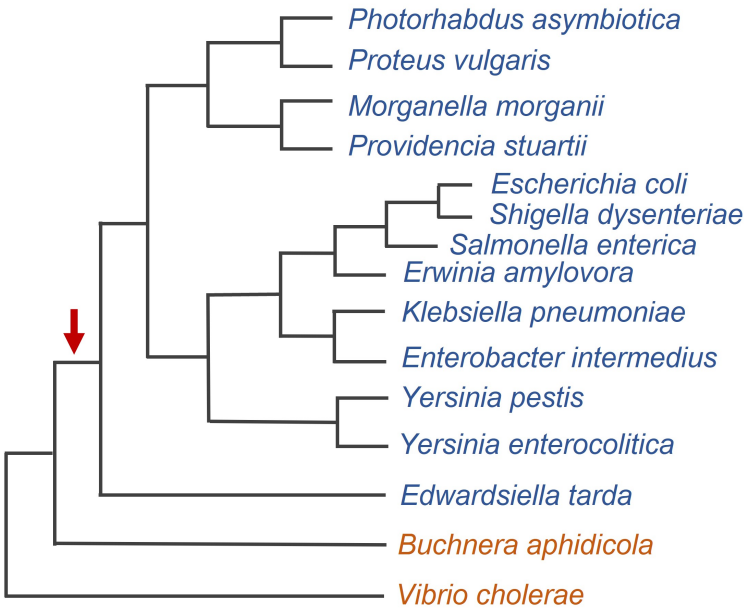

Figure 1.7: Insertion of *ptrA* gene within the *recC-recB-recD* operon shown on the phylogenetic tree of members of the family Enterobacteriaceae. The evolutional origin of *ptrA* insertion (red arrow) has been traced based on the phylogenetic analysis performed here [49]. Bacterial species before and after *ptrA* insertion are shown in orange and blue, respectively.

It is worth mentioning that there is no evidence of the protease III being directly involved in the DNA repair response in *E. coli*. In fact, the protease III has the only known function of degradation of small peptides and it has been found in the periplasmic space of a cell [60, 228]. However, as a result of the polycistronic organisation of the *ptrA-recB-recD* operon, *ptrA* can significantly affect transcription and translation of the downstream genes. First, as no consensus promoter has been identified in front of *recB* [74], its expression has been reported to be controlled by the promoter upstream of *ptrA* [48, 89]. Secondly, the ribosome binding site (RBS) of *recB* is located within the coding sequence of *ptrA* suggesting a potential translational coupling mechanism between these genes. A phenomenon of translational interdependence for the genes encoded within the same polycistronic transcript has been previously observed for bacterial operons [165, 202, 139, 130]. Two main mechanisms of translational coupling have been proposed in the literature. The first one is based on unfolding of the secondary structure of the downstream mRNAs [139, 128]. For instance, in the *atp* operon a stem loop is formed at the end of the upstream gene (*atpH*) which inhibits translation of the downstream gene (*atpA*). However, when *atpH* is translated, a ribosome translocation resolves the secondary structure of the transcript making RBS for the *atpA* gene available [139, 128]. In another translational coupling mechanism, termination of translation of the upstream gene is coupled to the initi-

ation of translation of the downstream gene due to the overlap or close proximity of the second gene's RBS and the first gene's reading frame. If the RBS of the second gene is placed near the termination of the upstream gene, ribosome scanning model has been suggested [11]. A ribosome completes translation of the upstream gene and, without falling off the mRNA, gets directly recruited to the RBS of the downstream gene [202]. Location of the second gene's RBS before the stop codon of the upstream gene, which is the case for the *ptrA-recB* transcript, has even stronger evidence for interdependence of translational efficiencies between the genes of a polycistronic mRNA. Since the RBS of *recB* gene is located upstream of the stop codon of *ptrA*, it strongly suggests translational coupling between these genes 1.7.

Previous attempts to explore RecBCD expression upon DNA damage conditions did not reveal any evidence of transcriptional regulation [90, 91]. Although an aberrant version of LexA-binding sequence upstream of the promoter of *ptrA* gene has been found, no evidence of transcriptional activation upon a global SOS response to DNA damage was detected [249]. Therefore, it seems to be unlikely that the *recBCD* genes take part in the bacterial SOS response. No studies have reported on investigation of RecBCD expression at the post-transcriptional or translational levels.

Finally, the low abundance of the RecBCD molecules has been discussed in the literature. In earlier populational studies, RecBCD abundance was estimated at only $\sim$10 molecules per cell [91, 234] and was recently corrected by single-molecule measurements to an even lower quantity ($\sim$5 molecules per cell) [119]. Such small numbers are known to be subjected to high relative fluctuations which, in turn, might result in large cell-to-cell variability [66, 168, 242]. Although those fluctuations might not play a significant role during normal cell growth, they may become crucial and determine cell fates upon stress conditions, environmental change and DNA damage. Therefore, for vital processes, a regulation mechanism may be needed to control gene expression noise of low-abundant components involved in a pathway.

1.5 Post-transcriptional regulation in bacteria

1.5.1 RNA-binding proteins

Traditionally, bacterial genes were believed to be mainly controlled at the level of transcription. However, over the last two decades prokaryotic RNA-binding proteins (RBPs) and small non-coding RNAs have been shown to be extensively involved in co- and post-transcriptional regulation of gene expression during stress responses and adaptation to environmental changes [47]. More specifically, post-transcriptional regulators have been demonstrated to control transcript stability and translation efficiency by a variety of molecular mechanisms such as altering accessibility of a target ribosome binding site (RBS), modulating termination of transcription, and facilitating or preventing RNA degradation by ribonucleases [162, 243].

Several RBPs have been identified as global regulators as a result of their essential roles in major cell responses and metabolic pathways. For instance, carbon storage regulator, CsrA, affects glycogen metabolism, bacterial motility and biofilm formation [102, 182]. Another global RNA chaperone, ProQ, initially discovered as a protein involved in proline uptake regulation, has been recently shown to control virulence and host immune pathways in *Salmonella enterica* [95, 208]. The major cold-shock RBP, CspA, regulates protein expression at low temperatures mainly via destabilization of mRNA targets [23]. Among all global post-transcriptional regulators, Hfq is one of the most studied in *Escherichia coli* [62, 107, 241].

1.5.2 Global regulator Hfq

An Sm-like hexameric ring-shaped protein Hfq, also known as HF-1 (Host Factor I), was originally discovered as a host factor for replication of the bacteriophage Qβ in *E. coli* [72]. However, the pleiotropy of *hfq* mutants pointed to its involvement in many other metabolic and stress-response pathways. When the mechanisms by which Hfq acts on its RNA targets were still unclear, Hfq-facilitated interaction between a small RNA, DsrA, and its target mRNA, *rpoS*, was demonstrated with a real-time fluorescence resonance energy transfer (FRET) assay [18]. Additionally to the modulation of base pairing between DsrA and *rpoS*, Arluison *et al.* showed the role of Hfq as an RNA chaperone. To date, Hfq is known to regulate a vast number of RNAs through different molecular mechanisms [97, 247].

Hfq mainly facilitates intermolecular base pairing between sRNAs and target mRNAs which typically results in altering mRNA stability and/or mRNA translation [42, 107, 162, 241, 243]. Specifically, Hfq can modulate initiation of translation of a target mRNA by competing with ribosomes for RBS accessibility. Direct Hfq binding to the RBS of an mRNA makes the RBS less likely to be bound by a ribosome, and thus, decreases the probability of translation initiation (Fig. 1.8A). Although Hfq mainly inhibits RNA translation, a few examples of positive regulation have been discovered. For instance, Hfq/sRNA binding near the RBS of *rpoS* changes the sec-

ondary structure of the mRNA and opens up an inhibitory hairpin [221]. This, in turn, releases the RBS sequence for ribosome binding and further promotes RpoS translation. A simplified molecular mechanism, where translation initiation is positively regulated by Hfq, is illustrated in Figure 1.8B.

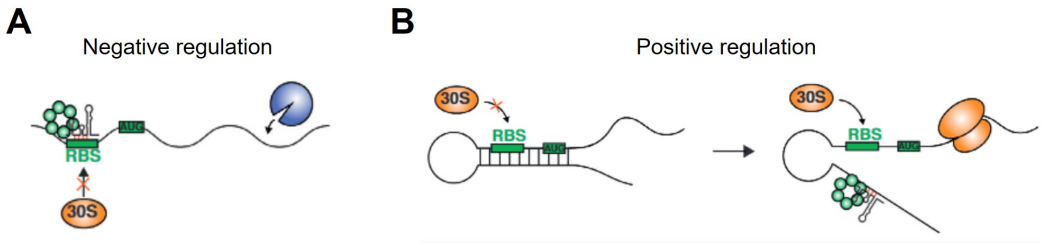

Figure 1.8: Negative and positive mechanisms of Hfq-mediated regulation: inhibition of translation by blocking ribosome access (**A**); changing the structure of RNA inhibitory hairpin to allow ribosome access (**B**). The image was modified from Kavita *et al.* [107].

Furthermore, Hfq can modify the stability of a target mRNA by competing with ribonucleases for the same recognition sequence. After it was shown that in bacterial genes where RNase cleavage sites coincide with Hfq-binding sites [153], Hfq was proposed to preserve an mRNA from degradation either by direct binding to an RNase recognition site or by reshaping secondary structure of the mRNA nearby. For example, Hfq binding to poly(A) tails of polyadenylated mRNAs of *rpsO* protects the transcripts from degradation by ribonucleases [75, 118]. Moreover, Hfq is able to interact with a major endoribonuclease RNase E that results in recruitment of RNase E to a target mRNA followed by rapid decay of the target [34, 136]. Taken together, these diverse functions of Hfq protein emphasize its importance for mRNA turnover in prokaryotes.

Structural and mutational studies have given invaluable insights into the mechanistic aspects of Hfq functions and molecular mechanisms of Hfq-mediated RNA–mRNA pairing. In order to carry out its numerous functions, Hfq proteins assemble into a ring-like structure with four surfaces available for RNA binding: proximal; distal; rim; and C-terminal tail [200, 265]. Various configurations of these binding sites allow for the interactions with different RNA types. Thus, mRNA molecules have high affinity to the distal surface of Hfq through repeated trinucleotide motifs of A-R-N (A-A/G-any nucleotide) [125] while sRNAs (Class I) preferentially bind to proximal and rim sites [197]. Interactions with small RNAs of Class II depend on the proximal and distal Hfq surfaces [200]. Hfq binding surfaces and alternative modes of its interaction with mRNA and different sRNA classes are illustrated in Figure 1.9.

Although multiple binding surfaces point to the significance of Hfq in mRNA and sRNA stability control, there is increasing evidence that Hfq itself can regulate mRNA translation [21, 107, 220]. In the conventional model, Hfq alters the accessibility of the RBS of a target mRNA by direct binding nearby leading to less effective translation [45, 65, 194, 224]. An important

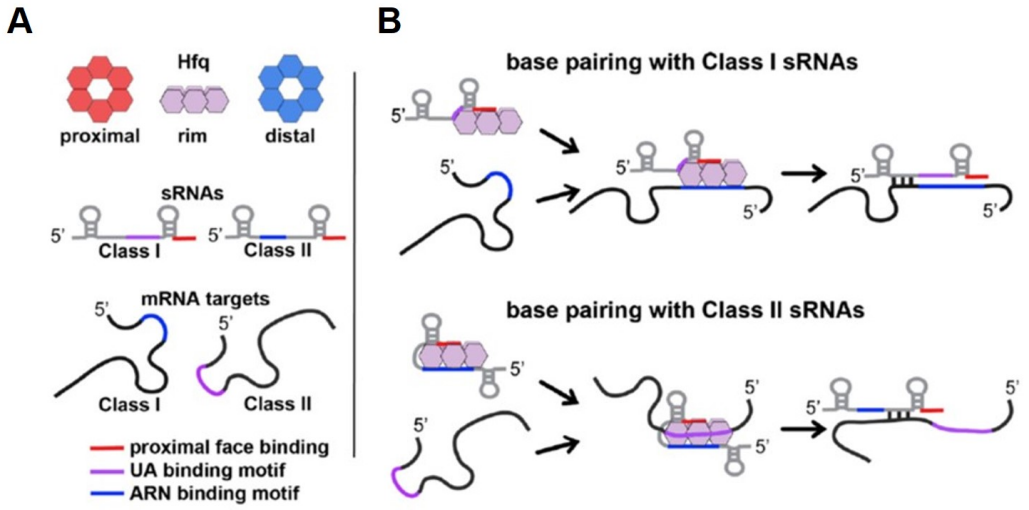

Figure 1.9: Hfq binding surfaces (proximal, distal, rim) (**A**) and alternative modes of its interaction with mRNA and different sRNAs of Class I and Class II (**B**). The figure was borrowed from Schu *et al.* [200].

example of sRNA-independent inhibition of translation is autoregulation of Hfq synthesis. Hfq represses its own translation by direct binding to its own mRNA through its distal face [157, 246]. Such a mechanism allows maintenance of cellular concentration of Hfq proteins even when *hfq* mRNAs are overproduced [157]. However, it is worth noting that, despite being limited by a post-transcriptional negative feedback loop, Hfq proteins are present in high abundance: there are approximately 10,000 hexamers of Hfq in exponentially growing *E. coli* cells [12, 14, 106].

Being recognized as a regulator of cell physiology, Hfq is involved in many pathways, including stress response programs. Therefore, as expected, a disruption of *hfq* gene results in pleiotropic phenotypes. These include reduced fitness, increased cell size, decreased viability upon stress conditions (such as ultraviolet radiation and osmotic shock) and various growth defects [159, 160, 239]. Interestingly, *hfq* deletion does not always lead to a decrease of cell growth [132, 239]. This effect has been shown to depend upon growth medium composition and might result in a populational growth of *hfq* mutants that is similar or even faster than the growth of wild-type cells [1].

1.5.3 Detection of protein-RNA and RNA-RNA interactions

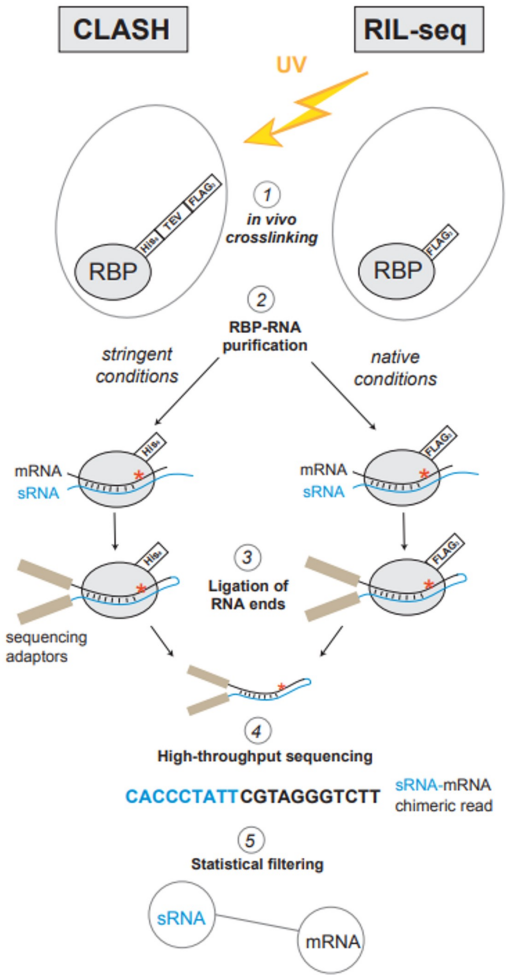

Figure 1.10: Overview of experimental techniques, CLASH and RIL-seq, to capture RNA-protein and RNA-RNA interactions *in vivo*. The figure was borrowed from Iosub [99].

The initial maps of Hfq-RNA interactions, detected in RNA immunoprecipitation studies, revealed thousands of mRNA and sRNA targets [32, 43, 264]. Since then, a number of high-throughput techniques have been developed in order to verify these protein-RNA and RNA-RNA interactions and have identified the RBP-RNA recognition patterns more precisely [83, 86, 97, 142, 163, 229, 238, 240, 253]. Description of CLASH (UV-crosslinking, ligation and sequencing of hybrids) and RIL-seq (RNA interaction by ligation and sequencing) methodologies are reviewed below (Fig. 1.10).

In both methods, CLASH and RIL-seq, the interactions between proteins and directly bound RNAs are stabilized by *in vivo* UV crosslinking (Fig. 1.10, step 1). In CLASH, the RBP of interest is double-tagged (for example, with His6-TEV-3XFLAG) to allow for stringent purification (Fig. 1.10, step 2). This ensures that only RNAs covalently bound to the RBPs are purified. RIL-seq utilizes a single tag (such as a FLAG tag) and purification is performed under milder conditions what is believed to preserve the native structure of Hfq hexamer, and hence, allow for detection of all native Hfq-mediated RNA-RNA interactions (Fig. 1.10, step 2). Then, in both protocols, RNA ends in duplexes are ligated with an RNA ligase (Fig. 1.10, step 3). Finally, cross-linked RNAs are identified by high-throughput sequencing and can be mapped to RBP-RNA or RNA-RNA interactions using bioinformatics pipelines (Fig. 1.10, step 4 and 5). Because of non-stringent purification conditions, RIL-seq outcomes have a higher chance of detecting unspecific bindings, and thus, require more rigorous statistical analysis, compared to the ones uncovered with CLASH [142].

1.6 The scope of this work

RecBCD enzyme is a key player of DNA double-strand break repair pathway in *Escherichia coli*. While being essential, RecBCD over-expression decreases cell viability upon DNA damage. This suggests that RecBCD molecule numbers need to be controlled within an optimal range. A few studies attempted to explore the details of RecBCD transcription however no evidence of regulation has been found. A potential post-transcriptional control mechanism of RecBCD expression is the focus of this work.

Chapter 3 revisits RecBCD expression in wild-type cells using a single-cell approach. Both mRNAs and proteins have been found at low levels: less than one mRNA molecule and five proteins per cell, respectively. Furthermore, propagation of fluctuations within the gene expression network have been explored using a simple stochastic model. Curiously, while *recB* mRNA levels are highly variable, cell-to-cell RecB protein noise is substantially suppressed. This observation led to the hypothesis of a post-transcriptional regulation mechanism actively reducing RecB protein fluctuations.

In Chapter 4, RecB expression in stress conditions is investigated. Specifically, mRNA and protein concentrations have been quantified upon DNA double-strand break induction. Interestingly, despite the decrease in *recB* mRNA concentration upon DNA damage, RecB protein levels have been found to remain unchanged. This implies that lower mRNA levels are compensated by more efficient translation upon stress. Consistent with the previous result, the uncoupling between mRNA and protein production further supports a post-transcriptional regulation mechanism maintaining RecB protein level within an optimal range.

In Chapter 5 a novel Hfq-mediated regulation mechanism of RecB expression is suggested. Firstly, the interaction between a post-transcriptional regulator Hfq and the *recBCD* transcript has been uncovered in the available high-throughput CLASH dataset [100]. Secondly, down-regulation of RecB expression by Hfq *in vivo* is established based on a small (1.3-fold) yet significant level of RecB protein accumulation in *hfq* mutants. Thirdly, based on the localization of Hfq enrichment in the CLASH dataset, a molecular mechanism whereby Hfq inhibits the initiation of RecB translation by binding to the RBS of the *recB* mRNA has been proposed. This fine-tuning mechanism has been further verified by modulating Hfq cellular levels *in vivo*.

Overall, these observations suggest that a global regulator, Hfq, has a key role in the post-transcriptional regulation of RecB expression. Furthermore, when Hfq-mediated regulation is perturbed, (i) less effective noise suppression has been observed and (ii) cells were found to reduce DNA repair capacity. This Hfq-mediated fine-tuning has been proposed to contribute to RecB protein noise reduction and protects cells from having too many RecBCD molecules.

Chapter 2

Materials and Methods

2.1 Materials

The recipes for growth media and solutions used in this study are provided in Table 10.2. The details and final concentrations of supplements and antibiotics are reported in the tables: 10.3 and 10.4, respectively. Compositions of all stock solutions and buffers can be found in Table 10.1. Preparation procedures of hybridization (smFISH), fixation (smFISH and HaloTag-labelling) and RNA extraction solutions are described in detail. Also, RNA FISH probe and Janelia Fluor 549 dilutions are described. RNA FISH probe sequences for *recB*, *recC* and *recD* genes are listed in Tables: 10.9 – 10.11. Finally, the preparation of agarose pads for microscopy experiments are described.

2.1.1 Growth media, supplements and antibiotics

In all microscopy, RT-qPCR and growth curve experiments, cell cultures were grown in M9-based medium supplemented with 0.2% glucose, 2 mM $\mathrm{MgSO_4}$, 0.1 mM $\mathrm{CaCl_2}$ and either 10% Luria broth (Imaging Medium) or 1X MEM Essential and MEM Non-Essential Amino Acids (Gibco®) (Generic Medium). Luria broth (LB) was used for serial dilution assays and cloning. Liquid media were stored at room temperature with the exception of microscopy experiments where media were warmed to 37℃ prior to an experiment. LB agar, used for plating in cloning and serial dilution assays, was heated in a microwave and stored at 55℃. Compositions of M9-based media and salt solutions are provided in Appendix 10.1 (Tables: 10.1 and 10.2).

In the *recB* induction smFISH experiments, 0.2% glucose was replaced with 0.2% glycerol, and arabinose ranging from $10^{-5}\%$ to 1% was added to the medium. In the mRNA lifetime measurements with smFISH, rifampicin was added to cell cultures to give a final concentration of 500 µg/ml. In the DSB induction experiments, a sub-lethal concentration of ciprofloxacin, 4 ng/ml, was used (minimal inhibitory concentration (MIC) is 16 ng/ml).

In the phenotypical assays, plates were supplemented with nalidixic acid (1 – 2.5 µg/ml) or ciprofloxacin (2 – 12 ng/ml). In the other populational experiments and cloning protocols, ampicillin (100 µg/ml), kanamycin (50 µg/ml) or chloramphenicol (30 µg/ml) were added to cultures when appropriate. All supplement and antibiotic concentrations are summarized in Appendix 10.1 (Tables: 10.3 and 10.4).

2.1.2 Hybridization and wash solutions (smFISH)

To prepare hybridization solution for smFISH experiments, the following reagents were mixed by vigorous vortexting in DEPC-treated water: 0.1 g/ml of dextran sulphate, 40% (w/v) (for *recB* and *recD* experiments) or 45% (w/v) (for *recC* experiments) of formamide, 1 mg/ml of *E. coli* tRNA, 2X SSC, and 2 mM ribonucleoside vanadyl complex. The solution was mixed and sterilized with a 0.22-µm filter.

To make up 40% wash solution, 2X SSC and 40% (w/v) of formamide were mixed in DEPC-treated water. The 40% wash solution was used for *recB*, *recC* and *recD* genes.

2.1.3 Fixation solution (smFISH and HaloTag-labelling)

To prepare 2.5% (v/v) fixation solution used in HaloTag-labelling experiments, 1X PBS, 2.5% (v/v) formaldehyde were thoroughly mixed in $\mathrm{ddH_2O}$. 3.2% (v/v) fixation solution for smFISH was made up similarly by mixing 3.2% (v/v) formaldehyde in 1X PBS.

2.1.4 RNA FISH probe and Janelia Fluor 549 dilutions

RNA FISH probes with TAMRA dye were ordered from Stellaris RNA FISH (LGC Biosearch Technologies). The sequences for *recB*, *recC* and *recD* genes are listed in Tables: 10.9, 10.10 and 10.11, respectively (Appendix 10.1). The working 25 µM stocks of RNA FISH probes were prepared by resuspending the initial 100 µM FISH probe stocks in RNase-free TE buffer.

Janelia Fluor HaloTag ligand 549 (JF549) (Promega) was resuspended in 39 µl of 100% DMSO and aliquoted in 5 µl eppendorf tubes. The aliquots were stored at – 20℃ in the dark.

2.1.5 Agarose pads

To make agarose pads, 2% agarose was dissolved in 1X PBS by heating the mixture in a microwave and constant swirling. To dissolve the agarose, slow heating regime was used in order to prevent its boiling evaporation. Then, $\sim$ 400 µl of the warm agarose solution was poured into a defined squared area of a microscope slide, covered by another slide and gently pressed on top. Agarose solution between the slide was left to solidify for $\sim$ 1 min. Then, the upper slide was carefully removed by sliding and the excess of the agarose was excised

with a scalpel. Finally, 2 – 4 µl of a sample was added to the middle of a tilted agar pad, allowed to dry for $\sim$ 10 – 15 min and covered by a coverslip. Adding cells onto a tilted agar pad allowed to achieve a gradient in cell concentration. Because cell density at the end of the protocol varies between experimental repeats and growth conditions, it is useful to achieve different concentration of cells onto an agar pad to find the optimal cell coverage once imaging with microscopy. Coverslips were previously cleaned by heating in a kiln (up to $\sim$ 400 – 500℃, overnight), then kept in KOH solution (overnight), rinsed in distilled water and dried with nitrogen gas.

2.2 Methods

Here, growth conditions for the microscopy and populational experiments performed in the study are described. The details of the strains and plasmids constructed for this study are provided. The entire list of strains and plasmids used in this work can be found in Appendix 10.1 (Tables: 10.5 and 10.6). Protocols for populational experiments such as growth measurements, phenotypical assays, bacterial RNA extractions, RT-qPCR are described. The detailed single-molecule microscopy protocols for RNA FISH and HaloTag-labelling with the accompanied image analysis pipelines are provided. Lastly, modelling, statistical and simulation approaches used in the study are outlined.

2.2.1 Growth conditions

In microscopy experiments, cell cultures from frozen – 80℃ stocks were grown in a medium overnight (14 – 16 hours) at 37℃. The overnight cultures were diluted (1:300) into 15 ml medium and grown at 37℃ to the mid-exponential phase (optical density OD_{600} = 0.2 – 0.3). The optical density was measured with spectrophotometer. Unless otherwise noted, cells were further treated according to the smFISH or HaloTag-labelling procedures described below.

In the *recB* induction smFISH experiments, arabinose was added after one hour of overday growth, the cultures were grown for one more hour, and then samples were collected.

For the smFISH experiments with rifampicin, the antibiotic was added once the cultures had reached OD_{600} = 0.2 ($\mathrm{t} = 0$). Cells were harvested at 1 min intervals and fixed directly in 3.2% formaldehyde (Fig. 2.1). Further steps of smFISH procedure were followed as described below.

In the DSB induction experiments, ciprofloxacin was added at $\mathrm{OD}_{600} \sim 0.1$ and cell cultures continued growing for one hour. For mRNA quantification with smFISH, the cultures were grown for one more hour before being harvested and fixed in the final concentration of 3.2% formaldehyde. For protein quantification, the HaloTag-labelling protocol was followed as described below with ciprofloxacin being kept in the medium during labelling and washing steps.

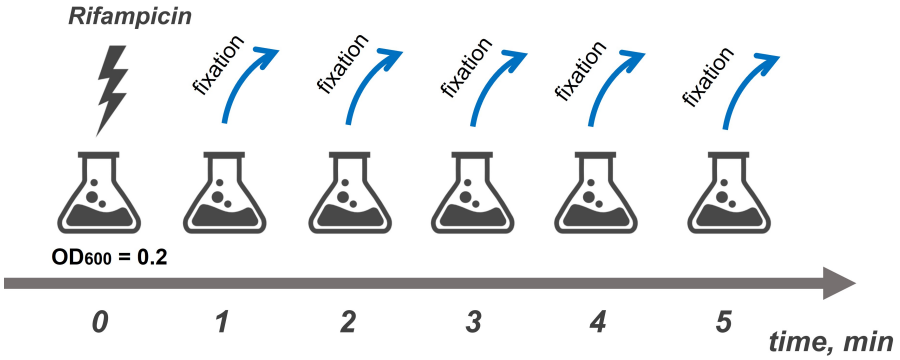

Figure 2.1: Illustration of a time course experiment with inhibition of transcription with rifampicin. The antibiotic was added in the final concentration of 500 µg/m and the samples were collected with 1 min time intervals for 6-8 min.

In the sRNA over-expression experiments, kanamycin (50 µg/ml) was added to cell cultures to select for the cells carrying an expression plasmid. In the experiments where Hfq point mutants were expressed, low concentration of ampicillin (50 µg/ml) was added to select for the cells carrying an expression plasmid. 100 µg/ml of ampicillin was added to a medium in the *recBCD* overproduction experiments with pDWS2 plasmid. In the *ptrA* over-expression experiments, chloramphenicol (30 µg/ml) was added to select for the cells carrying an expression plasmid. In all cases, an antibiotic was added to overnight cultures and kept during overday growth until cells were fixed in formaldehyde.

For the HaloTag pulse-chase experiments, cell cultures were grown till $\mathrm{OD}_{600} \sim 0.2$ and incubated with the dye JF549 for another hour. Then, the excess of the dye was removed by extensive washing and the number of the labelled molecules over time was measured using epifluorescent microscopy with the time intervals of $\sim$ 15 min (Fig. 2.2).

2.2.2 Strains and plasmids construction

All strains and plasmids used in the study are listed in Appendix 10.1 (Tables: 10.5 and 10.6). The *Escherichia coli* MG1655 and its derivatives were used for all conducted experiments (except for one with arabinose-inducible expression where the *Escherichia coli* BW27783 background strain was used). The strain with RecB-HaloTag fusion, MEK65 [119], was used in Halo-labelling experiments while the corresponding parental strain, MG1655, was used in smFISH experiments.

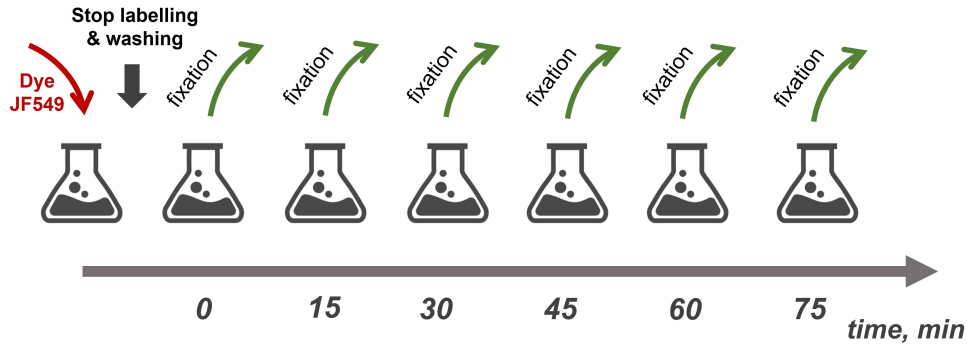

Figure 2.2: Illustration of the HaloTag pulse-chase experiment. Cells were grown until an early exponential phase and incubated with the JF549 dye for 1 hour. Then, the HaloTag labelling was stopped by removing the excess of the dye with extensive washing. The cell cultures continued growing and the samples were collected in a time-course with intervals of $\sim$ 15 min for 75 min in total.

MEK1329, MEK1345 and MEK1349 were built with plasmid-mediated gene replacement (PMGR) method [124] with pTOF24 derivative plasmids, pTOFΔ*recC* and pTOFΔ*recB* [51]. The PMGR protocol is described in detail in Appendix 7. The deletions of *recB* or *recC* genes were verified using a phenotypical assay, polymerase chain reaction (PCR) and Sanger sequencing of the locus of interest.

The lambda red cloning technique [52] was used to construct *hfq* (MEK1902 and MEK1457), *chiX* (MEK1888 and MEK1449), *cyaR* (MEK1887 and MEK1453) and *arcZ* (MEK1919 and MEK1921) deletions in the MG1655 and MG1655 $recB165 :: halotag$ backgrounds. The *hfq*, *chiX*, *cyaR* and *arcZ* deletions were verified by PCR. The primers used for the lambda red protocol are listed in Table 10.7 (Appendix 10.1). Details of the protocol are provided in Appendix 7.

pIK01 and pIK02 were constructed to allow arabinose-inducible expression of *recC* and *recB*, respectively. A fragment of the backbone plasmid was amplified from pBAD33 vector with the oIK03/oIK04 primers and treated with DpnI (NEB) restriction enzyme. *recB* and *recC* genes were amplified from *E. coli* chromosome using the oligos oIK10/oIK11 (for *recB*) and oIK01/oIK02 (for *recC*). The purified PCR products were then inserted into the backbone plasmid using Gibson assembly and the output was transformed into DH5α cells. Correct plasmid constructions were verified by PCR and Sanger sequencing.

pZA21-ChiX plasmid was constructed by Dr Ira Alexandra Iosub to allow overproduction of a small RNA ChiX. The plasmid, derived from the pZA21MCS (Expressys), was generated according to the protocol, described here [100], using chiX_ZA21 and pZA21MCS_5P primers.

pIK06 and pIK07 were constructed to allow arabinose-inducible expression of *ptrA* and a truncated *ptrA* fragment, respectively. The backbone was amplified from pBAD33 using oIK03/ oIK04 primers and treated with DpnI restriction enzyme. *ptrA* gene and a truncated *ptrA* fragment were amplified from *E. coli* chromosome using the oligos oIK66/oIK67 and oIK66/oIK68, respectively. The purified PCR products were then inserted into the backbone plasmid using Gibson assembly and the output was transformed into DH5α strain. Plasmid constructions were verified by PCR and Sanger sequencing.

All oligonucleotide sequences, used in the study, are listed in Table 10.7 (Appendix 10.1).

2.2.3 Population growth measurements

Bacterial cells, seeded by inoculating from fresh colonies, were grown in the appropriate medium at 37°C overnight. The cultures were diluted (1:1000) into 20 ml of the medium and grown in a shaking incubator at 37℃. Optical density measurements were taken overday at an absorbance wavelength of 600 nm (OD_{600}) with a time interval of 15 – 20 min. The exponential phase of growth, OD_{600} = [0.08, 0.4], was used for the fitting procedure. Fitting was performed with a linear regression model in MATLAB (`fitlm`) on the OD_{600} measurements and normalized to the OD_{600} at $\mathrm{t} = 0$.

2.2.4 Serial dilution assay

In serial dilution assays, cell cultures were grown in LB at 37°C for $\sim$ 15 hrs. The overnight cultures were normalized to the optical density of OD_{600} = 1.0, spotted in 5- or 10-fold serial dilutions (started from OD_{600} = 1.0 $a.u.$) onto LB plates supplemented with 0 – 12 ng/ml of ciprofloxacin or 0 – 2.5 µg/ml of nalidixic acid and incubated at 37°C overnight.

2.2.5 Bacterial RNA extraction and RT-qPCR

Bacterial cells were grown as described earlier, harvested in the equivalent of $\mathrm{vol} \times \mathrm{OD}_{600}$ = 2.5 ml and flash-frozen in liquid nitrogen. In the HaloTag-labelling experiments, samples for RNA extraction were grown at 37℃ without the dye and collected right before the fixation step. Cells were lysed and total RNA was extracted by the Guanidinium thiocyanate-phenol-chloroform protocol (Table 10.1, [101]). 10 µg of the extracted RNA was treated with 1 µl of $\mathrm{TURBO}^{\mathrm{TM}}$ DNase (Thermo Fisher Scientific) and 0.1 µl of $\mathrm{SUPERase} \cdot \mathrm{In}^{\mathrm{TM}}$ RNase Inhibitor (Thermo Fisher Scientific) in a 10 µl reaction for 1 hour at 37℃ and purified by ethanol precipitation.

Primers for real-time qPCR experiments were designed using the OligoPerfect Primer Designer (Thermo Fisher Scientific) and listed in Table 10.8. RT-qPCR quantification was performed on 20 ng of RNA with the $\mathrm{Luna}^{®}$ Universal One-Step RT-qPCR Kit (NEB). All qPCRs were performed in technical triplicates alongside no template and no reverse transcriptase controls.

The qPCR reactions were performed using a $\mathrm{LightCycler}^{\circledR}96$ (Roche) with a 40 cycles of two-step amplification (95℃, 10 s, 4.4℃/s; 60℃, 30 s, 2.2℃/s) followed by a melting curve (65℃, 60 s, 2.2℃/s; 97℃, continuous acquisition). Initial steps of the data analysis were performed in the $\mathrm{LightCycler}^{\circledR}96$ application software (v1.1.0.1320, Roche) with the built-in protocols: (i) *Absolute Quantification (Fit Point)* for determination of threshold cycles (Ct) and (ii) *High Resolution Melting* for evaluation of primer specificity. The *rrfD* gene (5S rRNA) was used as a reference gene to compute a fold-change relative to a control sample with the $2^{-\Delta\Delta Ct}$ method. Outlier samples with the standard deviation $std(Ct) > 0.3$ across technical replicates were removed from the analysis.

2.2.6 Single-molecule RNA FISH (smFISH)

Single-molecule fluorescent *in situ* hybridization (smFISH) experiments were carried out according to the established protocol [217]. Bacterial cells were grown in 500-ml flasks in an orbital shaker, harvested in the equivalent of $\mathrm{vol} \times \mathrm{OD}_{600}$ = 1 ml, centrifuged at 3500 r.p.m for 5 min at 4℃ and incubated in 3.2% formaldehyde solution for 30 min. Then, the samples were washed with 1X PBS and further permeabilized in 70% ethanol overnight at 4℃. The next day, the samples were centrifuged at 4000 r.p.m for 7 min at room temperature, hybridized with a gene-specific set of TAMRA-labelled RNA FISH probes in 40% (*recB* and *recD*) or 45% (*recC*) hybridization solution and kept at 30℃ overnight. Probe sequences are listed in the tables: 10.9, 10.10 and 10.11 (Appendix 10.1). Then, unbound RNA FISH probes were removed with multiple washes in 40% washing solution. Finally, the samples were visualized with epifluorescent microscopy as described below. In each experiment, a mutant strain (Δ*recB*, Δ*recC* or Δ*recD*) was subjected to the protocol in parallel with the primary sample. The protocol is illustrated in Figure 2.3.

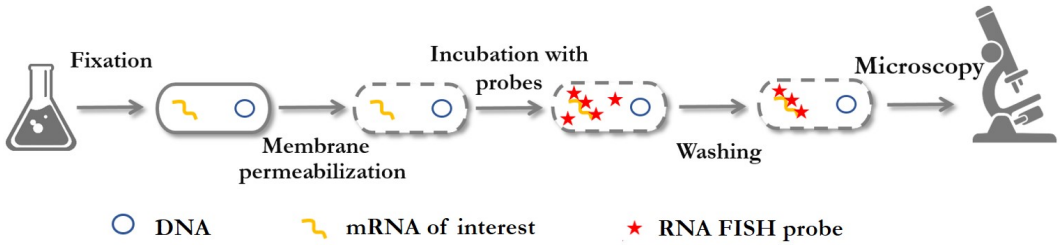

Figure 2.3: The main steps of single-molecule fluorescent *in situ* hybridization (smFISH) protocol: cell fixation, membrane permeabilization, incubation with RNA FISH probes, washing to remove unbound probes, imaging.

2.2.7 HaloTag-labelling

For single-molecule RecB protein quantification, the HaloTag-labelling protocol was followed as described in detail by Lepore *et al.* [119]. *E. coli* strains with RecB-HaloTag fusion were grown in 500-ml flasks in an orbital shaker, harvested in the equivalent of $\mathrm{vol} \times \mathrm{OD}_{600}$ = 0.2 ml, centrifuged at 8000 r.p.m for 5 min at 4 ℃ and incubated with $\mathrm{JaneliaFluor}^{\circledR}549$ (JF549) dye (Promega) in the final concentration of 5 µM for one hour (from this step cells were kept in low light to prevent bleaching of the dye). Then, the cells were washed with aspirator (4 – 5 washes) and incubated in 2.5% formaldehyde for 1 hour on a roller. Finally, the samples were washed in 1X PBS two times and imaged on 2% agarose pads (the details of imaging procedure are provided below). In each experiment, a parental strain that does not have HaloTag fusion ('no HaloTag') was subjected to the protocol in parallel with the primary samples. The main steps of the HaloTag-labelling protocol is illustrated in Figure 2.4.

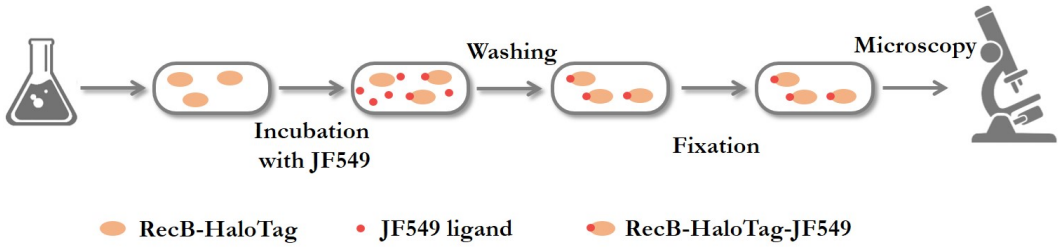

Figure 2.4: The main stages of the HaloTag-labelling protocol: incubation with JF549 ligand, washing to remove unbound dye molecules, cell fixation, imaging.

2.2.8 Microscopy set-up and conditions

Image acquisition was performed using an inverted fluorescence microscope (Nikon Ti-E) equipped with an EMCCD Camera (iXion Ultra 897, Andor), a Spectra X Line engine (Lumencor), dichronic mirror T590LPXR, 100X oil-immersion Plan Apo objective (NA 1.45, Nikon) and 1.5X magnification lens (Nikon). A TRITC (ET545/30nm, Nikon) filter was used for imaging HaloTag-labelling experiments while smFISH data were acquired with an mCherry (ET572/35nm, Nikon) filter.

Once a protocol (smFISH or HaloTag-labelling) had been carried out, fixed cells were resuspended in 1X PBS and mounted onto a 2% agarose pad. The snapshots were acquired in bright-field and fluorescence channels. In all experiments, an electron-multiplying (EM) gain of 4 and an exposure time of 30 ms were used for bright-field imaging. In the fluorescence channel, smFISH data were acquired with a 2 s exposure time and a gain of 4, while RecB

proteins were visualized with an exposure time of 2 s and EM gain 200. For each XY position on an agarose pad, a Z-stack of 6 images centered around the focal plane (total range of 1 µm with 0.2 µm step) was obtained in both channels. A set of multiple XY positions were acquired for each slide to visualize ∼1000 cells per sample.

2.2.9 Image processing: Cell segmentation

Microscopy bright-field images were used to identify positions of bacterial cells with an automated MATLAB-based pipeline. The segmentation algorithm was based on detection of cell edges by passing an image through a low-pass filter and was applied to defocused bright-field images (Fig. 2.5, left). In the pipeline, a defocused image is convoluted with a low-pass filter and a score of each pixel of the image is computed based on the relative intensity values of the neighbouring pixels. Then, the intensity threshold is applied to the intensity landscape and cells are identified as individual connected components. Cell positions are saved in a binary mask as shown in Figure 2.5 (middle image). The cells located close to the image edges were discarded with an additional post-processing filter. The binary masks are transformed into final masks where each cell has its ID number assigned to each cell within an image (Fig. 2.5, right).

The segmentation outputs were manually corrected by discarding misidentified cells in the accompanied graphical interface. It is worth mentioning that the algorithm identifies well the cells which are separated from each other, however, mistakes frequently occur when cells clump together. The corrected segmentation results were saved as MATLAB matrices (cell masks) containing mapping information between pixels and cell ID numbers. The cell masks were used as an input for the spot-finding software described below. Bright-field segmentation pipeline is available from an open-access repository [2].

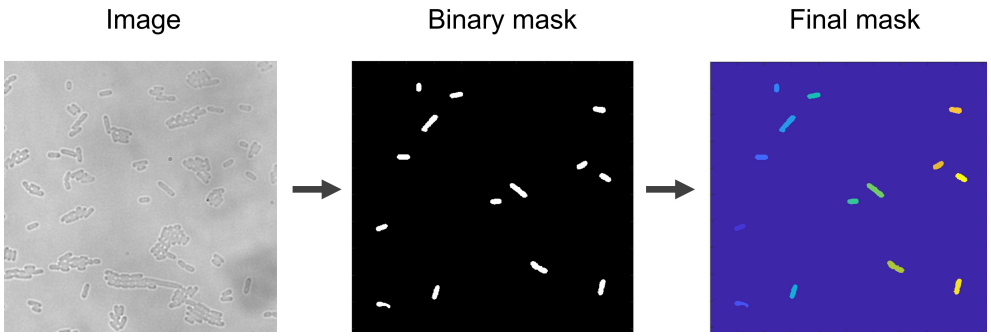

Figure 2.5: Cell segmentation algorithm based on the convolution of a bright-field defocused image with a low-pass filter. The positions of identified cells are saved in binary masks (with zeros and ones) and final masks (different colors correspond to cell identification numbers).

2.2.10 Image processing: Spot detection

The Spätzcell package (MATLAB) was utilized to detect foci in fluorescence smFISH images [3, 217]. Principally, the analysis consists of the following steps: (i) identification of local maxima above a chosen intensity threshold, (ii) matching the maxima across frames in a Z-stack, (iii) performing 2D-Gaussian fitting of the detected maxima. The main steps of the Spatzcells software are illustrated in Figure 2.6.

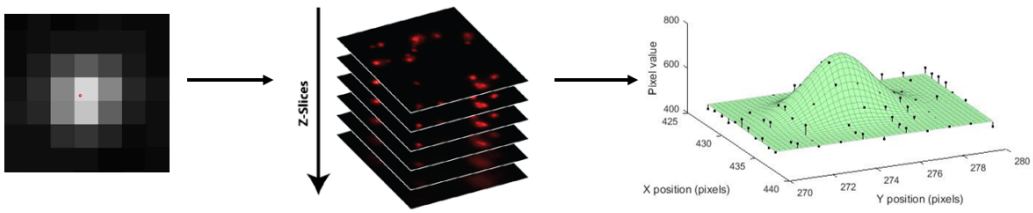

Figure 2.6: The illustration of the major steps of the Spätzcell spot detection software: detection of local maxima, connecting the spots across Z-stacks, 2D-Gaussian fitting of each spot. The image was modified from Skinner *et al.* [3].

Based on the peak height and spot intensity computed from the fitting output, specific signal was separated from false positive spots. Then, the integrated spot intensity histograms were fitted to the sum of two Gaussian distributions (see equation 2.1, where a, b, c and d are the fitting parameters), which correspond to one and two mRNA molecules per focus. An intensity equivalent corresponding to the integrated intensity of the average number of FISH probes bound to one mRNA was computed as a result of multiple-Gaussian fitting procedure, and all identified spots were normalized by the one-mRNA equivalent.

$$f(x) = ae^{-\frac{(x-b)^2}{c^2}} + de^{-\frac{(x-2b)^2}{2c^2}} \tag{2.1}$$

To detect fluorescence single-molecule foci in the HaloTag-labelling experiments, the first step of Spätzcell analysis was modified (similar to the modification implemented by Wang *et al.* [251]). Two modifications specific to HaloTag-labelling were tested: (i) removing a Gaussian filter (that was used to smooth raw signal in the original version) or (ii) calculating the Laplacian of a Gaussian-filtered image. Both versions gave similar and consistent quantification results. Peak height intensity profiles allowed separation of specific signal from false positive spots. As the result of low-abundance of the protein of interest and single-molecule labelling, the analysis of integrated intensity was omitted for HaloTag-labelling images. All other steps of the Spätzcell software remained unchanged. The modified version of the analysis used for spot detection in HaloTag-labelling experiments is available from an open-access repository [4].

Total (mRNA or protein) molecule numbers per cell were obtained by matching spot positions to cell segmentation masks and quantifying the number of (mRNA or protein) spots within each cell (Fig. 2.7). Molecule concentration was calculated for each cell as total number of mRNA or protein molecules per cell divided by the area of the cell ($1/\mu m^2$).

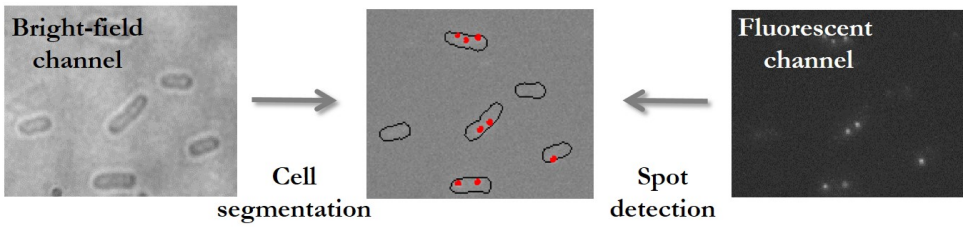

Figure 2.7: The illustration of the entire image analysis pipeline composed of bright-field cell segmentation and spot detection with Spätzcell.

2.2.11 Statistical analysis

Unless otherwise stated, histograms and RT-qPCR results represent the average across at least three repeated experiments and error bars reflect the maximum difference between the repeats.

Fitting of mRNA and protein distributions were performed with the maximum likelihood estimation method in MATLAB (`fitdist`). The Kullback–Leibler divergence (D_{KL}) between simulated and experimental distributions was calculated with the MATLAB function [5]. Comparisons between the average protein or mRNA concentrations among different conditions were performed with a two-sample *t*-test and *P* values were calculated in the MATLAB built-in function `ttest2`. To visualize data with box plots, MATLAB function `boxplot` was utilized.

2.2.12 Stochastic simulations

Single trajectories of mRNA and protein abundance over time were generated with the Gillespie's algorithm (direct method) [78], executed in MATLAB environment [6]. Parameters used in simulations are listed in Table 10.15. A total of 10,000 replicate simulations has been performed for a given condition, and probabilistic steady-state mRNA and protein distributions were recovered for population snapshot data.

2.2.13 CLASH and RIL-seq data analysis

The CLASH data (NCBI Gene Expression Omnibus ID GSE123050, [100]) and RIL-seq data (ArrayExpress ID E-MTAB-3910, [142]) were analysed for this study by Dr Ira Alexandra Iosub and Dr Sander Granneman. The entire data analysis pipelines are available from open-access repositories [7, 8].

Chapter 3

RecBCD mRNA and protein quantification in wild-type cells

The aim of this chapter is to quantify RecBCD expression in single wild-type *Escherichia coli* cells. Initially, *recBCD* transcription was analysed and precise quantification of *recB*, *recC* and *recD* mRNAs was performed with single-molecule RNA FISH (smFISH) technique. It is shown that *recBCD* mRNAs are present in low abundance: less than one molecule per cell. Next, *recB* and *recC* mRNA degradation rates were measured in a time-course experiment. It is concluded that the extremely low abundance of *recBCD* mRNAs can be explained by superposition of fast mRNA decay and infrequent (relative to other mRNAs in a cell) transcription.

Further quantification of RecB subunits using a HaloTag-labelling protocol was performed with single-cell resolution and confirmed a previously reported quantification result: ~ 5 copies per cell [119]. Additionally, the RecB protein removal rate was measured in a modified pulse-chase HaloTag-labelling experiment. These results were summarized in a two-stage stochastic model of gene expression and the parameters (such as mRNA burst size, transcription and translation rates) were inferred from the experimental data.

Finally, the gene expression model was further applied to track propagation of fluctuations within the network. It is shown that *recB* mRNA production is noisy while cell-to-cell RecB protein variability is lower than expected for a constitutive gene. Thus, it is hypothesized that there is a post-transcriptional regulatory mechanism of RecB expression that actively suppresses fluctuations at the protein level.

3.1 *recBCD* mRNA quantification

To investigate RecBCD transcription in *E. coli*, precise quantification of *recB*, *recC* and *recD* mRNAs with fluorescent *in situ* hybridization (smFISH) was performed [217]. Specificity of the hybridization conditions and sensitivity of the protocol in a low abundant regime were verified. The Spätzcell pipeline [3] was used for image analysis: parameters were optimized to allow for separation of specific signal from the background and for identification of co-localized mRNAs. As the result of quantification, similarity and extremely low abundance of *recBCD* transcripts was demonstrated. To further describe *recBCD* transcription with a stochastic gene expression model, a gene-dosage effect was taken into account by identifying cells ('newborns') which had not yet replicated the *recBCD* locus. Secondly, degradation rates of *recB* and *recC* mRNAs were measured in independent time-course experiments. Finally, modelling analysis by inferring mRNA burst size and transcription rate from the smFISH data was performed.

3.1.1 smFISH: technical controls and image analysis

In order to verify the specificity of the hybridization conditions for *recB*, *recC* and *recD* mRNAs, the wild type and corresponding mutant (Δ*recB*, Δ*recC* or Δ*recD*) strains were subjected to the smFISH protocol and then visualized with epifluorescent microscopy in bright-field and fluorescence channels. While well-defined fluorescent foci were detected in wild-type cells, high-intensity specific spots were rarely seen in the mutants (Fig. 3.1). As seen from the representative images, very low numbers of foci were detected: from 0 to 5 focus/foci per single cell (consistently with preliminary results [235]).

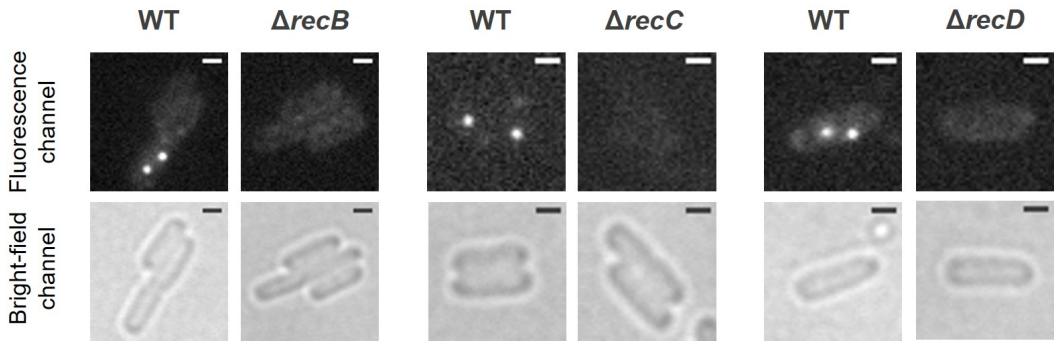

Figure 3.1: Examples of fluorescence and bright-field images of *recB*, *recC* and *recD* mRNA FISH experiments in wild-type and mutant strains. Foci correspond to specific mRNA FISH signal. All fluorescence images are shown in the same intensity range. Scale bars represent 1 µm.

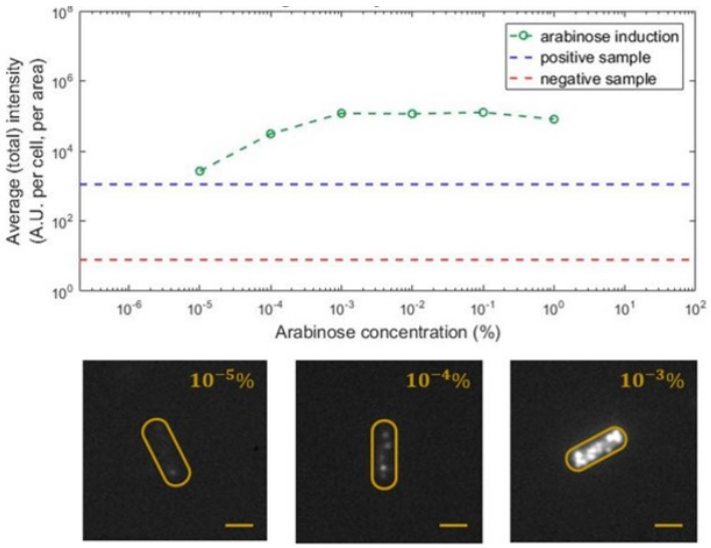

Figure 3.2: Top: Fluorescence signal, detected in the samples with *recB* over-expression from an arabinose-inducible plasmid (pIK02), presented as a function of arabinose concentration (green dashed line). The fluorescence intensities in the wild type and Δ*recB* strains are shown with blue dashed and red dashed lines, respectively. Total number of analysed cells in each sample: 440 for $10^{-5}\%$; 529 for $10^{-4}\%$; 567 for $10^{-3}\%$; 475 for $10^{-2}\%$; 434 for $10^{-1}\%$; 372 for 1%. **Bottom:** Representative fluorescence images of the samples where *recB* expression was induced with $10^{-3}\%$, $10^{-4}\%$ or $10^{-5}\%$ arabinose. All fluorescence images are shown in the same intensity range. Scale bars represent 1 μm.

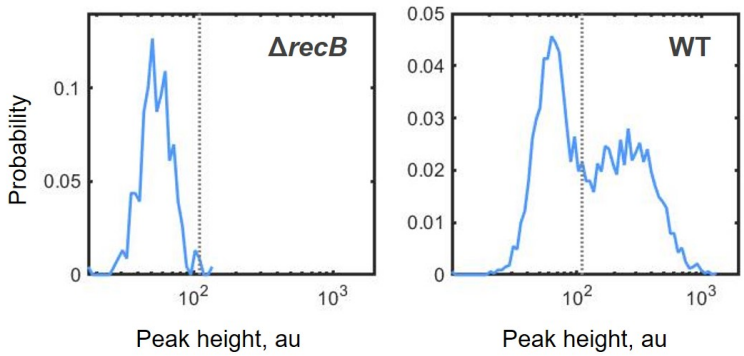

Figure 3.3: Peak height intensity profiles for foci detected in the wild-type and Δ*recB* samples. The dashed line indicates the intensity threshold (99.9%) that separates a specific signal from the background.

A range of hybridization conditions (the percentage of hybridization solution was varied from 35% to 55%) was tested in order to increase and optimize the labelling efficiency of the smFISH protocol. The details of the smFISH protocol with optimized hybridization conditions are provided in Chapter 2. Additionally, the smFISH results were verified with RT-qPCR quantification and resulted in $\sim$ 0.1 molecules per cell for all three genes. While estimation of single molecule numbers from a population experiment such as qPCR may not provide the most accurate values, the similarity of magnitudes of mRNA abundance measured by two independent methods strongly suggests that most of the transcripts have been labelled and detected in the optimized smFISH protocol.

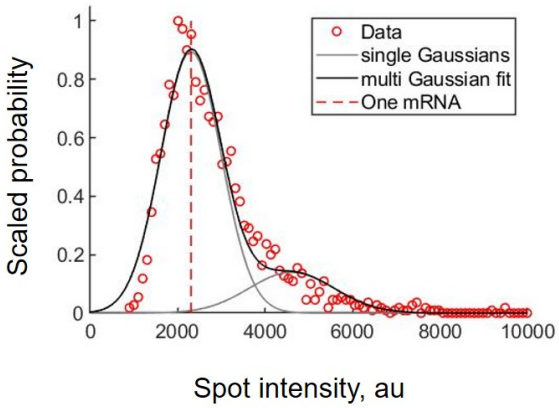

Figure 3.4: Integrated spot intensity histogram for foci detected in wild-type cells. The data, binned and averaged in groups, are shown in red circles. The grey lines correspond to single Gaussian distributions, while the black solid curve is the sum of these two Gaussians. The red dashed line indicates the intensity equivalent corresponding to the integrated intensity of FISH probes (in average) bound to a single mRNA.

cells.

To test whether the optimized smFISH protocol enables detection of such low mRNA levels, over-expression of *recB* mRNAs from an arabinose-inducible low-copy plasmid in the *E. coli* BW27783 strain was carried out. The expression plasmid with *recB* gene under the control of pBAD promoter, pIK02, rescued Δ*recB* mutant phenotype in a phenotypic assay. It is worth noting that the BW27783 background allows for gradual induction of expression from the pBAD promoter with arabinose used as an inducer in contrast to the MG1655 background where a gene under control of pBAD promoter is expressed in an all-or-none manner. In the BW27783 strain, the native arabinose-dependent promoter of the gene encoding the high-capacity transporter *araE* is replaced by a constitutive arabinose-independent promoter [110]. Thus, with the increase of arabinose concentration in the medium, gradual induction of *recB* expression was expected in BW27783 bacterial

As seen in Figure 3.2, a corresponding gradual amplification of the fluorescence signal was detected for an arabinose concentration increase from $10^{-5}\%$ to $10^{-3}\%$. In addition, a fluorescence signal in wild type (blue dashed line) was similar to the $10^{-5}\%$ arabinose induction sample. This confirms that the smFISH protocol is sensitive enough to reliably detect mRNAs present in $\sim$0.1 molecule per cell.

Next, two questions were posed: first, is it possible to separate a specific signal from the background? And, second, can co-localized mRNAs be identified? In order to perform detailed image analysis, an available package Spätzcell was utilized [3, 217]. The package was specifically designed for detection of foci in fluorescence smFISH images.

In the Spätzcell pipeline, each local maximum with the intensity above a chosen threshold was fitted by a 2D-Gaussian and spot characteristics (such as peak height and integrated intensity) were computed from the fitting output. The peak height histograms for wild type and Δ*recB* are shown in Figure 3.3. The histogram for the wild-type sample showed two clearly-separated modes while only one mode was present in the Δ*recB* peak height intensity distribution. This suggested that the mode with higher peak intensity corresponds to a specific signal while the other mode relates to the background spots. Based on this interpretation, a specific signal was separated from the background foci by applying the 99.9% threshold (gray dashed line in Fig. 3.3).

To answer the second question, an integrated spot intensity profile in the wild-type sample was analysed. Having assumed that (i) probe hybridization is a probabilistic process, (ii) binding of single RNA FISH probes happens independently, and (iii) in the majority of cases there is one mRNA per spot, it was expected that integrated intensities of FISH probes bound to one mRNA are Gaussian distributed. In the case of two co-localized mRNAs, there are two independent binding processes and, therefore, a wider Gaussian distribution with twice higher mean and twice larger variance will be expected. In fact, the integrated spot intensity profile has the main mode that corresponds to a single mRNA per focus, and a 'shoulder' that represents a population of spots with two co-localized mRNAs (Fig. 3.4). Based on the underlying model, the spot intensity histogram was fitted to the sum of two Gaussian distributions (multi Gaussian fit) and an intensity equivalent corresponding to the integrated intensity of average number of FISH probes bound to one mRNA was computed (shown as red dashed line in Fig. 3.4).

3.1.2 All-cell *recBCD* mRNA distributions

Based on the results of image analysis, intensity signals of all identified spots were normalized by a one-mRNA equivalent. Then, focus positions were matched with bright-field cell segmentation masks and *recB*, *recC* and *recD* mRNA distributions were presented for the whole range of cell sizes (Fig. 3.5). The corresponding mutant strains (Δ*recB*, Δ*recC* or Δ*recD*) were used for evaluation of the amount of non-specific spots per cell resulting in negligible errors of false positives from 0.007 to 0.016 per cell (Fig. 3.5: inserts). In wild-type cells, very low average levels of *recB*, *recC* and *recD* transcripts were detected, ~ 0.62, ~ 0.50, and ~ 0.59 molecules per cell, respectively. The average mRNA levels for all three genes are fairly similar with *recB* and *recD* mRNA levels being closer to each other than to *recC*. This observation is consistent with the genetic organisation of the locus: *recB* and *recD* genes are controlled by the same promoter while *recC* gene is under control of a different one.

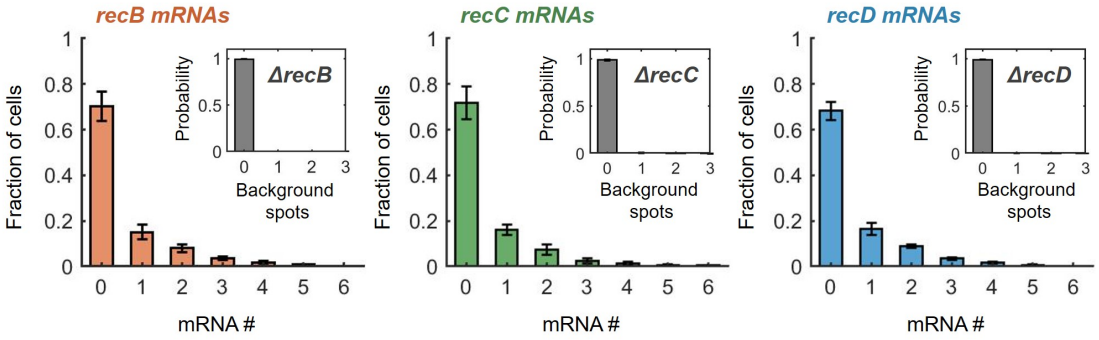

Figure 3.5: All-cell *recB*, *recC* and *recD* mRNA distributions presented in molecule numbers per cell. The histograms represent the average across three replicated experiments; error bars reflect the maximum difference between the repeats. Total number of cells included in the distributions are 15,638 for *recB*; 12,029 for *recC*; 9,365 for *recD*. **Inserts:** Estimation of false positives in the mutants resulted in ~ 0.007 (*recB*), ~ 0.012 (*recC*), ~ 0.016 (*recD*) molecules/cell. Total number of cells included in the distributions are 8,439 for Δ*recB*; 6,432 for Δ*recC*; 5,067 Δ*recD*.

Noticeably, $\sim$70% of cells did not have any *recB*, *recC* or *recD* mRNA molecules. However, it may be argued that it is unlikely to be a result of low sensitivity of the smFISH protocol, as demonstrated above (Fig. 3.2).

3.1.3 The effect of gene replication

Next, *recBCD* transcription is described with a stochastic model of gene expression. However, as has been recently highlighted, one of the caveats of interpreting all-cell distributions emerges from the variation of gene-copy numbers during the cell cycle [39, 174, 251]. Based on the position of the *recBCD* locus on the chromosome, it is expected that the *recB*, *recC* and *recD* genes should experience significant fluctuations in the copy number during the cell cycle.

The synchronization between DNA replication and cell growth cycles in unperturbed conditions [250] allows to use the cell size as a measure of progression through the cell cycle [251]. By analysing mRNA abundance in length-sorted cells, transcription from a constitutive promoter was demonstrated to follow a gene-dosage trend while a gene controlled by a repressor was shown to have an increased peak of mRNA production solely around replication of the gene [251]. The latter effect can be explained, for instance, by replication-induced transcription where the repressor is displaced from DNA by replication machinery [80, 84].

To test whether the *recBCD* transcriptional activity is similar to that of a constitutive or a repressed gene, we analysed the acquired RNA FISH data as a function of cell size (Fig. 3.6). The data, binned by cell area and averaged in size-sorted groups, showed gene-dosage trends for all three genes. Remarkably, *recBCD* mRNA numbers approximately doubled within a narrow cell-size interval ($\sim$ 3.0 – 3.5 μm^2), in agreement with the hypothesis of *recBCD* locus replication. This observation indicates that the *recBCD* mRNA levels follow gene-dosage trends as expected (i) for constitutive

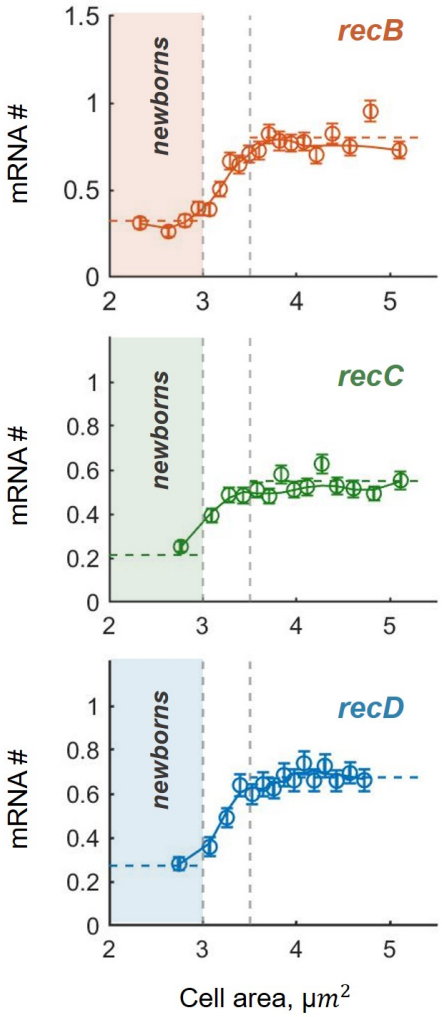

Figure 3.6: *recBCD* mRNA counts as functions of cell area. The circles represent the data binned by cell size and averaged in each group (mean+/-s.e.m.). Solid lines connect the averages across three neighboring groups. Total number of cells: 8,439 (Δ*recB*); 6,432 (Δ*recC*); 5,067 (Δ*recD*).

expression and (ii) for the case of gene copy number
being a limiting factor of transcription. Indeed, if the amount of transcription machinery components such as RNA polymerases is a limiting factor of transcription for a particular gene, that is expected to result in a more gradual increase of mRNA abundance with the volume [123].

In addition, the observation emphasizes that the all-cell mRNA distributions represent the mixture of multiple subpopulations with different gene-copy numbers. In order to separate populations with equal gene-copy number, all cells were partitioned into three size-sorted groups: *newborns* (< 3.0 μm^2), cells in *transition* (between 3.0 μm^2 and 3.5 μm^2) and *adults* (> 3.5 μm^2).

3.1.4 Measurement of mRNA degradation rates

In order to proceed with the analysis of the entire gene expression model, it was initially required to investigate characteristic time-scales of mRNA response. One way of answering this is to scale down the full network to a simpler decay process and measure a degradation rate. This design can be implemented experimentally in a standard protocol, whereby mRNA transcription is inhibited in a pulse manner and mRNA levels are measured in a time-course [30, 44, 205]. Specifically, the initiation of transcription was inhibited with high concentration of an antibiotic, rifampicin, and *recB* and *recC* mRNA distributions were quantified by smFISH at subsequent time points. Assuming a first-order RNA degradation, evolution of mean mRNA numbers was fitted to an exponential function (Fig. 3.7). As the result, *recB* and *recC* degradation rates were estimated as $\gamma_m = 0.62\ \mathrm{min}^{-1}$ for *recB* and $\gamma_m = 0.60\ \mathrm{min}^{-1}$ for *recC*. These values correspond to *recB* mRNA and *recC* mRNA lifetimes of $\sim$ 1.6 min and $\sim$ 1.7 min, respectively. Remarkably, these evaluations are consistent with the genome-wide studies in *E. coli* where transcript lifetimes were shown to be in average $\sim$ 2.5 min across all mRNA species [44, 151].

Taking advantage of the single-cell data, the second moments or variances or the mRNA distributions were further analysed. Specifically, it was investigated whether an evolution of variances is in agreement with the assumption of an exponential decay for *recB* and *recC* mRNAs. To this end, a time-dependent solution describing the evolution of the variance for a given mRNA distribution was derived (Appendix 9.1, Eq. 9.5). The experimental values of the variances were calculated from the full distributions for each time point presented as a function of time and further compared to the analytical predictions (Fig. 3.8). As shown, theoretical predictions describe the experimental data relatively well. Thus, two first moments (means and variances) of *recB* and *recC* mRNA distributions behave as expected for an underlying hypothesis of an exponential decay. This strongly suggests that *recB* and *recC* mRNA degradation can be modelled as a first-order chemical reaction. It is worth noting that exponential mRNA decay is a generally accepted view of transcript degradation in bacterial cells, supported by large-scale experimental studies [158, 203, 205].

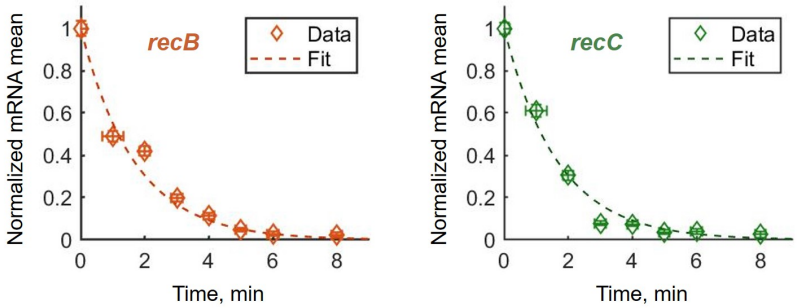

Figure 3.7: *recB* and *recC* degradation rates measured in a time-course experiment where transcription initiation was inhibited with rifampicin. Mean mRNA counts, calculated from the total mRNA distributions for each time point, were normalized to the mean mRNA number at time $t = 0$. Vertical error bars represent the standard error of the mean (s.e.m.); horizontal errors are given by experimental uncertainty on time measurements. Total number of cells for all time points were 20,593 (for *recB*) and 18,292 (for *recC*). The dashed lines are the fitted exponential functions. Based on the fitting functions, degradation rates were calculated with 95% confidence intervals: $\gamma_m = 0.62\ \mathrm{min}^{-1}$ [0.48, 0.75] for *recB*; $\gamma_m = 0.60\ \mathrm{min}^{-1}$ [0.41, 0.78] for *recC*.

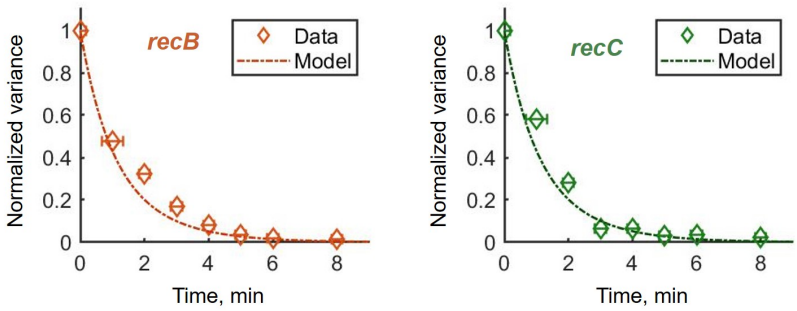

Figure 3.8: *recB* and *recC* mRNA variances, calculated from the entire mRNA distributions for each time point, were normalized to the mRNA variance at time $t = 0$. Horizontal error bars represent experimental uncertainty on time measurements. The dot dashed lines are the predicted analytical functions (Appendix 9.1, Eq. 9.5) based on the fitting results from Figure 3.7. Total number of cells for all time points were 20,593 (for *recB*) and 18,292 (for *recC*).

3.1.5 Stochastic modelling of *recBCD* transcription

Based on the acquired experimental results, *recBCD* transcription was characterized with a stochastic model of gene expression. Specifically, the observed mRNA distributions (quantified in newborns) were compared to theoretical expectation for two basic models of transcription: one-state promoter and two-state promoter in the approximation of bursty transcription. As

discussed in Chapter 1, the constitutive expression model predicts a steady-state mRNA statistics as described by a Poisson distribution, while the model with bursty transcription results in a negative binomial distribution for transcript copy numbers per cell. Therefore, smFISH data from Figure 3.5 for each gene was conditioned on cell size (cell area < 3.0 μm^2, newborns) and the updated distributions were fitted by a Poisson and negative binomial distributions (Fig. 3.9). In each case, the best fit was chosen with the maximum likelihood estimation method. The fitted parameters, λ, r, and p, with the 95% confidence intervals are listed in Table 10.12.

To perform model comparison with different degrees of freedom, the Bayesian and Akaike information criteria were applied. The details are described in Appendix 9.3. The calculated BIC and AIC scores are listed in Table 10.14: a model with a lower score has a better fit. According to both tests (AIC and BIC), bursty transcription model gives a better fit for all three genes regardless of the higher number of parameters. Next, the average mRNA burst size b was calculated using the fitted parameter of a negative binomial p according to the analytical expression Eq. 3.1. It resulted in $\sim$ 0.95, $\sim$ 1.01, and $\sim$ 0.76 molecules for *recB*, *recC*, and *recD* genes, respectively.

$$b = \frac{1-p}{p} \tag{3.1}$$

From the biological perspective, these results can be interpreted as follows: similarity with a negative binomial fit indicates a burst nature of mRNA transcription. Thus, *recB*, *recC* and *recD* genes, being well-fitted by a two-state promoter model, show evidence of burst behaviour. Although the burst sizes are close to one molecule (per mRNA production), strictly speaking these genes cannot be considered as truly constitutive because the time intervals between mRNA productions are not exponentially distributed (as would be expected for a Poisson process). However, such small burst sizes do not strongly support the idea of *recBCD* genes being actively regulated with a transcriptional factor, and hence, these genes can be considered as nearly constitutive.

Finally, the rate of transcription was inferred using *recB* mRNA degradation rate, measured in the time-course experiment (Fig. 3.7). This resulted in $k_m = \gamma_m r = 0.21\,\mathrm{min}^{-1}$ which is remarkably low and close to the lower bound of the reported range of transcription rates in *E. coli*, 0.17-1.7 min^{-1} [38, 184, 248]. Therefore, it is concluded that low abundance of *recBCD* mRNAs can be explained by fast mRNA decay and less frequent transcription in comparison to other mRNA species in a cell.

Correlation of expression from individual copies of the gene has been demonstrated to be dependent on growth rate, expression level, gene location and the details of gene expression network [66, 251]. For example, deviation from independent transcription of each copy of a regulated gene may come from the competition for repressor or activator molecules. To investigate the coupling between gene copies, mRNA expression in the population of adult cells was

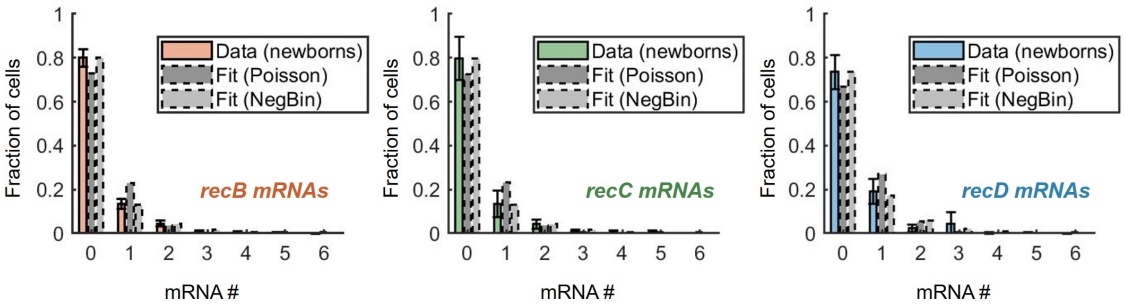

Figure 3.9: Experimental *recBCD* mRNA distributions for newborns (Data) fitted by Poisson (Poisson) and negative binomial (NegBin) distributions. The experimental histograms represent the data from Fig. 3.5 conditioned on cell size (newborns); the average across three replicated experiments are shown; error bars reflect the maximum difference between the repeats. Total number of cells included in the experimental distributions are 2,180 for *recB*, 861 for *recC*, 591 for *recD*. The Kullback–Leibler divergence scores are provided in Table 10.13.

analysed, in particular, whether *recB*, *recC* or *recD* gene activities depend on the presence of the other identical gene copy in a cell. It was hypothesized that transcription from the copies of the same gene occurs independently and the probability functions for mRNA distributions were analytically predicted (Appendix 9.4). The model parameters, inferred from the mRNA statistics in the newborns for each gene, were used for the corresponding analytical predictions in adults. As demonstrated above, the *recBCD* mRNA distributions are well-described by negative binomial distributions, $NB(r,p)$. To make the next step, the following mathematical result was used: the sum of two independent variables, each of which is distributed as $NB(r,p)$, is also described by a negative binomial distribution with parameterization $NB(2r,p)$ (Appendix 9.4). As seen from Figure 3.10, analytically-expected probability distributions ('Model (NegBin)') correspond well to the experimental histograms ('Data (adults)') for *recB* and *recC* genes. These results suggest that the *recB* and *recC* genes, once replicated, are transcribed independently from the gene-copies allocated on the sister chromosomes.

Curiously, there was not such a good match between analytically predicted and experimentally obtained distributions for the *recD* gene (D_{KL} = 0.03, Table 10.13). The prediction based on the fitting to a Poisson distribution for newborns did not result in a better distribution overlay (D_{KL} = 0.15, Table 10.13). This suggests that it might be a result of either (i) a further downstream location of the gene within the *ptrA-recB-recD* operon or (ii) a second promoter additionally controlling *recD* transcription. However, these hypotheses need to be further investigated.

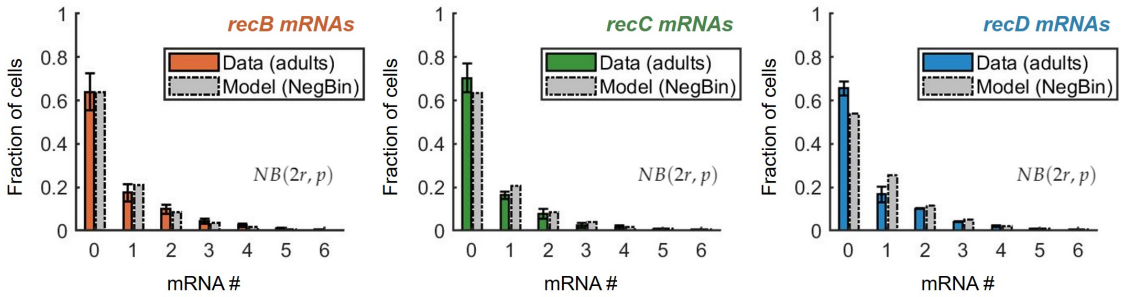

Figure 3.10: Experimental (Data) and analytical (Model) *recBCD* mRNA distributions for adults. The experimental histograms represent the data from Fig. 3.5 conditioned on cell size (adults). Total number of cells used for the experimental distributions: 6,413 for *recB*; 8,876 for *recC*; 5,306 for *recD*. The average across three replicated experiments are shown; error bars reflect the maximum difference between the repeats. The analytical functions are given by predicted negative binomial distributions, $NB(2r, p)$. The Kullback–Leibler divergence scores are provided in Table 10.13.

Variations in mRNA molecule numbers as a result of low abundance and short lifetimes can significantly contribute to gene expression noise [168, 195]. In order to estimate the level of fluctuations for *recBCD* transcription, the Fano factor was used. In our data, the average FF was found to be $\sim$ two times larger than a lower bound given by a Poisson process (FF = 1) for all three genes. A similar observation was made even when analysing the FF in size-sorted groups. As seen from Figure 3.11, Fano factor is fairly constant over a cell cycle and exceeds the Poisson limit in all cell-size groups. In other words, this can be seen as further evidence of burst transcription or two-state promoter model that is known to increase noise of a gene expression network.

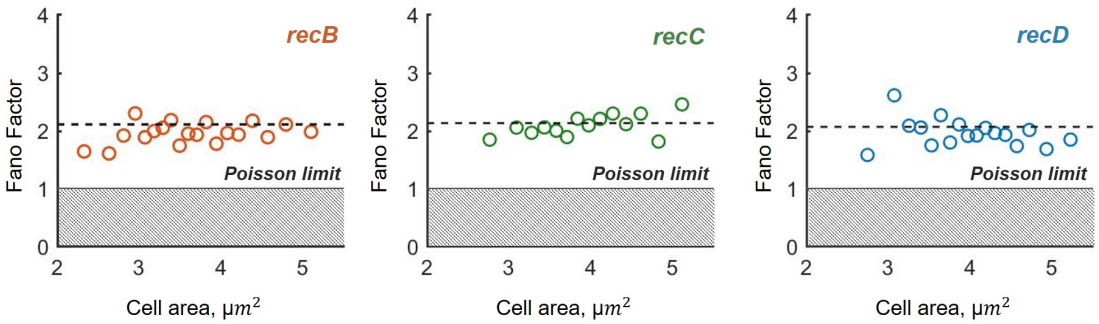

Figure 3.11: The Fano factors of *recB*, *recC* and *recD* mRNA FISH data presented as functions of cell area. The circles represent the data binned by cell size and averaged in each group. The dashed black lines show the average Fano factor across all cells. The average values of Fano factors are ~ 2.11 for *recB*; ~ 2.14 for *recC*; ~ 2.07 for *recD*. The black solid line, $FF = 1$, cuts off the area below the Poisson limit (shaded).

3.2 RecB protein quantification

To investigate RecBCD expression further, proteins were quantified with single-molecule accuracy in individual cells using the HaloTag-labelling technique [119]. The RecB subunit was chosen as a protein of interest because (i) it is an essential component for the complex formation with both helicase and nuclease activity and (ii) fluctuations in RecB numbers are transmitted to the noise in the complex abundance. With a modified image analysis pipeline, low abundance of RecB proteins, reported previously, was initially confirmed [119, 235]. Then, the absence of an active degradation mechanism of RecB subunits was demonstrated by measuring protein removal rate in a pulse-chase experiment. A two-stage stochastic model of RecB expression was further completed by inferring translation rate on the population of newborn cells. Based on the modelling results, it was shown that the level of fluctuations at the RecB protein level exceeds the expected. This observation strongly suggests a regulation mechanism controlling protein production.

3.2.1 HaloTag-labelling set-up and analysis

To perform precise quantification of RecB proteins in single chemically-fixed cells, the HaloTag-labelling protocol which was carried out and optimized, as described in detail by Lepore *et al.* [119]. The study has also shown that the translational RecB-HaloTag insertion does not affect the activity of the RecBCD enzyme. Additionally, the detection efficiency of RecB proteins quantified with the HaloTag labelling in similar conditions has been shown to reach at least 80% [119]. Representative examples of fluorescence images for the strain with RecB-HaloTag fusion and the parental wild-type strain, both labelled the JF549, are shown in Figure 3.12.

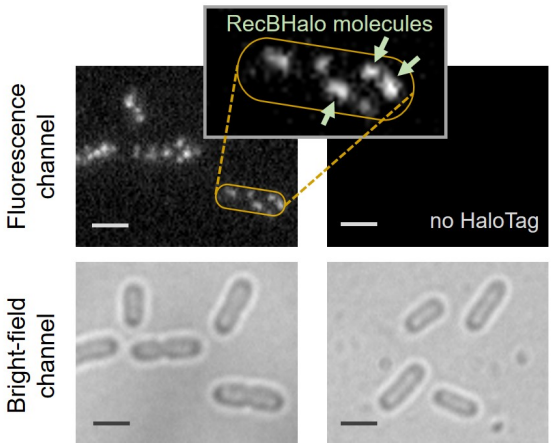

Figure 3.12: Examples of fluorescence and bright-field images for the strain with the RecBHalo fusion and its parental (no HaloTag) strain. Both samples were labelled with JF549 dye and the images are shown in the same intensity range. Scale bars represent 2 µm. **Zoom-in:** An example of a cell with five RecBHalo molecules (several single molecules are shown with light green arrows).

To detect fluorescent single-molecule foci, the Spätzcell pipeline was modified specifically for HaloTag-JF549 spots (see Chapter 2 for the details). In brief, the first step of spot processing has been adjusted for single-molecule spots. Two options have been tested: (i) removing a Guassian filtering;

(ii) adding a Laplace deconvolution filter. It is worth noting that a similar modification was recently implemented for protein detection by FROS (FROS: Fluorescent Repressor Operator System: an operator array was inserted next to the gene of interest and was bound by the cognate fluorescently tagged repressor allowing for visualisation of the gene locus) [251]. As a result of low-abundance of the protein of interest and single-molecule labelling, the analysis of integrated intensity was skipped for HaloTag-labelled images. All other steps of the Spätzcell software remained unchanged and did not require significant optimizations.

Similar to smFISH analysis, peak height intensity profiles in the RecB-HaloTag and the no-HaloTag samples allowed us to separate specific signal from the background (Fig. 3.13, left). Indeed, peak height distribution is bimodal in the RecB-HaloTag sample while no-HaloTag control has a single mode. The separation between two populations was further confirmed with a correlation plot between spot area and peak height (Fig. 3.13, right). In fact, the RecBHalo sample has two subpopulations of spots that are well separated in the two-dimensional space (spot area versus peak height), whereas false positive spots display high diversity in spot size. However, it is argued that based on the intensity threshold only, the population of background spots can be removed from both the peak intensity histogram and the correlation plot.

After removing the foci which are below the threshold, spot positions with cell segmentation masks were matched and the copy number of RecB spots per cell was calculated. As a result, RecB protein distribution for cells of all sizes was obtained and the low abundance of molecule numbers was confirmed: 3.9 ± 0.6 molecules/cell. These results are in agreement with previous quantification [119] and it is worth mentioning that the study quantified RecB proteins using two independent single-molecule methods: HaloTag-labelling with TMR; and MACS-based with the RecB-sfGFP fusion (MACS: Microfluidics Assisted Cell Screening: a microfluidic device that allows for mechanical compression of bacterial cells. This slows down the diffusion of cytoplasmic molecules and thus improves its detection with single-molecule fluorescence microscopy [164]). The resulting data from the methods are quantitatively similar. Consistency of the presented quantification with the reported data from [119] confirms the accuracy of the utilized protocols and image analysis pipelines.

3.2.2 RecB protein distributions

Similar to the mRNA FISH distribution analysis, two groups based on cell size were analysed: newborns and adults. The distributions for these two groups are shown in Figure 3.14 alongside the RecB distribution for all cells. While many false positive spots were removed based on the intensity threshold, it is unavoidable to have some unspecific signal left. Indeed, these false positive spots might correspond to remaining dye molecules in the no HaloTag control, and hence, have the same characteristics as specific foci present in the RecBHalo sample. The estimation of such false positives in the no HaloTag control resulted in $\sim$0.3 molecules/cell.

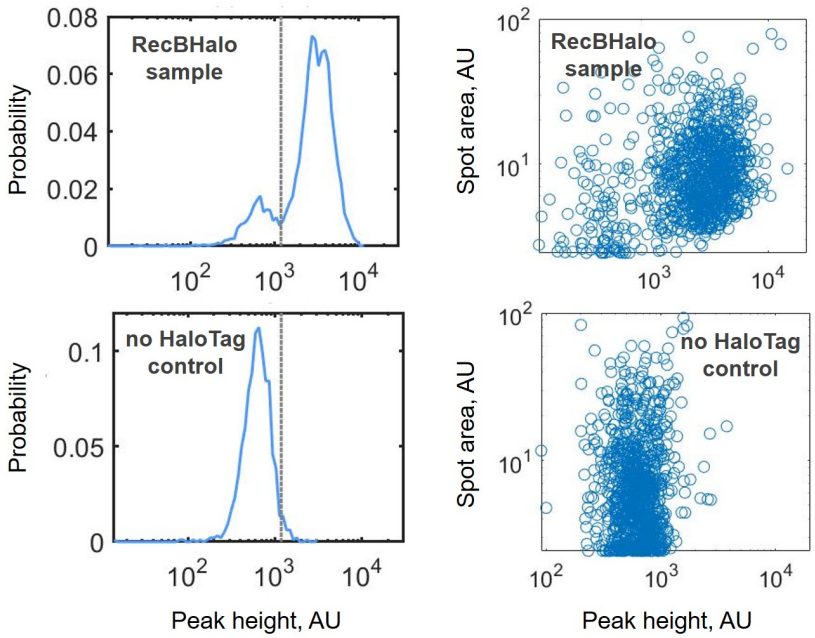

Figure 3.13: Left: Peak height intensity profiles for the spots detected in the strain with RecB-HaloTag fusion (RecBHalo sample) and its parental strain (no HaloTag control). The dashed line indicates a chosen intensity threshold in order to separate a specific signal from the background (96%). **Right**: Correlation plots between peak height intensity and spot area for the same samples. Each circle corresponds to a single spot. Random selection of spots were plotted for visualization purposes.

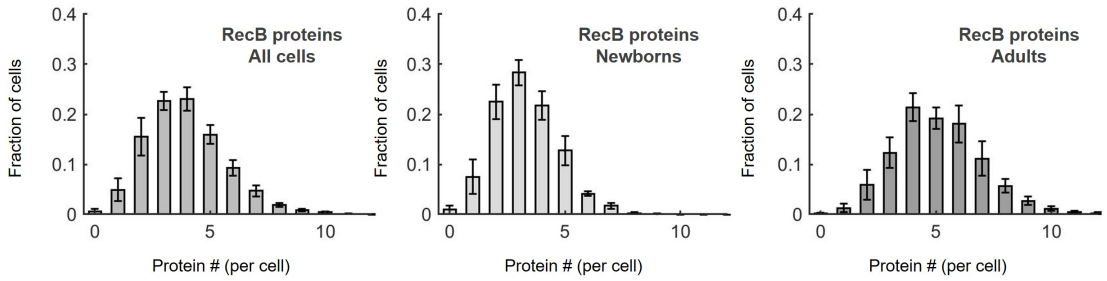

Figure 3.14: RecB distributions for all cells, newborns and adults given in molecules per cell. The histograms represent the average of three experiments; error bars reflect the maximum difference between the repeats. Total number of analysed cells is 10,699 (all cells); 5,477 (newborns); 2,781 (adults).

3.2.3 Measurement of the removal rate

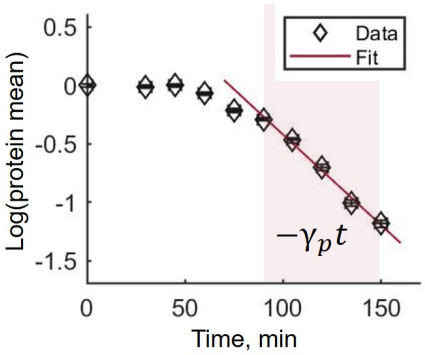

Figure 3.15: RecB removal rate measured in a pulse-chase HaloTag-labelling experiment. Mean protein counts, normalized to the mean protein number at $t = 0$, presented in the natural logarithmic scale. Total number of cells for all time points was 10,548. Shaded area shows the time interval used for fitting. The red line is the fitted function.

To test whether RecB is actively degraded by the proteasome or mainly diluted as a result of cell division, the RecB protein removal rate was measured. This is usually obtained from radiolabelled amino acid or translation inhibitor pulse-chase assays [59, 63, 262] which either requires radioactive labels/antibodies or leads to strong cellular perturbations. While the first approach might not be sensitive enough to accurately detect the changes for a low abundant protein, the latter may affect the measurements on long timescales.

Therefore, an alternative approach was adopted [144, 261], where the HaloTag ligand was used to pulse-label RecB molecules, then the excess dye was removed by extensive washing and finally the abundance of the labelled molecules over time was measured at the single-cell level. Assuming protein removal to be described by a first order reaction, mean protein counts were calculated from the total distributions for each time point, normalized to the mean protein number at time $t = 0$ and represented in the natural logarithmic scale. Then, mean protein counts over time were fitted to an exponential function on a subset of time points (Fig. 3.15, shaded area). The decay rate was extracted from the fit and resulted in $\gamma_p = 0.015\ \mathrm{min}^{-1}$ with the 95% confidence interval [0.011, 0.019] min^{-1}. Comparison to the population growth rate in these growth conditions ($\sim$ 0.017 min^{-1}, 95% confidence interval [0.016, 0.018] min^{-1}) suggests that RecB protein is stable and effectively degraded only as a result of dilution and molecule partitioning between daughter cells.

3.2.4 Stochastic modelling of RecBCD expression

In the final part of this chapter, the transcription model, described above, was extended by taking into account protein synthesis and removal. As discussed earlier, only the newborns were analysed and the rate of translation was estimated as following. The average protein count number in newborns ($\sim$ 3.2 molecules per cell) was matched to the theoretical expression for the protein mean. The rate of translation k_p was calculated (here the protein removal rate, γ_p = 0.015 min^{-1}, and mean mRNA number in newborns, $\langle m \rangle = 0.323$ were used):

$$k_p = \frac{\gamma_p \langle p \rangle}{\langle m \rangle} = 0.15\,\mathrm{min}^{-1} \tag{3.2}$$

This value fell below the lower bound of the reported translation rates in *E. coli* (0.25 – 1 min^{-1}) [38, 184]. Therefore, it was concluded that *recB* mRNAs are translated at low levels compared to other proteins in a cell.

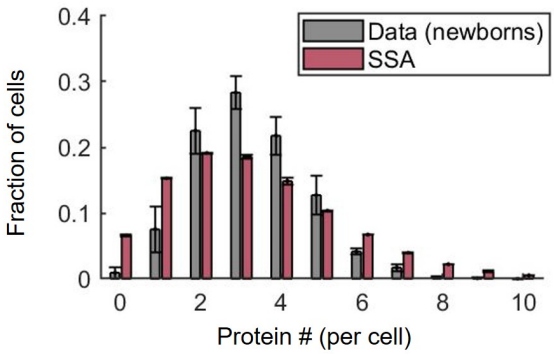

Figure 3.16: Comparison between the experimental RecB molecule distribution for newborns (Data) and the results of Gillespie's simulations (SSA) for the two-stage model. Total number of cells is 5,477; error bars represent the maximum difference between the repeated experiments. Parameters used in simulations shown in Table 10.15. The Kullback–Leibler divergence between the distributions is $D_{KL} = 0.15$.

Based on the constructed model of gene expression, it was decided to examine whether the model can reproduce the cell-to-cell variability in the protein levels observed in our experiments. To this end, the simulation approach was utilized and single trajectories of protein abundance over time were generated using the Gillespie's algorithm for the two-stage model with the estimated parameters. Strikingly, the simulated distribution had larger fluctuations than the experimentally measured one (Fig. 3.16). While both protein means were equal as a result of the first-moment-based parameter estimation, it was found that the standard deviations were different: 2.2 (simulations) and 1.5 (experiment). In other words, a two-stage model of protein production predicts a higher level of fluctuation than that detected experimentally.

This phenomenon can be also analysed using the squared coefficient of variation of the protein distribution, CV_p^2, defined as $\sigma_p^2/\langle p \rangle^2$. For a two-stage gene expression model with burst transcription, it has been established that the squared coefficient of variation is given by the following expression [167, 171, 207, 236]:

$$CV_p^2 = \frac{1}{\langle p \rangle} + \frac{1+b}{\langle m \rangle} \frac{\gamma_p}{\gamma_m + \gamma_p} \tag{3.3}$$

where $\langle p \rangle$ and $\langle m \rangle$ are the mean protein and mRNA counts (per cell), b is the average (mRNA) burst size, γ_m and γ_p are the mRNA degradation and protein dilution rates, respectively.

It was found that the experimental coefficient of variation (CV^2_{Data} = 0.19 $\pm$ 0.06) is approximately half of the value that analytically predicted (CV^2_{Theory} = 0.45). As expected, both coefficients of variation, theoretically predicted, CV^2_{Theory}, and computed from the simulated data, CV^2_{SSA}, are in a good agreement while the experimental value, CV^2_{Data}, significantly differs (Fig. 3.17). This deviation from the theoretical prediction implies a potential regulatory mechanism of RecB expression that actively suppresses variation in protein production in order to maintain protein levels within a certain range.

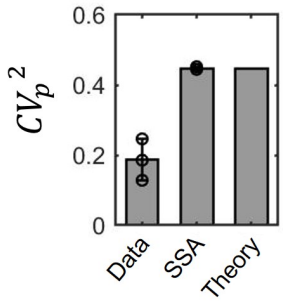

Figure 3.17: Comparison of the coefficients of variation, for experimental (Data), simulated (SSA) data and analytical prediction (Theory). Error bars represents standard deviations across three experiments/simulations; the black circles show the mean CV^2_p in each experiment/simulation. Theoretical prediction was calculated using Equation 3.3.

3.3 Discussion

RecBCD is a key enzyme in DNA DSB repair in *Escherichia coli*. Nevertheless, the details of its expression remain largely unknown. In this chapter, RecBCD expression in wild-type cells has been revisited with single-molecule methods. Consistent with several studies reported on the abundance of RecBCD proteins [91, 119, 234], RecB subunits were demonstrated to be present in average in $\sim$4-5 molecules per cell. In addition to the average levels, usage of a single-molecule approach made possible to evaluate protein noise. The variability of protein copies per cell was found to be unexpectedly low. As outlined in Chapter 1, an effective strategy to reduce protein fluctuations is to have high mRNA levels which are inefficiently translated. Surprisingly, that turned out not to be the case for RecB. According single-molecule RNA FISH quantification, *recBCD* mRNA levels were found to be (i) even less abundant (less than one molecule per cell) and (ii) noisier than expected for a random process. These results taken together, led to the hypothesis of a post-transcriptional regulation mechanism suppressing noise at the protein level.

A couple of additional observations may be worthy attention for future studies. First, taking the advantage of single-molecule quantification, verification of an exponential mRNA decay can be extended beyond the first two moments. An exact time-dependent solution for the entire mRNA probability distribution was derived for several initial conditions (Appendix 9.2). Specifically, a Poisson and negative binomial were considered as initial distributions at time $t = 0$ to account for the steady-state distributions of a constitutive and bursty transcription models. While analytical derivations have been performed, the final results need to be verified with simulations. Then, fitting to the full experimental distributions need to be carried out. Another way to test a first-order *recBCD* mRNA decay is to consider an alternative model with double exponential degradation [24]. In this case, further comparison to a single exponential decay process needs to be performed taking into account the difference in the number of fitting parameters with, for example, Bayesian information criteria. Such detailed investigation of *recBCD* degradation might be useful in the context of perturbed conditions.

Secondly, stochastic modelling of *recBCD* transcription may have been addressed in a more general way. Indeed, one might argue that an exact solution for the probability function derived for the two-state promoter model (Eq. 1.7, [170]) should be used instead of an approximated regime of bursty transcription. While this seems to be reasonable at first glance, fitting procedure to a hypergeometric function is computationally difficult as many various parameter sets describe the experimental distribution similarly well. This appears to be due to the fact that the experimental mRNA distributions for *recBCD* genes are mainly concentrated around two values: zero molecule per cell ($\sim$ 70% of the data) and one molecule per cell ($\sim$ 20%). While these values and, especially, the difference between them, is still informative for one- and two-state promoter model comparison, it does not provide sufficient amount of information to identify all four parameters of the initial model [186]. In order to be able to describe the system with the full model, more data needs to be obtained. Preferably, time course experiments shall be performed in order to capture the dynamic response of the system.

It is worth emphasising again that stating a biological interpretation based on the fitting to analytical expressions needs to be done very carefully. Because many mechanisms can be formulated in similar mathematical terms, biologically-different models might end up being described by similar analytical expressions. From this perspective, it is impossible to identify which mechanism is behind of an experimental distribution without any other additional information.

As previously discussed, mRNA abundance analyzed as a function of cell size can be informative to narrow down molecular details of a transcriptional mechanism. In a recent study [251], Wang *et al.* quantified nascent mRNA molecules per cell for two promoters, the phage lambda $\mathrm{P_R}$ and the lactose promoter $\mathrm{P_{lac}}$, and analyzed the results in length-sorted groups (Fig. 3.18, A). Two different trends were found. In the case of a strong constitutive promoter $\mathrm{P_R}$, transcription followed the gene dosage. In contrast, for a gene controlled by a repressor, the

increased peak of mRNA synthesis occured around replication of the *lac* gene (Fig. 3.18, B). As *recBCD* transcription shows a clear gene dosage trend (Fig. 3.6), it leads to the interpretation that these genes are constitutively expressed and are not regulated with a repressor at the transriptional level.

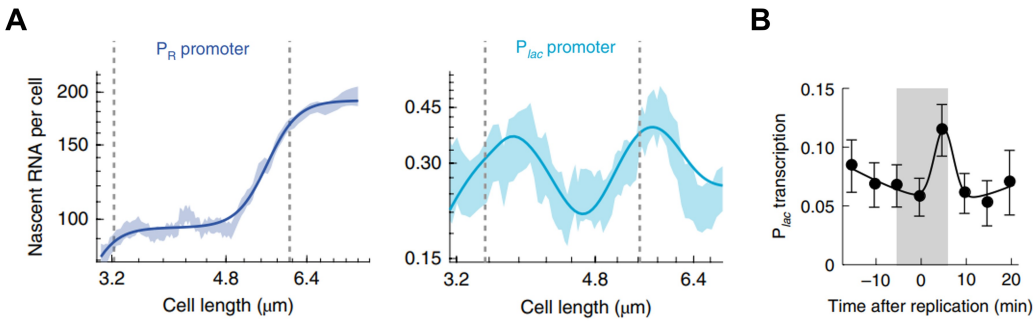

Figure 3.18: mRNA abundance as a function of cell length for two different promoters ($\mathrm{P_R}$ and $\mathrm{P_{lac}}$) and $\mathrm{P_{lac}}$ transcription activity in relation to the event of gene replication. The figures have been taken from the study by Wang *et al.* [251]. **A:** Nascent mRNA number per cell were measured with smFISH for the phage promoter $\mathrm{P_R}$ (left) and the lactose promoter $\mathrm{P_{lac}}$ (right). The data were binned in groups by cell length. **B:** Transcription activity of the $\mathrm{P_{lac}}$ promoter shows an increased peak right after replication of the *lac* gene.

Several points regarding protein quantification with HaloTag-labelling should be made. As reported in the study [119], labelling efficiency of HaloTag-labelling protocol (for RecB proteins) was demonstrated to be at least 80%. However, that can be argued that while imperfect labelling affects the absolute quantification, that should not significantly influence the relative quantification such as CV_p^2. Next, if should be noted that the absolute quantification of protein numbers is complicated by threshold uncertainty in the image analysis pipeline. Even in the cases when the peaks corresponding to the background and HaloTag-specific signals are well separated, it is difficult to choose an unequivocal threshold value because some foci in the no-HaloTag control would have the same characteristics as specific spots. Other labelling strategies and/or spot detection algorithms may be able to address this issue in the future.

In the presented stochastic model, experimental noise was assumed to be negligible. However, experimental protocols and data acquisition are never noiseless and might significantly affect detected variability of final outcomes. Therefore, modelling of data acquisition process might be worth considering in more detail and further taking into account. This is especially relevant when the interpretation of results is mainly based on average values and variances of the data.

Finally, the evidence of noise suppression for RecB subunits would be stronger established with an additional experimental control. For example, single-molecule quantification of an unregulated low-copy protein, which is expected to have high level of fluctuations, may be recommended. A trial to build such construct was attempted where a Halo protein was placed under control of a very weak synthetically-designed promoter/RBS [114] and integrated into *E. coli* chromosome. The strength of promoter/RBS region was calculated based on ribosome calculator estimator [9]. Unfortunately, the construct resulted in high production of Halo molecules what did not allow one-to-one side comparison to the RecB quantification.

Chapter 4

RecB mRNA and protein quantification upon DNA damage

In this section, the question whether there is any evidence of RecBCD expression regulation once *Escherichia coli* cells are exposed to stress is addressed. The initial hypothesis is that RecB mRNA and/or protein levels may be upregulated upon DNA damage in order to enhance their chromosome repair capability. To test this, bacterial cells were treated with a sub-lethal concentration of ciprofloxacin, an antibiotic that produces DNA double-strand breaks, and RecB protein and mRNA levels were quantified by HaloTag-labelling and smFISH, respectively. Regardless of the decrease in mRNA concentration upon DNA damage, RecB protein levels remain surprisingly constant. In other words, RecB proteins are translated more efficiently upon DSB induction. In line with the results of the previous chapter, the uncoupling between mRNA and protein production further supports a post-transcriptional regulation mechanism that maintains RecB protein level within an optimal range.

4.1 Experimental set-up

Aiming to investigate the behaviour of RecB expression network upon a perturbation, bacterial cultures were exposed to an antibiotic, ciprofloxacin, and protein and mRNA quantification was performed in single cells (Fig. 4.1). Ciprofloxacin belongs to the group of quinolone antibiotics and is known to target prokaryotic topoisomerases by forming irreversible DNA-topoisomerase complexes. A sub-lethal concentration of ciprofloxacin was used to induce DSBs while keeping cells alive and allowing for DNA repair to occur. After exposure to ciprofloxacin for two hours, cell cultures were separated and mRNA or protein levels were quantified according to smFISH or HaloTag-labelling protocol. While cell cultures were separated into two fractions, in both protocols cells were collected at a similar phase of growth. Then, the mean protein and mRNA levels were quantified from the corresponding distributions and the average protein-to-mRNA ratio was calculated in both samples. This ratio was named the average translational efficiency, which reflects the number of proteins in average produced per mRNA.

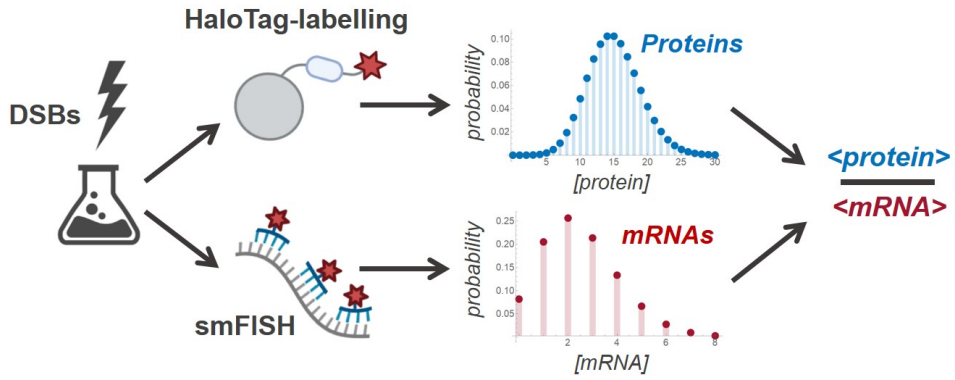

Figure 4.1: Schematic of the experimental workflow. Cell cultures were exposed to a sub-lethal concentration of ciprofloxacin for two hours, then separated into two fractions, and mRNA or protein levels were quantified according to smFISH or HaloTag-labelling protocol. Finally, the average protein-to-mRNA ratio was calculated in intact and perturbed conditions.

4.1.1 Cell elongation and growth rate

Although the vast majority of bacterial cells survive sub-lethal concentration of ciprofloxacin (4 ng/ml was used while the minimum inhibitory concentration is 16 ng/ml), cells may experience significant physiological changes. As has been previously reported, *E. coli* bacteria inhibit cell division and start filamenting upon activation of the bacterial SOS response, a global response to DNA damage [69, 93, 127]. An example of such elongation occurring when cells are treated with 4 ng/ml of ciprofloxacin for two hours in comparison to the intact control is shown in Figure 4.2.

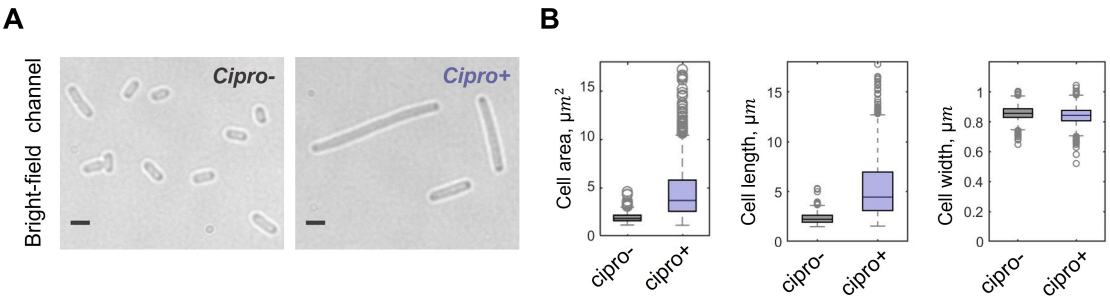

Figure 4.2: A: Representative examples of bright-field images for unperturbed (Cipro^{-}) and perturbed (Cipro^{+}) conditions. Scale bars represent 2 µm. **B:** Box plots with cell area, cell length and cell width distributions for both samples. Total number of cells included into box plots was 3000 for each sample.

To evaluate cell size difference between the samples, statistical analysis of cell area was performed (Fig. 4.2B). As expected, a larger cell area median in the perturbed sample was observed: 3.6 μm^2 upon the treatment compared to 1.9 μm^2 in the control. Additionally, a larger variability was detected in the perturbed cells. This might partially arise from different levels of DNA damage in single cells [104]. While cell area and length were increased upon ciprofloxacin treatment, cell width was not affected by the perturbation. This is consistent with many previous observations of *E. coli* cells maintaining their diameter while elongating in length.

To take into account cell size difference between damaged and intact samples, single-cell distributions of mRNA and protein concentrations were determined instead of absolute molecule numbers. mRNA and protein concentrations were calculated as molecule number normalized by cell area. It is worth noting that cell area is a reasonable proxy for cell volume because of *E. coli* maintenance of cell diameter while elongating in length (Fig. 4.2B).

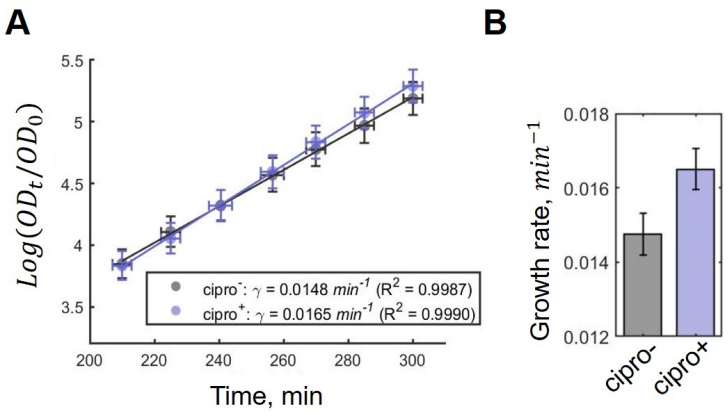

Figure 4.3: Growth curves, obtained from OD_{600} measurements and represented in logarithmic scale (**A**), and estimation of growth rates at the exponential phase (**B**). Ciprofloxacin (4 ng/ml) was added to the cells at 180 min. Three replicated experiments have been performed for each condition; error bars represent 95% confidence intervals.

Next, the effect of ciprofloxacin treatment on populational growth rate was examined. Ciprofloxacin (4 ng/ml) was added to wild-type cell culture at an early exponential phase and further growth was monitored with optical density measurements (Fig. 4.3). An exponential phase of growth was used to perform fitting and infer growth rates in Figure 4.3B. As a result, the exponential growth rates were estimated as $\sim$ 0.0148 min^{-1} [0.0142, 0.0153] min^{-1} in intact conditions and $\sim$ 0.0165 min^{-1} [0.01560, 0.0170] min^{-1} upon ciprofloxacin treatment.

The growth rate upon ciprofloxacin treatment was found to be slightly increased compared to the intact conditions. However, this is most likely to be an artefact of populational optical density measurements for elongated cells [223]. As demonstrated by Stevenson *et al.*, calibration function between optical density and cell number is significantly altered for filamentous cells. Therefore, it would be recommended to improve growth rate measurements by either a direct counting of cells with bright-field microscopy or tracking single cells on an agar pad/in a mother machine [104].

4.1.2 Time-scales of mRNA and protein response

As was shown above, RecB proteins and *recB* mRNAs have substantially different lifetimes. This results in different time-scales in response to a perturbation. Thus, the following question was addressed: how long does the DSB induction need to be applied for to see changes (if any) in both species? To answer this, the experimental workflow was described with a simple deterministic model (Appendix 9.5). Based on the parameters measured above, a time-dependent solutions for mean RecB mRNA and protein concentrations after a perturbation were derived (Eq. 9.39 and Eq. 9.41). An elementary analysis of the obtained expressions shows that mRNA solution has one mode, $e^{-(\gamma_m+\gamma)t}$, and hence, a characteristic response time-scale is defined as $1/(\gamma_m+\gamma) \sim 1.6$ min, where γ is the growth rate. In contrast, the expression for protein concentration is given by two time-scales: $e^{-\gamma t}$ and $e^{-(\gamma_m+\gamma)t}$, where the slowest one defines the response time as $1/\gamma \sim 61$ min. This means that a perturbation needs to be applied for at least for one hour in order to be able to detect changes in protein concentrations. Therefore, in the described experiments, DSBs were induced for two hours to guarantee the fulfillment of the this requirement.

4.2 The RecB protein-to-mRNA ratio is decoupled upon DSB induction

4.2.1 RecB protein quantification by HaloTag-labelling

Following the experimental workflow described above, RecB proteins were quantified in single cells and the results were normalized by cell area. It was observed that RecB protein concentration does not change upon DSB induction (Fig. 4.4). In fact, the average cellular RecB concentration was surprisingly very conserved: 2.28 mol/μm^2 (95% confidence interval is [2.21, 2.32] mol/μm^2) in perturbed condition and 2.12 mol/μm^2 (95% confidence interval is [2.01, 2.32] mol/μm^2) in unperturbed one. This non-significant difference between the average RecB concentrations was verified by two-sample *t*-test (*P* value = 0.36).

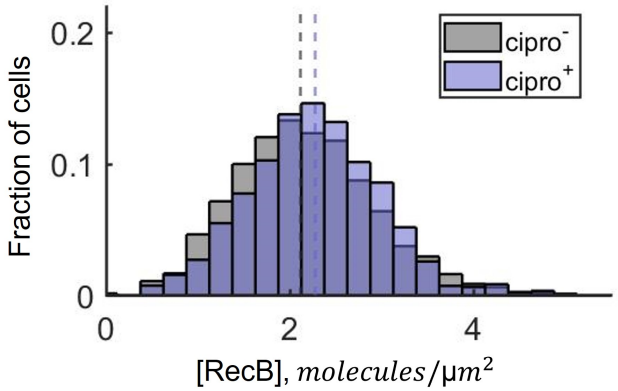

Figure 4.4: RecB protein concentration distributions in unperturbed (grey) and perturbed (blue) conditions. The histograms represent the average of three experiments. The medians are shown by dashed lines: 2.12 mol/μm^2 (cipro^-) and 2.28 mol/μm^2 (cipro^+). Total number of analysed cells are 1,534 (cipro^-) and 683 (cipro^+).

Additionally, RecB concentration was analysed in size-sorted groups of cells (Fig. 4.5). Regardless of moderate fluctuations in protein concentration over a cell cycle, both samples significantly overlap along the whole range of cell sizes. This observation confirms that RecB protein concentration remains fairly constant over a cell cycle and similar to the wild type in perturbed conditions.

4.2.2 *recB* mRNA quantification by smFISH

Next, *recB* mRNA levels were analysed. The obtained mRNA concentration distributions are shown in Figure 4.6. In contrast to the results of protein quantification, *recB* mRNA concentration showed a sharp decrease upon DNA damage. Specifically, mRNA concentrations upon DSB induction is $\sim$2.4-fold lower than in the unperturbed control: 0.12 mol/μm^2 (95% confidence interval is [0.06, 0.17]) upon DNA damage versus 0.28 mol/μm^2 (95% confidence interval is [0.23, 0.32]) in the control. Verification of the difference between the average *recB* mRNA concentrations by two-sample *t*-test confirmed statistical significance (*P* value = 0.0023).

This population-observed decrease was further confirmed by analysing mRNA concentration in cell-size sorted groups (Fig. 4.7). The results showed that the *recB* mRNA level under ciprofloxacin exposure was lower than the one in unperturbed cells in all cell-size sorted groups.

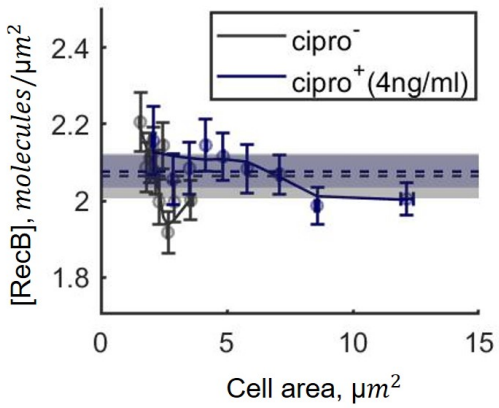

Figure 4.5: RecB protein concentration in unperturbed (grey) and perturbed (blue) samples shown as a function of cell area. The data from Figure 4.4 were binned by cell size and averaged in each group (mean±s.e.m.). The solid lines connect the averages while the dashed lines and shaded areas show mean±s.e.m. calculated across all cells in each sample. Total number of analysed cells are 1,534 (cipro^-) and 683 (cipro^+).

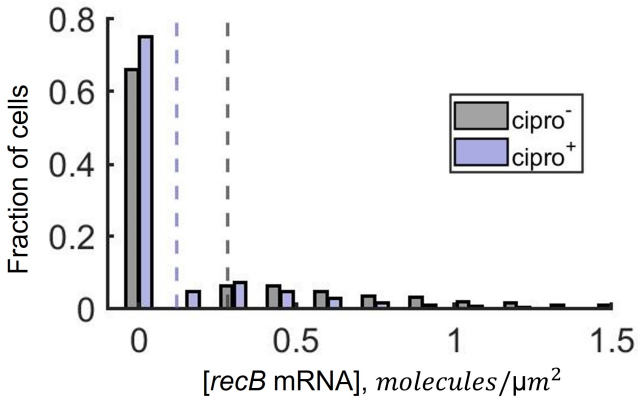

Figure 4.6: *recB* mRNA concentration distributions for intact (grey) and damaged (blue) samples. The histograms represent the average of three experiments. The medians of the *recB* mRNA concentrations are shown by dashed lines: 0.28 mol/μm^2 (cipro^-) and 0.12 mol/μm^2 (cipro^+). Total numbers of analysed cells are 11,700 (cipro^-) and 6,510 (cipro^+).

4.2.3 Estimation of translational efficiency

Finally, the average number of proteins produced per mRNA (translational efficiency) in both conditions was determined. The ratio between the average protein and mRNA concentrations in damaged cells is 19.7 ± 5.3, whereas it is 8.1 ± 2.0 proteins/mRNA in intact cells (Fig. 4.8). Thus, a $\sim$2.4 increase in translational efficiency was detected upon DSB induction.

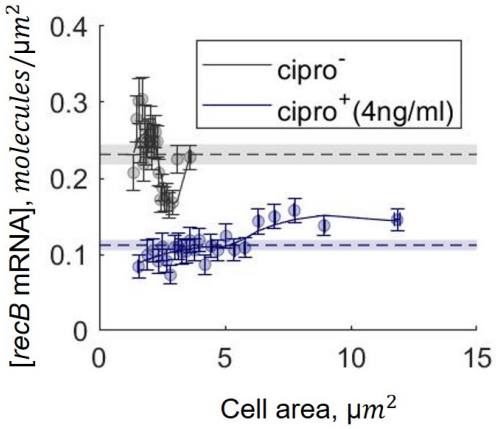

Figure 4.7: *recB* mRNA concentration for perturbed (blue) and unperturbed (grey) samples represented as a function of cell area. The data from Figure 4.6 were binned by cell size and averaged in each group (mean±s.e.m.). The solid lines connect the averages while the dashed lines and shaded areas show mean±s.e.m. calculated across all cells in each sample. Total numbers of analysed cells are 11,700 (cipro^-) and 6,510 (cipro^+).

Taken together, the results suggest a post-transcriptional regulatory mechanism that maintains a constant RecB protein level in response to the fluctuations in mRNA production.

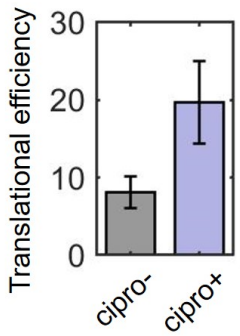

Figure 4.8: The average number of proteins produced per one mRNA in intact (grey) and damaged (blue) conditions (translational efficiency). Average translational efficiency, TE, (for each condition) was calculated as the ratio between the mean protein concentration, c_p, and the mean mRNA concentration, c_m. The uncertainty on translational efficiency values (shown with error bars) was calculated as $\Delta TE = TE\sqrt{(\Delta c_p/c_p)^2 + (\Delta c_m/c_m)^2}$, where Δc_m and Δc_p are standard deviations of mRNA and protein concentrations, respectively, across three repeated experiments.

4.3 Discussion

Based on the evidence of noise suppression in intact conditions, RecBCD expression has been investigated upon stress. It was found that upon ciprofloxacin treatment *recB* mRNA concentration decreases while RecB protein level remains the same as in unperturbed conditions. In other words, in average twice more RecB proteins produced per a *recB* transcript upon DNA DSB induction.

As previously reported, *recB* gene is not included into the SOS regulon [249], and hence, not necessarily expected to be up-regulated upon DNA damage. However, down-regulation of *recB* transcription is an unexpected observation. Several explanations can be proposed. First, once SOS response programme is triggered, $\sim$40 genes, that are normally repressed by LexA, would promptly require RNA polymerases and other components of transcription machinery. That might result in non-SOS response genes being down-regulated due to depletion of shared pool of resources.

Another potential explanation is based on the idea that elongated cells may have lower DNA concentration content. Upon sub-lethal treatment with ciprofloxacin, bacterial cells inhibit cell division but keep increasing their volume at a similar pace. In contrast, DNA replication is highly affected by DSB induction, activation of SOS response and cell division inhibition. In fact, treated cells were demonstrated to have multiple unsegregated chromosomes [183]. Intuitively, perturbation of DNA synthesis is likely to result in lower concentration of genomic DNA in filamented cells. However, to test this hypothesis genomic content upon ciprofloxacin exposure needs to be quantified with a single-molecule approach.

In addition to the decrease in average mRNA level, mRNA-cell size dependence was notably different for the ciprofloxacin-treated sample (Fig. 4.7). When cells are not damaged, a periodic trend in the average mRNA concentration was observed and explained by *recB* gene replication and cell division. In contrast, *recB* mRNA level in bacterial cells treated with ciprofloxacin shows a homeostatic trend: especially for cell sizes from 1 to 5 μm^2, mRNA level remains fairly stable. It is likely to be a consequence of disruption of a normal cell cycle. Although it still remains unknown how ciprofloxacin influences replication and transcription in detail, it clearly affects the interlink between replication and division cycles leading to less synchronized patterns.

Finally, RecB protein concentration remains constant in both, treated and untreated, samples suggesting maintenance of RecB concentration in different conditions. Protein homeostasis has been described mainly in intact cells and potential mechanisms behind it have been investigated [123, 152, 232]. For example, cell size homeostasis has been recently proposed as one of potential hypothesis for the conservation of protein concentration [211]. Interestingly, RecB protein concentration distribution was found to be even narrower upon ciprofloxacin treatment (Fig. 4.4). This may be a consequence of spatial self-averaging when cells increase their size [193].

Chapter 5

Post-transcriptional regulation of RecB expression

The idea of a potential post-transcriptional regulation mechanism was further investigated. Specifically, Hfq protein attracted attention as a possible regulator of RecBCD expression because (i) it is shown to regulate a vast number of mRNAs [97, 238, 265] and (ii) its activity has been directly linked to DNA damage and SOS responses [26, 45, 183].

In collaboration with Dr Ira Alexandra Iosub and Dr Sander Granneman, University of Edinburgh, an available high-throughput Hfq-CLASH dataset with the interactions between Hfq protein and bound RNAs has been analysed [100]. Two Hfq-binding sites were identified within *ptrA-recB-recD* mRNA with the major peak being located close to the ribosome binding site (RBS) of *ptrA-recB*. Down-regulation of RecB expression by Hfq was established *in vivo* by quantifying RecB mRNA and protein levels in *hfq* mutants. These observations were summarized in a molecular post-transcriptional regulation mechanism where Hfq inhibits the initiation of RecB translation by binding to the RBS of the *recB* mRNA.

The suggested model was further tested by modulating the Hfq cellular level *in vivo*. First, an expected decrease in RecB production has been detected when Hfq protein was expressed in Δ*hfq*. Secondly, Hfq molecules were sequestered by over-expression of a highly abundant small RNA called ChiX. Consistent with the proposed model, RecB protein distribution was shifted to higher values when ChiX was overproduced. Additionally, quantifying RecB expression in *chiX* mutants did not reveal any evidence of direct ChiX involvement in the regulation mechanism.

5.1 Identification of Hfq binding sites within *recBCD* locus

To identify RNA molecules interacting with RNA binding proteins (RBPs) *in vivo* and detect the RNA-binding sites, a number of high-throughput techniques have been developed in the past (reviewed in Chapter 1). To assess whether Hfq is involved in post-transcriptional regulation of RecBCD, an available Hfq-CLASH dataset was analysed. There, the transcriptome-wide Hfq binding sites were identified in *E. coli* using CLASH [100]. During the protocol, Hfq-RNA complexes are cross-linked *in vivo* and purified under highly stringent conditions, ensuring specific recovery of Hfq RNA targets.

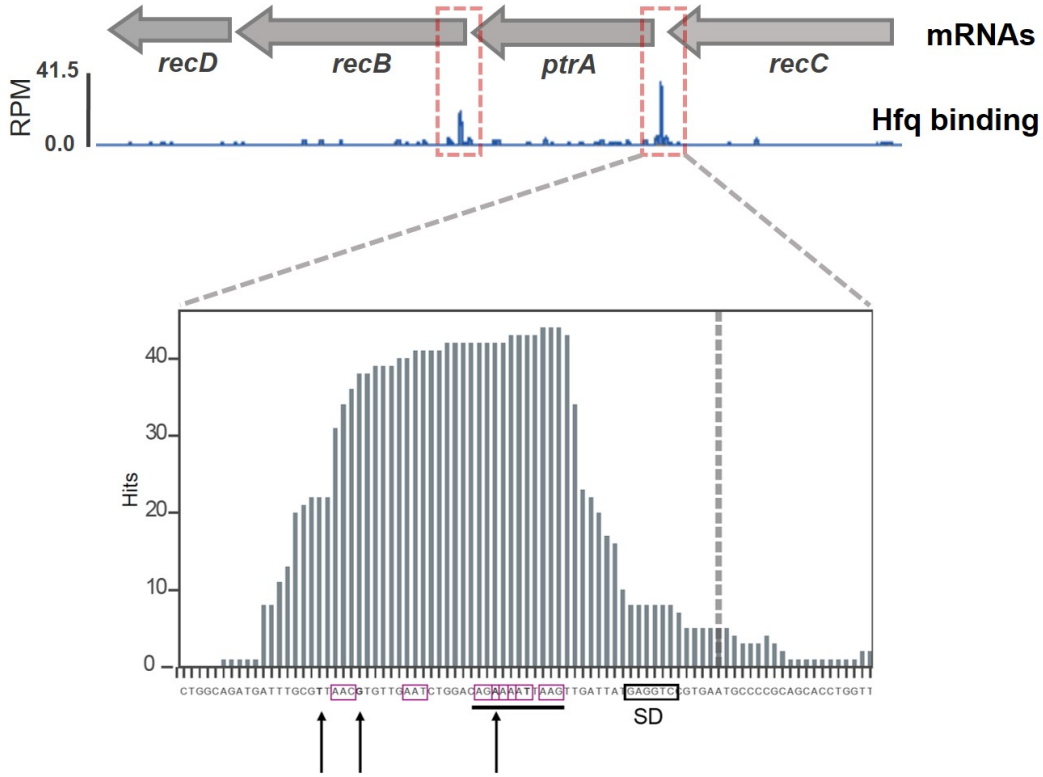

Figure 5.1: Genome browser track showing Hfq binding to *recC-ptrA-recB-recD* mRNAs identified in the Hfq-CLASH experiment in *E. coli* [100]. The coverage is normalized to reads per million (RPM). The major peaks of interest are highlighted by red dashed boxes. **Bottom:** The main binding peak magnified and presented in raw counts (Hits). The vertical dashed line and black box indicate a translation initiation site (ATG) and Shine-Dalgarno sequence (SD), respectively. The black vertical arrows show the nucleotides where direct Hfq cross-linking was detected in the CLASH data. Hfq-binding motifs, A-R(A/G)-N(any nucleotide) are highlighted in the red boxes while the cluster of Hfq binding motifs is underlined.

The Hfq binding profile across *recC-ptrA-recB-recD* transcripts showed that Hfq directly cross-links to the *ptrA-recB-recD* mRNA (Fig. 5.1). The Hfq binding mainly localized at two sites: one in close proximity to the RBS of the *ptrA* gene, and the second one further downstream in the coding sequence of *recB* gene. Interestingly, multiple trinucleotide Hfq-binding motifs, A-R(A/G)-N(any nucleotide) [125, 238] were found $\sim$8 nucleotides upstream of the RBS of the *ptrA* sequence (Fig. 5.1, zoom-in) suggesting that Hfq may control translation initiation or decay of the mRNA. Because of the potential translational coupling between *ptrA* and *recB*, this may directly impact RecB protein production. Alternatively, the other Hfq binding site might also affect *recB* translation. Taken together, these results led to the hypothesis that Hfq may control RecB expression *in vivo*.

5.2 RecB mRNA and protein quantification in *hfq* mutants

To test this hypothesis, RecB protein and mRNA levels were quantified in *E. coli* strains lacking Hfq. Because *hfq* mutants have been reported to have an increased average cell size and various growth defects [159, 160, 239], firstly, the effects of *hfq* deletion on cell size and growth rate were tested.

5.2.1 Cell size and growth rate measurements in Δ*hfq*

Consistent with previous observations, the strains lacking Hfq have larger average cell size in comparison to wild-type cells (Fig. 5.2A). While conserving the same width as in wild type, *hfq* mutants increase their average length (Fig. 5.2B). Indeed, the medians of cell area distributions, quantified using bright-field microscopy, are given as 3.2 μm^2 in Δ*hfq* and 2.8 μm^2 in wild type. Additionally, cell area distribution is wider in Δ*hfq* reflecting higher variability in cell size.

Secondly, *hfq* deletion has been shown to result in different effects on growth rate depending on nutrient availability in the medium [132, 239]. The growth rates was examined in M9-based medium supplemented with glucose and amino acids by standard populational optical density measurements OD_{600} (Fig. 5.3). No significant difference in growth rates was detected in an early/mid-exponential phase ($OD_{600} \sim$ [0.08-0.4]) (Fig. 5.3B). This agrees with similar populational measurements in minimal medium reported here [239]. A slight increase in the exponential growth rate of *hfq* mutants may be a consequence of optical density measurements for cells of different sizes [223].

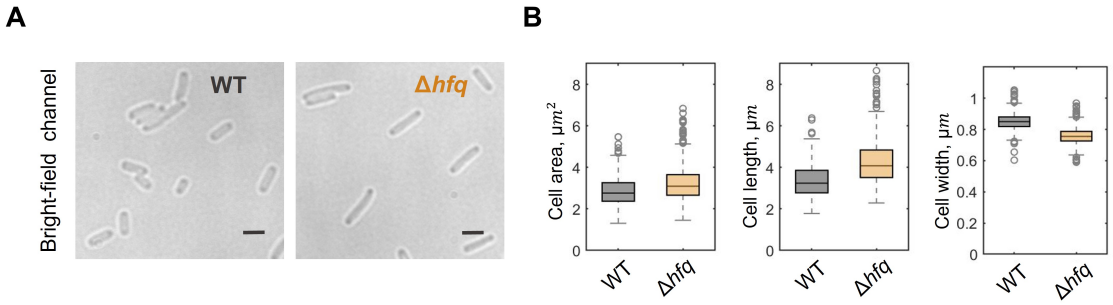

Figure 5.2: A: Examples of bright-field images for wild-type and Δ*hfq* strains. Scale bars represent 2 µm. **B:** Box plots with cell area, cell length and cell width distributions for both samples. Total number of cells included into box plots was 3000 for each sample.

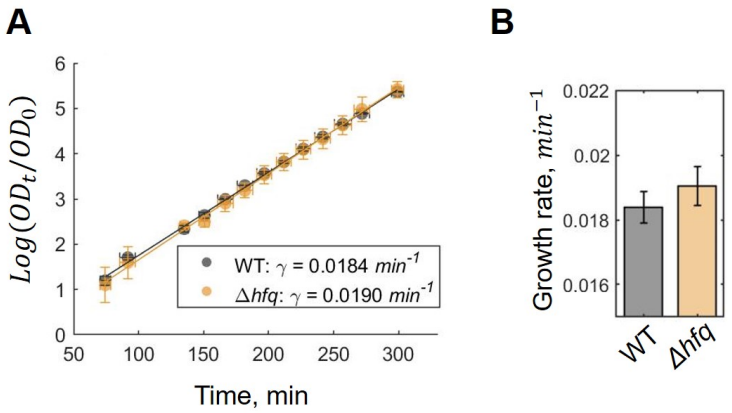

Figure 5.3: Growth curves, obtained from OD_{600} measurements and represented in logarithmic scale (**A**), and estimation of growth rates at the exponential phase (**B**) for Δ*hfq* and wild type. The growth rates, averaged across three experiments, are shown in the legend. Error bars represent 95% confidence intervals.

5.2.2 Quantification of *recB* mRNAs in Δ*hfq* by smFISH

Next, we addressed the question whether Hfq has any effect on RecBCD transcription. To this end, *recB* mRNA molecules were quantified in the Δ*hfq* strain using smFISH. Specific foci of a comparable intensity to the spots in the wild-type sample were detected in the Δ*hfq* cells (Fig. 5.4).

To account for the increased average cell size of *hfq* mutants, single-molecule RNA measurements were normalized by cell area and the data were represented in *recB* mRNA concentrations (Fig. 5.5). Although *hfq* mutants show a slight decrease in the average mRNA concentration, 0.21 mol/μm^2 (95% confidence interval is [0.08, 0.30]), compared to 0.28 mol/μm^2 (95% confidence interval is [0.23, 0.32]) in wild type, as seen from Figure 5.5 the distributions substantially overlap.

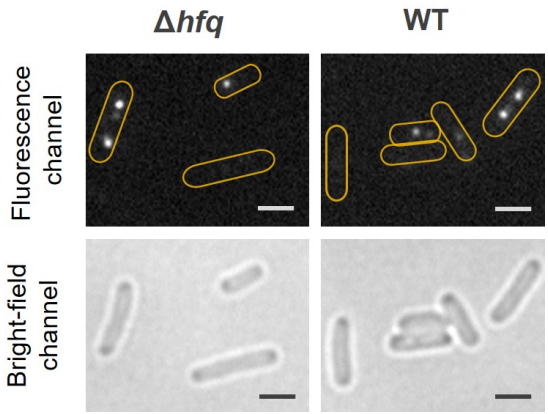

Figure 5.4: Examples of fluorescence and bright-field images of *recB* mRNA FISH experiments in Δ*hfq* and wild-type cells. Yellow outlines indicate rough positions of bacterial cells in the fluorescence channel. Scale bars represent 2 µm.

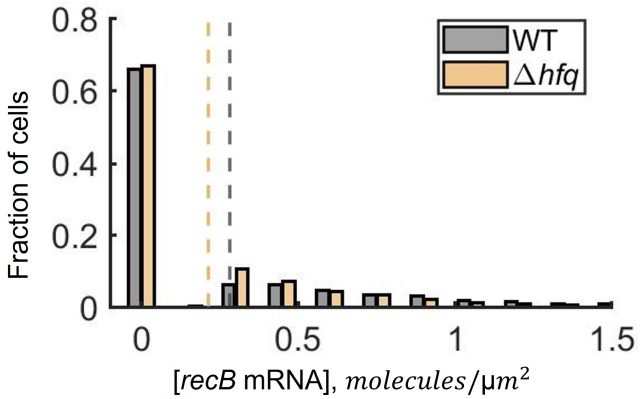

Figure 5.5: *recB* mRNA concentration distributions quantified with smFISH in *hfq* mutants (orange) and wild type (grey). The histograms represent the average of three repeated experiments. Total number of analysed cells are 7,270 (WT) and 5,486 (Δ*hfq*).

These results were confirmed by RT-qPCR measurements which did not show any significant changes for *recB* mRNA steady-state levels (Fig. 5.6). Additionally, no changes were detected for the other genes of the locus: *ptrA*, *recD*, *recC*. Taken together, it was concluded that Hfq binding to the *recBCD* mRNA does not substantially alter steady-state and stability of the *recB* transcript.

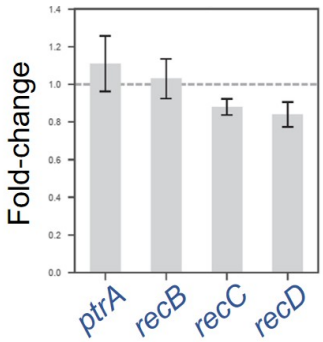

Figure 5.6: *ptrA*, *recB*, *recC* and *recD* transcripts quantified by RT-qPCR in *hfq* mutant strains and normalized to the corresponding expression levels in wild-type cells. The data represent averages and standard deviations across three replicates for each gene. *rrfD* was used as a reference gene.

5.2.3 Quantification of RecB subunits in Δ*hfq* by HaloTag-labelling

Next, RecB protein molecules were quantified in the Δ*hfq* strain with HaloTag-labelling. Diffraction-limited spots, corresponding to RecBHalo molecules labelled with JF549 dye, were detected in the Δ*hfq* mutant (Fig. 5.7).

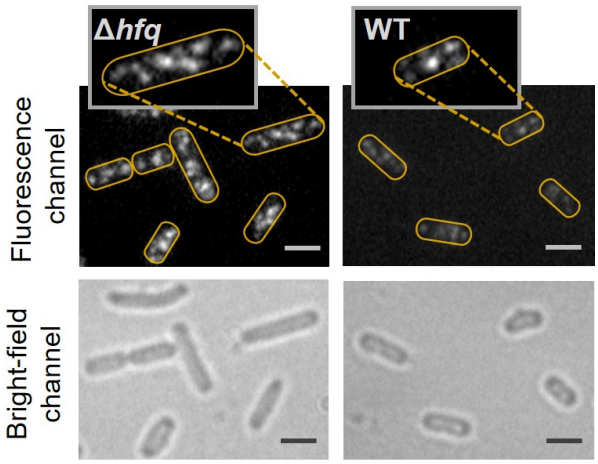

Figure 5.7: Examples of fluorescence and bright-field images of RecB HaloTag-labelling experiments in Δ*hfq* and wild type. Both fluorescence images are shown in the same intensity range while different background modulation was applied in the zoom-in figures (for better spots visualisation). Scale bars represent 2 μm.

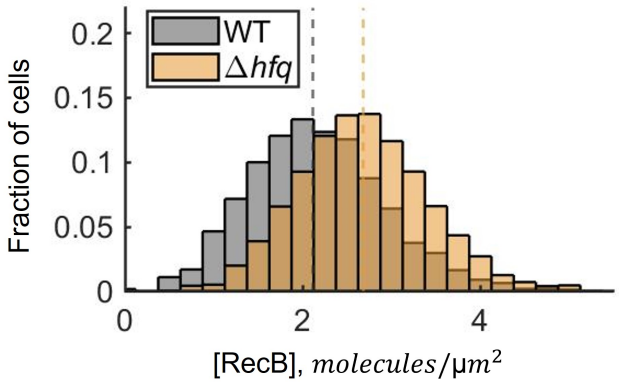

Figure 5.8: RecB concentration distributions in *hfq* mutants (orange) and wild type (grey). The histograms represent the average of three repeated experiments; the medians are shown by dashed lines. Total number of analysed cells: 3,324 (WT) and 2,185 (Δ*hfq*).

Following the same strategy as in the mRNA measurements, single-molecule data were normalized to cell area and presented in RecB protein concentration. In contrast to the transcriptional level, an increase of RecB protein concentration was observed in the Δ*hfq* cells. *hfq* mutants showed more fluorescent foci than detected in wild-type cells (Fig. 5.7). Even after normalization by cell area, RecB concentration distribution was shifted towards higher values (Fig. 5.8). Quantification across five repeated experiments showed a significant 1.3-fold increase in RecB protein concentration in the Δ*hfq* strain (Fig. 5.8, insert): 2.68 mol/μm^2 (95% confidence interval is [2.54, 2.80]) in Δ*hfq* compared to 2.12 mol/μm^2 (95% confidence interval is [2.01, 2.32]) in the control. Verification by two-sample *t*-test confirmed statistical significance (*P* value = 6.83×10^{-4}) between the average RecB concentrations in wild type and Δ*hfq* cells.

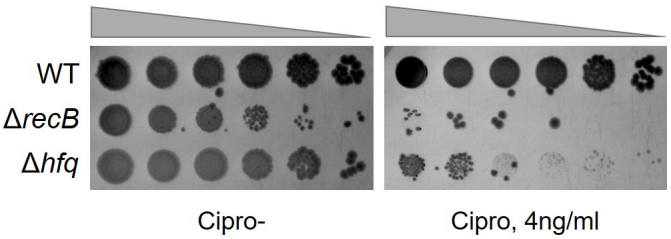

Figure 5.9: A 10-fold serial dilution assay for wild type (WT), Δ*recB* and Δ*hfq* strains. A plate was supplemented with no (left) or 4 ng/ml (right) of ciprofloxacin.

To test if the increase in RecB concentration has an impact on the DNA repair capacity, wild type and Δ*hfq* strains were exposed to sub-lethal levels of ciprofloxacin and the viability of cells was measured in a serial dilution assay (Fig. 5.9). The wild-type strain is fully viable upon exposure to 4 ng/ml of ciprofloxacin, whilst the viability of the Δ*recB* mutant is drastically reduced. Importantly, the *hfq* mutants were more sensitive to ciprofloxacin compared to the wild-type strain suggesting that the increase in RecB protein levels in the absence of Hfq may negatively impact the capacity of cells to repair DSBs.

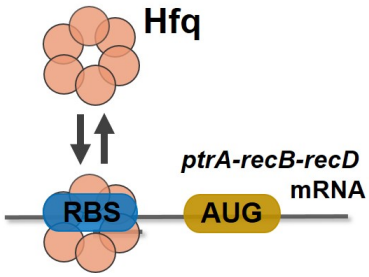

Figure 5.10: A model of Hfq down-regulating RecB translation by blocking the RBS of the *ptrA-recB-recD* transcript.

The observed results were summarized in a simple model (Fig. 5.10). Specifically, it is proposed that, a global post-transcriptional regulator, Hfq, controls gene expression of RecB subunit by binding to the polycistronic *ptrA-recB-recD* mRNA in a close proximity to the RBS of *ptrA*. The suggested model is strongly supported by the evidence of Hfq-binding detected in the CLASH experiment. The RBS contains a short sequence, known as Shine-Dalgarno, that is recognized by a ribosome, and thus, crucial for the initiation of translation [210]. Therefore, when this sequence is occupied by an RNA-binding protein, the ribosome

binding site is not accessible and translation of the mRNA can not be initiated. In other words, Hfq molecules and ribosomes are constantly competing for the same binding site within the *ptrA-recB-recD* transcripts. Thus, in the proposed model, the polycistronic transcript is less likely to be translated what will affect protein synthesis of PtrA. This will further impact the translation of the downstream protein, RecB, because of the overlap between *ptrA* and *recB* reading frames. A similar mechanism where Hfq takes part in modulating translational initiation efficiency has been uncovered in recent experimental studies [45, 65].

5.3 Hfq complementation experiment

In order to test the proposed mechanism further, Hfq expression level was altered *in vivo*. As discussed in Chapter 1, in native conditions cellular Hfq concentration is tightly autoregulated by a post-transcriptional negative feedback mechanism [157, 246]. This strict control, based on the repression of *hfq* mRNA translation by Hfq binding, makes it challenging to modulate Hfq levels in native conditions. Thus, Hfq protein was expressed from a multi-copy plasmid pQE-Hfq in the Δ*hfq* strain [157]. Using a phenotypical assay it was first confirmed that the expression plasmid pQE-Hfq rescues *hfq* deletion phenotype (Fig. 5.11).

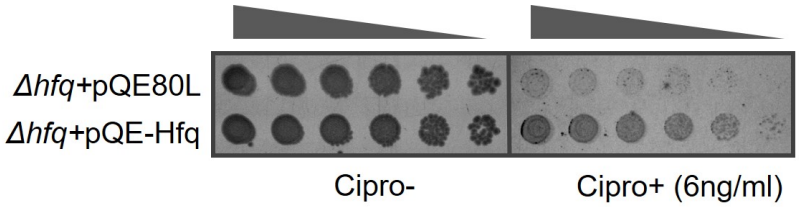

Figure 5.11: A 5-fold serial dilution assay of Δ*hfq* strain carrying pQE80L or pQE-Hfq plasmid. Cells were plated onto LB plates without or with 6 ng/ml of ciprofloxacin.

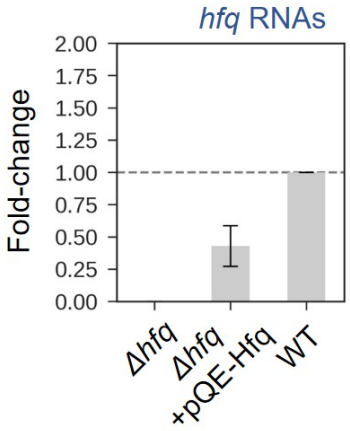

Figure 5.12: *hfq* RNA levels quantified by RT-qPCR in Δ*hfq* cells carrying pQE-Hfq plasmid and normalized to the *hfq* RNA level in wild type. The data represent averages and standard deviations across three replicates. *rrfD* was used as a reference gene.

Next, the following question was asked: to what extent pQE-Hfq expresses *hfq* in our conditions? By quantifying *hfq* mRNA levels by RT-qPCR in Δ*hfq* strain carrying pQE-Hfq, it has been shown that a third of *hfq* transcription level was achieved in these conditions (Fig. 5.12).

Further, it was tested whether the expression from the pQE-Hfq plasmid or the backbone plasmid pQE80L has any effect on the *recBCD* transcription. *ptrA*, *recB*, *recC* and *recD* levels were quantified by RT-qPCR in Δ*hfq* carrying pQE80L or pQE-Hfq plasmid and normalized to the corresponding expression in Δ*hfq* and Δ*hfq*+pQE80L, respectively (Fig. 5.13). No significant changes were detected at the transcriptional steady-state levels. In other words, it was demonstrated that harboring pQE80L or pQE-Hfq plasmid does not alter *recB* transcription.

We next proceeded with single-cell protein quantification. It is worth noting that the strain harboring the backbone plasmid pQE80L did not show any effect on RecB expression (Fig. 5.14, in blue). In contrast, it was found that in the sample, where Hfq production was complemented by expression from pQE-Hfq plasmid, RecB protein concentration distribution was

shifted toward the lower values compared to the RecB histogram in the *hfq* mutants. Statistical significance was evaluated with a two-sample *t*-test and resulted in the following *P* values: 0.045(*) for Δ*hfq* and Δ*hfq*+pQE-Hfq; 0.71(ns) for Δ*hfq* and Δ*hfq*+pQE80L. Interestingly, RecB protein level was not fully recovered to the wild-type level. This is consistent with partial recovery of *hfq* mRNA level measured by RT-qPCR (∼ one third of the *hfq* mRNA level in wild type, Fig. 5.12).

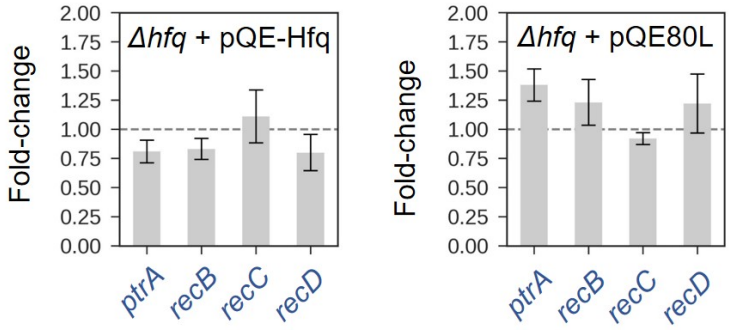

Figure 5.13: *ptrA*, *recB*, *recC* and *recD* transcripts quantified by RT-qPCR in Δ*hfq* carrying pQE-Hfq (left) or pQE80L (right) plasmid and normalized to the corresponding expression levels in Δ*hfq* and Δ*hfq*+pQE80L, respectively. The data represent averages and standard deviations across two replicates. *rrfD* was used as a reference gene.

According to the proposed model of RecB regulation, expression of Hfq in the *hfq* mutant leads to the binding of Hfq molecules to the *ptrA-recB* RBS which is expected to result in less effective RecB translation, and hence, a decrease in RecB protein concentration. Thus, this observation is in full agreement with the mechanism of RecB regulation suggested above (Fig. 5.10).

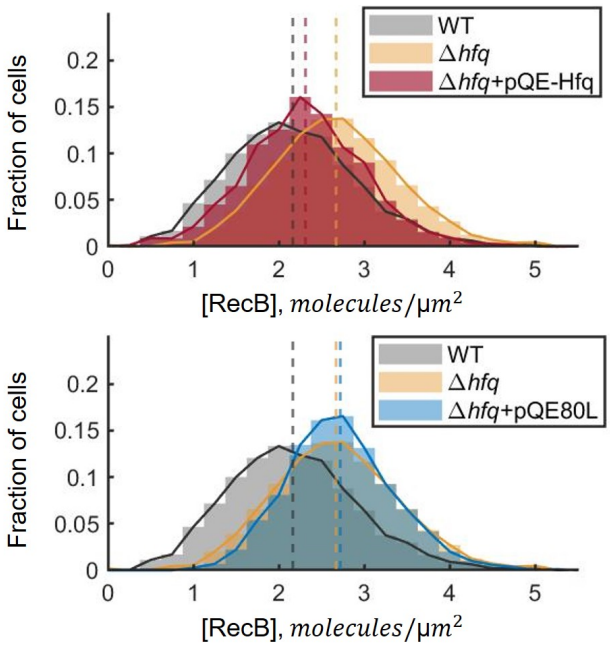

Figure 5.14: RecB protein concentration distributions quantified in wild type, Δ*hfq* and Δ*hfq* carrying pQE-Hfq or pQE80L plasmid. The histograms represent the average across two repeated experiments per each condition. The dashed lines represent mean RecB concentration for each sample. Total number of analysed cells are 2,221 (Δ*hfq*+pQE-Hfq) and 1,722 (Δ*hfq*+pQE80L).

5.4 Titration of Hfq by a small RNA ChiX

To further ascertain the role of Hfq in controlling RecB expression whilst avoiding potential indirect effects due to the pleiotropic phenotype of the *hfq* mutants, an experiment where Hfq proteins were sequestered from its *recB* mRNA targets was designed. Hfq is known to bind ∼100 sRNAs [97, 238], some of which can reach high expression levels and/or have higher affinity for Hfq, and thus, compete for Hfq binding [70, 155]. It is reasoned that because *recB* mRNA is present in a very low abundance, it could be out-competed for Hfq binding by over-expression of a small RNA that efficiently binds Hfq, such as ChiX [137, 155].

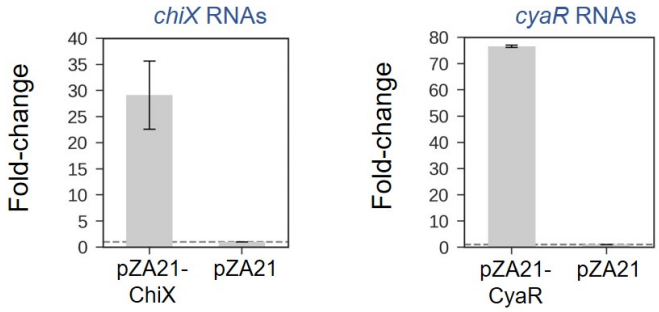

Figure 5.15: *chiX* (left) and *cyaR* (right) RNA levels quantified by RT-qPCR in wild type carrying pZA-ChiX or pZA-CyaR over-expression plasmid. The results were normalized to the corresponding expression in the cells carrying the backbone plasmid (pZA21). The data represent averages and standard deviations across three replicates for each gene. *rrfD* was used as a reference gene.

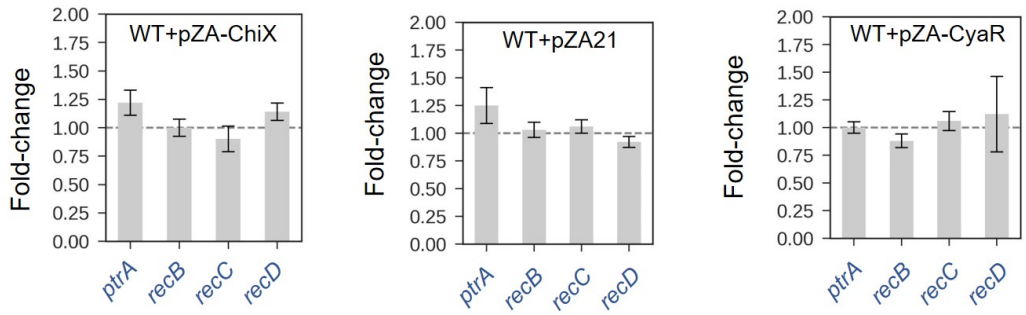

Figure 5.16: *ptrA*, *recB*, *recC* and *recD* transcripts quantified by RT-qPCR in wild type carrying pZA-ChiX, pZA21 or pZA-CyaR plasmid. RNA levels were normalized to the corresponding expression levels in wild type. The data represent averages and standard deviations across three replicates for each gene. *rrfD* was used as a reference gene.

To test this, a multi-copy plasmid pZA21-ChiX was constructed to over-express ChiX and a ∼30-fold increase was achieved in ChiX expression relative to its native expression in wild-type cells (Fig. 5.15). Given its already high abundance in wild-type cells, such overproduction is very likely to sequester a large number of Hfq molecules [65, 134, 224]. Additionally, a control with another small RNA was carried out. A sRNA CyaR, that does not display as high affinity as ChiX, and thus, is not expected to titrate Hfq molecules, was chosen. In these conditions, the CyaR over-expression led to a ∼75-fold higher *cyaR* level compared to the one in wild type (Fig. 5.15).

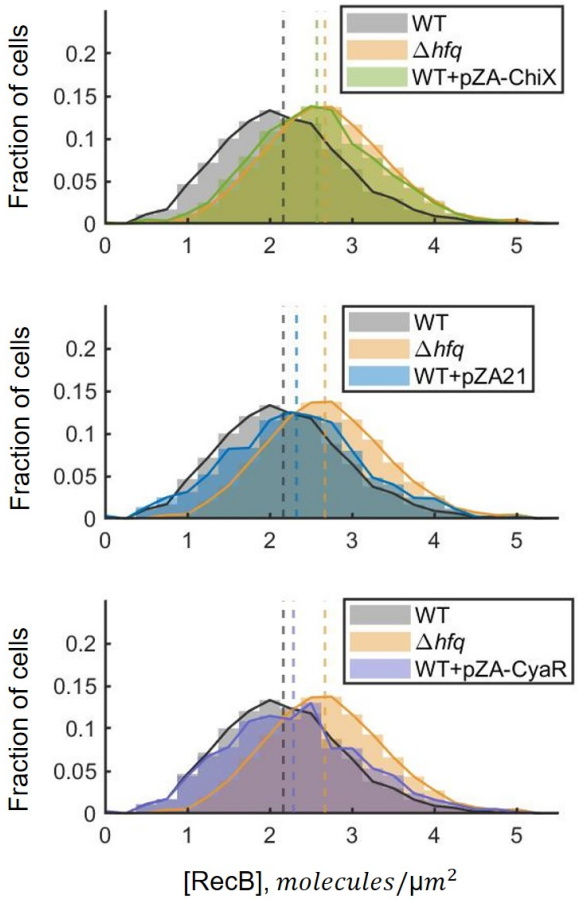

Figure 5.17: RecB protein concentration distributions quantified in wild type, Δ*hfq*, wild type carrying pZA-ChiX, pZA21 or pZA-CyaR plasmid. The histograms represent the average across two repeated experiments per each condition. The dashed lines represent the mean RecB concentration in each condition. Total number of analysed cells are 2,558 (WT+pZA-ChiX); 985 (WT+pZA21); 2,918 (WT+pZA-CyaR).

Further, we asked whether harboring any of the overproduction plasmids, pZA-ChiX or pZA-CyaR, or a backbone plasmid pZA21, impacts *recB* transcriptional level. To investigate this, *ptrA*, *recB*, *recC* and *recD* mRNAs were quantified by RT-qPCR (Fig. 5.16). No significant changes were detected at the *ptrA*, *recB*, *recD*, or *recC* transcript levels (i) when ChiX or CyaR was overproduced or (ii) when cells carried the backbone plasmid.

Next, the effect of ChiX overproduction on RecB protein production was tested. First, it was ensured that neither the backbone plasmid nor CyaR overproduction changed RecB protein steady-state distribution (Fig. 5.17). Statistical significance was evaluated with two-sample *t*-test (*P* values: 0.42(ns) for WT and WT+pZA21; 0.37(ns) for WT and WT+pZA-CyaR). In contrast, an increase in RecB protein production was detected when ChiX was over-expressed (Fig. 5.17, *P* value for WT and WT+pZA-ChiX, evaluated by two-sample *t*-test, is 0.021(*)). Surprisingly, RecB concentration in the cells where ChiX was over-expressed almost fully overlapped with the quantification in *hfq* mutants. This suggests that very few Hfq molecules are available when ChiX is overproduced to such extent ($\sim$30-fold) and that ChiX over-expression specifically impairs RecB post-transcriptional regulation.

To further test whether ChiX over-expression has an effect on the efficiency of DNA repair, DSBs were induced with 8–10 ng/ml of ciprofloxacin in a serial dilution assay (Fig. 5.18, upper panel). While ChiX overproduction does not affect cell size as much as the *hfq* deletion, it shows a phenotype upon DNA DSB damage. That may suggest that the increased RecB protein production decreases the ability to faithfully perform DNA repair. Interestingly, a striking similarity in decrease in DNA repair efficiency was observed when RecBCD is over-expressed from a plasmid encoding the entire *recBCD* locus (Fig. 5.18, lower panel). This observation supports the idea that increased number of RecBCD molecules is toxic when cells are exposed to DNA damage.

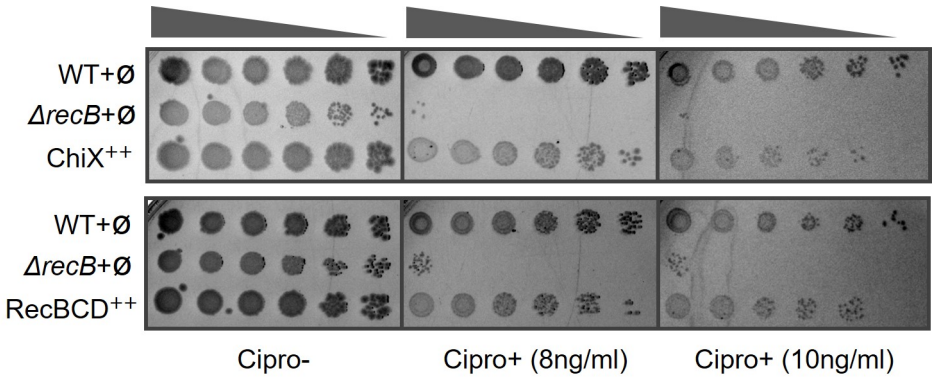

Figure 5.18: A 5-fold serial dilution assay for wild type, Δ*recB* and wild type carrying ChiX expression plasmid (ChiX^{++}) (top) or RecBCD expression plasmid (RecBCD^{++}) (bottom). A plate was supplemented with no (left), 8 ng/ml (middle) or 10 ng/ml (right) of ciprofloxacin.

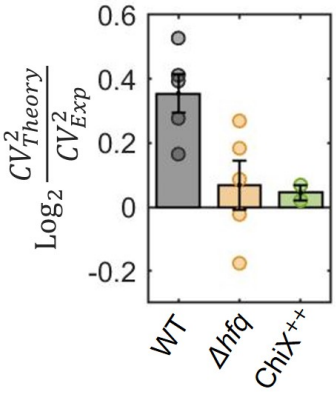

Figure 5.19: The ratio between theoretically predicted (CV^2_{Theory}) and experimentally measured (CV^2_{Exp}) coefficient of variation (squared) presented in logarithmic scale. The expected values were computed via the average protein copy number $\langle p \rangle$ as $1/\langle p \rangle$ for WT, Δ*hfq* and ChiX^{++} samples. Error bars represent s.e.m. calculated from the replicated experiments.

Finally, to test whether Hfq regulation contributes to the noise suppression detected in the wild-type cells (Chapter 3), the levels of fluctuations of RecB copy numbers were quantified in *hfq* mutants and under ChiX over-expression. To this end, the ratio between analytically predicted noise (CV^2_{Theory}) and experimentally measured (CV^2_{Exp}) has been computed. Theoretical values were calculated as $1/\langle p \rangle$, where $\langle p \rangle$ is the average RecB protein copy number in each condition. Here, only the first right-hand-side term of the equation (Eq. 3.3) was used to estimate protein noise as no significant change was detected in *recB* mRNA levels in the *hfq* mutants or ChiX^{++} compared to the wild type. Thus, it was assumed that mRNA contribution (the second right-hand-side term of Eq. 3.3) is equal for all three conditions and only the difference given by the change in average protein abundance $\langle p \rangle$ was considered. A less effective noise reduction was observed in Δ*hfq* and ChiX^{++} in comparison to wild type (Fig. 5.19) suggesting that post-transcriptional regulation by Hfq decreases fluctuations in RecB protein numbers.

5.5 Testing Hfq binding sites: proximal, distal, rim

While sRNA-independent regulation by Hfq is a re-emerging field, in many reported cases a small RNA is required for the specificity of an interaction to a target mRNA. Thus, in this section, an attempt to investigate whether any small RNAs are directly involved in the regulation of RecB expression is described. To this end, an interaction between Hfq and other potential players was aimed to be disrupted by expressing Hfq proteins (from a plasmid) that were point-mutated in one of its binding surfaces: proximal pQE-Hfq(K56A), distal pQE-Hfq(Y26D), or rim pQE-Hfq(R16A). These binding sites were shown to impact the interaction between Hfq and different RNA classes [265].

By using RT-qPCR, *hfq* transcription from each of the plasmids was measured. As seen from Figure 5.20, *hfq* RNAs were transcribed from all the plasmids. The level of expression from pQE-Hfq(Y26D) and pQE-Hfq(K56A) were close to the expression of a fully functional Hfq (pQE-Hfq), which gives approximately one third of the Hfq level in wild type (Fig. 5.12). Expression from pQE-Hfq(R16A) was found to be ∼50% lower what roughly results in one sixth of the *hfq* transcript steady state in wild-type cells.

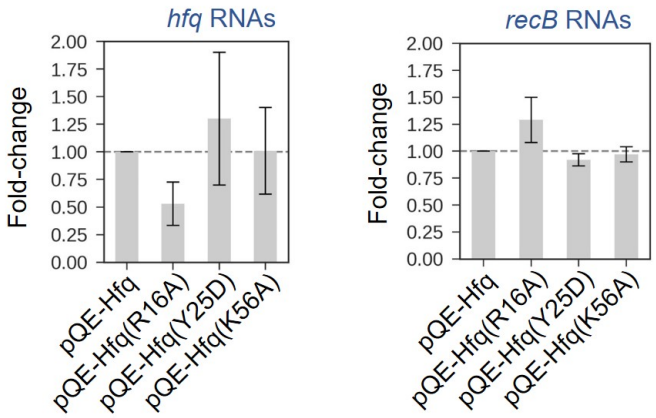

Figure 5.20: *hfq* (left) and *recB* (right) RNA levels quantified by RT-qPCR in Δ*hfq* carrying a pQE80L-derivative plasmid with a mutated Hfq protein. The results were normalized to the corresponding expression in Δ*hfq* cells carrying the backbone plasmid, pQE-Hfq. The data represent averages and standard deviations across two replicates for each gene. *rrfD* was used as a reference gene.

Next, RecB mRNAs and proteins were quantified upon expression of Hfq mutants. When expressing an Hfq mutant from a plasmid in Δ*hfq* cells, *recB* mRNA steady-state levels remained unchanged (Fig. 5.20, right plot). An intermediate effect on RecB protein distributions was detected in all the cases (Fig. 5.21). Statistical significance was evaluated with two-sample *t*-test and resulted in the following *P* values: 0.19(ns) for Δ*hfq* and Δ*hfq*+pQE-Hfq(R16A); 0.07(ns) for Δ*hfq* and Δ*hfq*+pQE-Hfq(Y25D); 0.30(ns) for Δ*hfq* and Δ*hfq*+pQE-Hfq(K56A).

It is worth noting that RecB protein distribution quantified in Δ*hfq* cells harboring a mutated Hfq version on a plasmid, fell between the extreme cases: Δ*hfq* and wild type. However, further interpretation is complicated by the fact that the difference between the extreme cases of RecB concentrations (in Δ*hfq* and wild type) is fairly small ($\sim$1.3-fold) what makes it challenging to confidently detect any changes within this range. Thus, while these results provide an additional confirmation of Hfq involvement in the regulation of RecB translation, they do not reject the involvement of a small RNA in the regulation mechanism. In fact, the similarity of RecB protein expression in all three cases to the RecB distribution in Δ*hfq* cells may suggest sRNA candidates which utilize more than one Hfq site for the interaction with the *ptrA-recB* transcript.

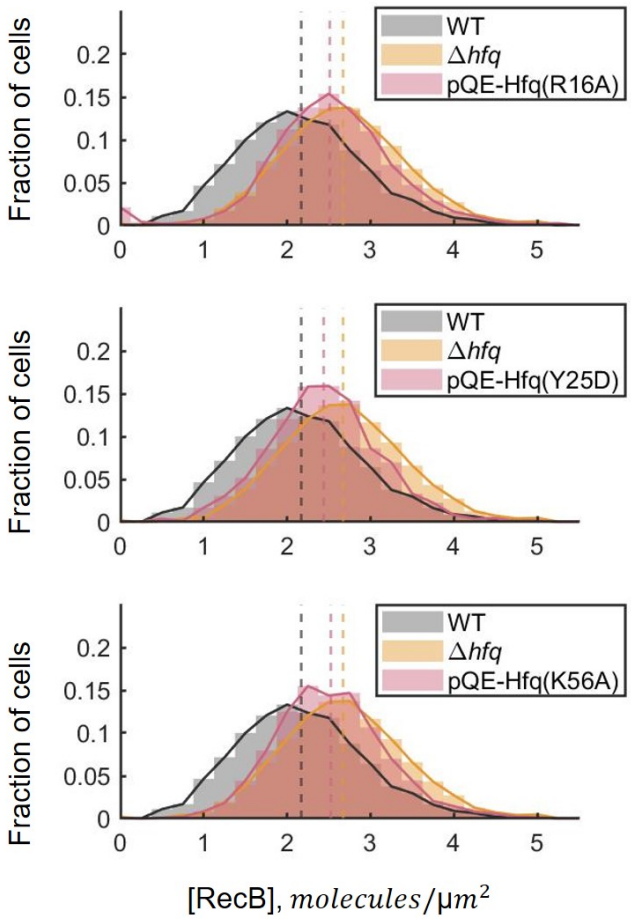

Figure 5.21: RecB protein concentration distributions quantified in Δ*hfq* carrying a pQE80L-derivative plasmid with a mutated Hfq protein. Point mutations were made in one of the Hfq binding sites: rim (R16A), distal (Y25D) or proximal (K56A) [265]. The histograms represent the average across two repeated experiments for each mutation. The RecB concentration histograms for wild type and *hfq* mutants are plotted as references in grey and orange, respectively. The dashed lines represent the mean RecB concentration in each condition. Total number of analysed cells are 1,630 for pQE-Hfq(R16A); 1,318 for pQE-Hfq(Y25D); 1,673 for pQE-Hfq(K56A).

5.6 Small RNA candidates: CyaR, ChiX, GcvB and ArcZ

Because no clear indication of a sRNA type was found with point mutations in Hfq, we attempted to investigate potential small RNAs with a different approach. To look for a sRNA candidates, we analysed an available RIL-seq (RIL) dataset where all RNA-RNA interactions were detected in *E. coli* [142]. As previously, the analysis was specifically focused on the *recBCD* locus. Several small RNAs were detected to bind within the *ptrA-recB-recD* mRNA: ArcZ, CyaR, GcvB, and ChiX. A potential role of these small RNAs in regulating RecB expression was further tested by quantifying RecB mRNA and protein levels in sRNA mutants and over-expression experiments. No evidence of a direct role of these sRNAs in the RecB regulation mechanism was found.

5.6.1 Identification of sRNA candidates

Alongside with large-scale approaches (such as CLASH and CLIP) to study RNA-binding protein interactions, RIL-seq is the transcriptome-wide method to detect *in vivo* RBP-bound pairs of small RNAs with their targets. As described in detail in Chapter 1, the RIL-seq methodology is based on *in vivo* protein-RNA crosslinking, purification of the RBP of interest and bound RNAs, RNA ligation, sequencing, and finally mapping of the sequenced fragments to the bacterial genome. This procedure results in identification of RNA-RNA interactions known as chimeric reads.

In the published RIL-seq dataset [142], Hfq protein was utilized as an RBP of interest and Hfq-bound sRNA-target pairs were detected at the genome-wide scale in *E. coli*. The data were analysed specifically for the *recC-ptrA-recB-recD* chromosomal region. The binding profiles for all RNAs (Hfq RIL-seq) as well as the tracks for each particular sRNA (ArcZ, CyaR, GcvB, ChiX) are shown in Figure 5.22. In addition, these sRNA binding were aligned with the Hfq-binding from Hfq-CLASH experiment (Hfq binding), discussed above (Fig. 5.1).

Interestingly, the binding peaks between CyaR, GcvB, or ChiX and *recC-ptrA-recB-recD* mRNAs were found to be mainly localized at the two Hfq-binding sites, identified in the Hfq-CLASH data. Similarly to Hfq interactions, the major sRNA binding peak is the one which is in close proximity to the RBS of *ptrA-recB*. It is worth noting that GcvB and ChiX reads were also detected at the position of the other Hfq-binding which is close to the beginning of the *recB* coding sequence.

Additionally, direct cross-linking between sRNA ArcZ and *ptrA-recB-recD* mRNA was found in the RIL-seq and Hfq-CLASH datasets [100, 142, 143]. This binding is localized further downstream of the *recB* reading frame. Although (i) the amount of detected reads significantly exceeds the background signal and (ii) it was detected in three datasets, the binding of ArcZ does not overlap with Hfq-binding sites within *ptrA-recB* (Fig. 5.22).

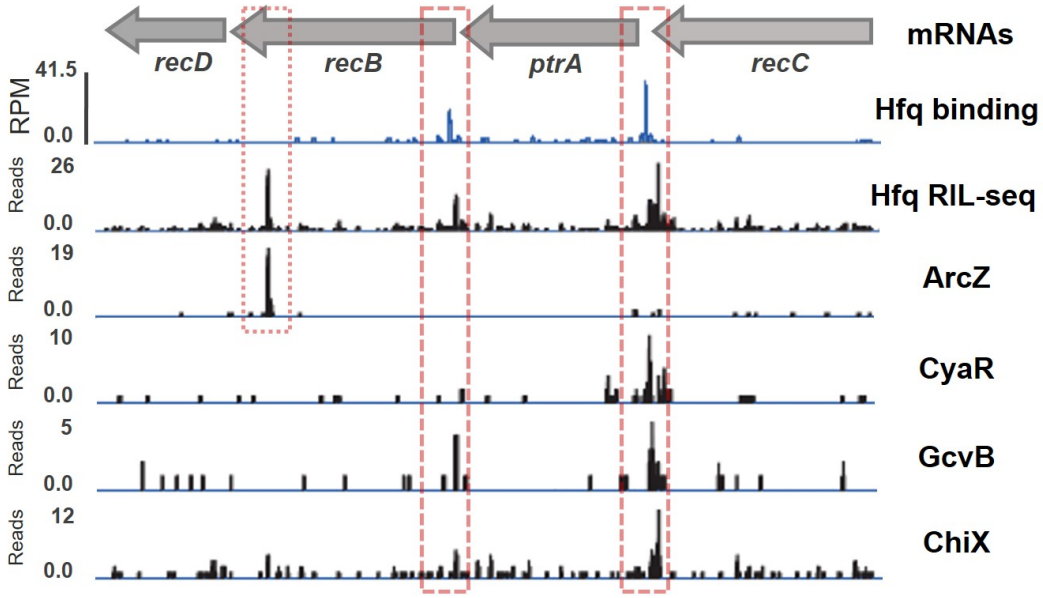

Figure 5.22: RNA binding within *recC-ptrA-recB-recD* transcripts identified in the RIL-seq experiment in *E. coli* [142]. sRNA binding coverage is shown in raw counts (Reads). Hfq binding detected in Hfq-CLASH experiment is shown in blue (the same as in Figure 5.1). The main binding sites are highlighted by red dashed boxes.

5.6.2 sRNA deletion and over-expression experiments

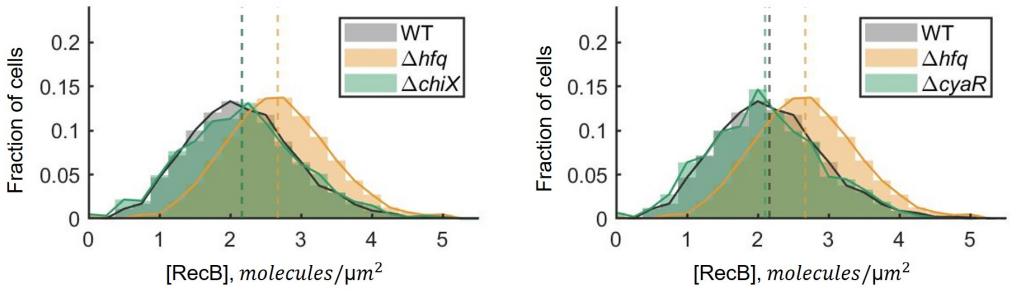

Figure 5.23: RecB protein quantification in Δ*chiX* (left) and Δ*cyaR* (right). The histograms represent the average across two repeated experiments. The RecB concentration histograms for wild type and Δ*hfq* mutants are shown for comparison. The dashed lines represent the mean RecB concentration in each condition. Total number of analysed cells are 2,238 (Δ*chiX*); 2,286 (Δ*cyaR*).

In order to test the role of small RNA candidates identified in RIL-seq dataset, RecB expression was characterized in sRNA deletion and over-expression studies. The *chiX*, *cyaR* and *arcZ* mutant strains with RecB-HaloTag fusion were constructed. Then, HaloTag-labelling quantification of RecB subunits was performed in these strains. Unfortunately, the *gcvB* deletion was not built however its potential role was addressed by an over-expression experiment (see below). As seen from the results for Δ*chiX*, Δ*cyaR* and Δ*arcZ* (Fig. 5.23 and 5.24), no significant changes were detected in RecB concentration distributions. Indeed, statistical significance was evaluated with two-sample *t*-test and resulted in the following *P* values: 0.96(ns) for WT and Δ*chiX*; 0.61(ns) for WT and Δ*cyaR*. Using RT-qPCR approach, we confirmed that *recC-ptrA-recB-recD* mRNA levels were not affected by sRNA deletions (Fig. 5.25).

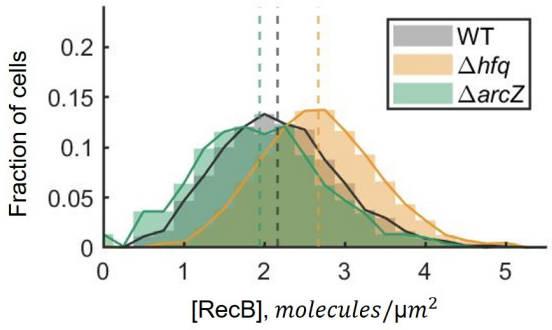

Figure 5.24: RecB protein quantification in Δ*arcZ*. The histogram represents the result of one experiment. Total number of analysed cells is 2,379 for Δ*arcZ*.

Because no changes of RecB mRNA or protein levels were detected in Δ*chiX* strain relative to wild-type control, it was concluded that sRNA ChiX is not directly involved in the regulation of RecB expression. In other words, these observation with the results of ChiX overproduction experiment (Fig. 5.17) aligns with the hypothesis of indirect affect on the RecB translation. Consistent with the suggested mechanism of RecB regulation (Fig. 5.10), titration of Hfq proteins from a shared pool by ChiX sRNAs leads to the increased availability of *ptrA-recB* RBS for ribosomes and, consequently, results in a more efficient RecB translation (Fig. 5.17).

Finally, RecB quantification was performed in the conditions where a small RNA candidate was constitutively over-expressed from a pZA21-derivative plasmid. The results for ChiX, CyaR and the control with a backbone plasmid were shown and discussed above (Fig. 5.17). Here, preliminary experiments with GcvB and ArcZ over-expression are presented (Fig. 5.26). While the average RecB protein concentration was found to be similar to the one in wild-type cells, the RecB distribution seems to be noisy when GcvB is overproduced. In addition, GcvB over-expression led to a substantial decrease in growth rate (data not shown). However, we did not achieve reliable statistics in this particular experiment. Thus, it was preliminary concluded that GcvB and ArcZ do not seem to have a direct regulatory role in RecB expression.

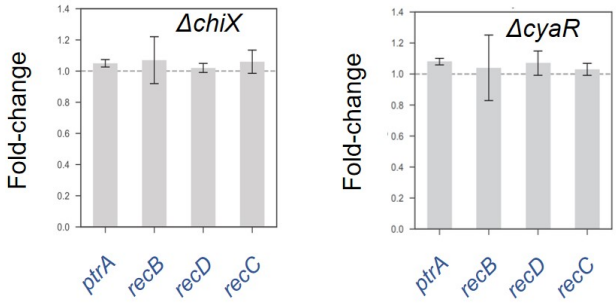

Figure 5.25: *ptrA*, *recB*, *recD* and *recC* transcripts quantified by RT-qPCR in Δ*chiX* (left) and Δ*cyaR* (right) and normalized to the corresponding expression levels in wild type. The data represent averages and standard deviations across three replicates for each gene. *rrfD* was used as a reference gene.

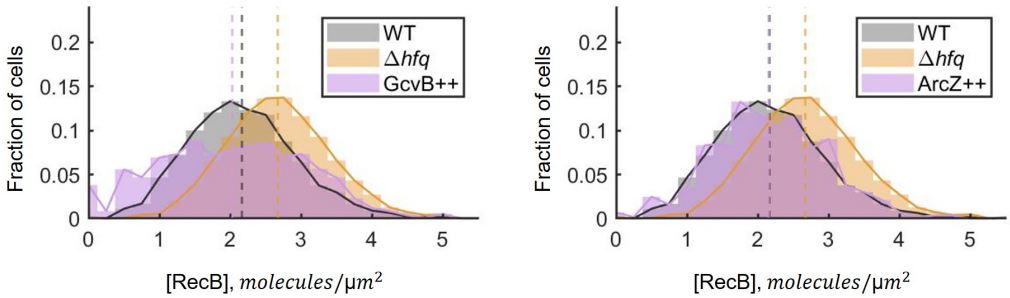

Figure 5.26: RecB protein concentration distributions quantified in wild type carrying pZA-GcvB (left) or pZA-ArcZ (right) plasmid. The histograms represent the average across two repeated experiments for GcvB and one experiment for ArcZ over-expression. The dashed lines represent the mean RecB concentration in each condition. Total number of analysed cells are 539 (GcvB^{++}) and 942 (ArcZ^{++}).

Taken together, apart from already discussed ChiX, that played a role of an Hfq-titrator, no significant changes in RecB protein or *recB* mRNA levels were detected when a potential sRNA candidate was overproduced or deleted. Based on these results, it is not possible to entirely exclude the possibility of a small RNA being involved into the network, however, it should be acknowledged that identification of one (if there are any) is a difficult task.

5.7 Discussion

Several evidence of potential regulation of RecB expression were provided in the previous chapters. Here, based on the interaction between a global regulator Hfq and the *recB* mRNA, a post-transcriptional regulation mechanism of RecB expression has been proposed. By quantifying mRNA and protein levels in the Δ*hfq* cells, Hfq was shown to be involved in the regulation of RecB expression *in vivo*. RecB subunits were found to accumulate in the absence of Hfq while the *recB* mRNA concentration was shown to be Hfq-independent. It was argued that the detected increase in RecB protein level cannot be explained by enlarged cell size or slower dilution in *hfq* mutants.

Because of the pleiotropic phenotype of Δ*hfq*, a different strategy has been utilized to further verify the proposed regulation mechanism. A sRNA ChiX was over-expressed to sequester Hfq molecules from other targets which bind Hfq with lower affinity. While ChiX over-expression system is still unspecific to the *recB* target, bacterial cells display less pronounced phenotypic changes in comparison to Δ*hfq*. This allows to consider ChiX overproduction as a less perturbed system.

While the experiments with point-mutated Hfq proteins are lacking a clear interpretation, the expectations are still worth discussing. As reported in the literature, Hfq is believed to mainly use the distal face for binding mRNA targets [125]. For instance, Hfq binds its own RNA through the distal surface. Therefore, it may be expected that expressing Hfq mutated in the distal site Hfq(Y25D) should disrupt the interaction with *recB* mRNA. Indeed, RecB distribution does not overlay with the one in wild type suggesting that the distal site is utilized in the regulation (Fig. 5.21, middle plot). On the other hand, if the distal site was the only one used in the interaction, it would be expected that the other two mutations (in rim and proximal surfaces) should have the same RecB protein distributions as in wild type (Fig. 5.21, upper and lower plots). This is not the case: all sites (rim, distal and proximal) seem to be utilized in some way. This may suggest direct or indirect involvement of some additional players facilitating the interaction.

To speculate further, it may be informative to recall the potential sRNA candidates identified in the RIL-seq dataset (Fig. 5.22, [142]): ArcZ, CyaR, GcvB and ChiX. In order to estimate the abundance of these sRNAs in native conditions, RT-qPCR quantification has been performed and compared to the *recB* mRNA level (Fig. 5.27). As seen from the figure, the small RNAs were found to be present in a wide range from less than one molecule per cell to $\sim$ 80% of the expression level of a ribosomal gene. Indeed, ChiX and GcvB are highly expressed (Fig. 5.27, right plot): ChiX is produced at half of the *rrfD* RNA level while GcvB is even more abundant ($\sim$ 80% of *rrfD* transcripts). Interestingly, transcription levels of ArcZ and CyaR were found to be comparable to the *recB* mRNA abundance (Fig. 5.27, left plot). Although the data are preliminary, this emphasizes the strength of detected interactions of *ptrA-recB* transcript with CyaR and ArcZ.

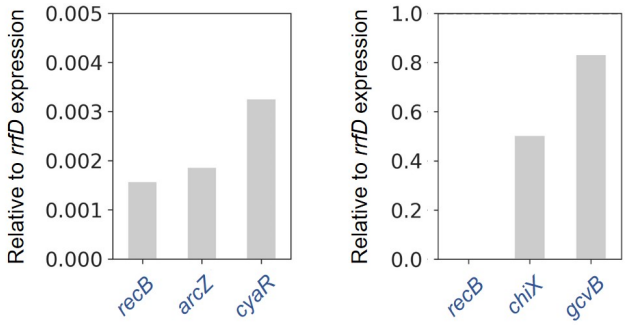

Figure 5.27: *recB*, *arcZ*, *cyaR*, *chiX* and *gcvB* RNA abundance relative to the level of a highly-expressed ribosomal gene. RNA levels were quantified by RT-qPCR in wild-type strain and normalized to the expression level of *rrfD* transcripts. The data represent the results of only one experiment. The plots are presented in different scales.

Finally, Hfq-facilitated regulation of RecB expression does not preclude that another RNA-binding protein may be involved. Although the difficulty remains the same: due to the low abundance of the *recBCD* targets, the interactions (if detected) can fall below a statistically significant threshold in a high-throughput screening. For instance, no reads were detected within the *recC-ptrA-recB-recD* transcripts in the available CLIP-seq experiment where a post-transcriptional regulator ProQ was used as a bait protein in *Escherichia coli* [96]. Similarly, no evidence of binding to the *recC-ptrA-recB-recD* mRNAs was found in a genome-wide CLIP-seq dataset with CsrA in *Salmonella* Typhimurium [97]. The interactions between ProQ and *recBCD* mRNAs were identified in a recent RIL-seq data in *E. coli* however they were not considered as statistically significant, and consequently, genuine [143].

Chapter 6

Closing remarks

RecBCD enzyme performs a key role of detecting dsDNA breaks and initiating the repair pathway in bacteria *Escherichia coli*. Besides the essential role in DSB repair, over-expression of RecBCD decreases cell viability upon DNA damage induction (Fig. 5.18, [55]). Such a dual impact on cell fitness hints at the existence of an underlying regulation that tightly controls RecBCD copy numbers within an optimal range.

Here, with single-molecule mRNA and protein quantification it was shown that, despite the large cell-to-cell variability in *recB* mRNA abundance, RecB protein levels remain remarkably constant. The propagation of noise from mRNA to protein level can be attenuated by a post-transcriptional regulation mechanism [226], for instance, by a tight control of the *recB* mRNA translation. In this work, it was shown that a global regulator Hfq binds to the *recB* mRNAs and a post-transcriptional down-regulation of RecB expression by Hfq was further established *in vivo*. Additionally, a less effective noise reduction was detected in the absence of Hfq providing the first experimental evidence of an RNA binding protein involved in noise suppression in bacterial cells. As a result of high abundance of Hfq molecules and low copy numbers of the target *recB* mRNAs, the effect of buffering mRNA fluctuations is likely to be robust for any changes in *recB* mRNA levels upon stress conditions. Finally, a reduced capacity to repair DNA was demonstrated when the Hfq-mediated regulation was perturbed. It was thus speculated that the inhibition of RecB translation by Hfq protects bacteria against the toxic consequences of high RecBCD molecule numbers.

Toxicity of RecBCD over-expression

The necessity for limitation of RecBCD production is consistent with previously reported results where DNA damage was induced with UV and gamma irradiation [55]. It is worth noting that another study did not show toxicity of RecBCD overproduction upon DNA damage [61], however this is likely because the level of UV irradiation was too low to detect a significant reduction in cell viability. The toxicity of RecBCD over-expression was only observed at relatively high level of ciprofloxacin exposure (10 ng/ml with the minimum inhibitory concentration

being at 16 ng/ml) (Fig. 5.18). Notably, no toxic effect was detected when a single subunit (RecB or RecC) was over-expressed (Appendix 8.1). These observations suggest that toxicity of RecBCD overproduction is detected when the enzyme is expressed from the originally encoded locus.

Molecular mechanism: details and challenges

The analysis of Hfq-CLASH data showed direct Hfq cross-linking with the *ptrA-recB-recD* mRNA. Further investigation of the exact Hfq binding site by mutational analysis is complicated by the extremely close location to the RBS site of the target mRNA ($\sim$ 8 nt): any modification of the 5'UTR region is likely to have an impact on the mRNA translation and stability regardless its role in Hfq-binding. However, it is argued that because the main Hfq-binding site is localized in close proximity to the RBS of *ptrA* (and not *recB*) mRNA, it may be possible to test the specificity of Hfq regulation by mutating the binding sequence within *ptrA*. Once the Hfq binding is disrupted, it is expected to result in more efficient translation of RecB as a result of translational coupling between *ptrA* and *recB*. Quantification of the *recB* transcriptional levels will be needed to control for any changes in mRNA abundance, which might be caused by modification of the sequence.

Is a small RNA involved in the regulation mechanism of RecB expression? While in the majority of the characterized cases a small RNA is needed for specificity of the interaction, recent studies reported on Hfq performing regulation independently of sRNAs. For instance, *mutS* and *Tn10* in *Escherichia coli*, *cirA* and *amiE* in *Pseudomonas aeruginosa* were shown to be regulated by Hfq in the absence of any sRNA [45, 220, 65, 194]. Although RecB quantification in the Hfq point-mutated and sRNA over-expression/deletion experiments do not exclude the possibility of additional player(s) of the network, no sRNA was identified as a strong candidate. Therefore, there is a chance of another sRNA(s) being involved however the very low abundance of the *ptrA-recB-recD* transcripts is likely to make it particularly difficult to detect RNA-RNA interactions in bulk genome-wide experiments. Thus, in order to investigate the regulation of low-copy enzymes, a novel single-molecule methodology is necessary to be developed.

Potential role of Hfq in DNA repair and SOS response

The Hfq-mediated (with a potential assistance of a sRNA) control of RecB expression may be just a hint of a larger bacterial stress-response program to DNA damage. In fact, post-transcriptional regulation has been mentioned in relation to DNA repair and genome maintenance pathways. For instance, a DNA mismatch binding protein, MutS was shown to be mainly regulated by an sRNA-independent Hfq binding and its translation repression was linked to the increased mutagenesis in the stationary phase [45]. In addition, small RNAs (RybC, IsrA, GlmZ) showed evidence of involvement in SOS response however no mechanism was identified yet [133, 149]. Interestingly, two sRNAs (GcvB and ArcZ), which bind *ptrA-recB-recD* mRNAs [142], have been shown to promote mutagenic break repair [26, 183].

A modest (but not significant) decrease ($\sim$30%) in Hfq protein abundance has been seen in a proteomic study in *E. coli* upon DSB induction with ciprofloxacin [122]. However, it is conceivable that even modest changes in Hfq availability could result in significant changes in gene expression and this could explain the increased translational efficiency of RecB upon DBS induction (Fig. 4.8). If Hfq indeed directly represses RecB translation (Fig. 5.10), a reduction in Hfq protein abundance is expected to result in more efficient RecB synthesis. Taken together, this suggests that Hfq-mediated regulation might play a direct role in DNA repair pathway however it needs to be further investigated at both, transcriptional and translational, levels.

Potential role of Hfq in noise suppression

A possibility of the post-transcriptional regulation reducing protein fluctuations has been raised in theoretical and experimental studies [120, 213, 226]. Because mRNA molecules are among the lowest abundant components in a network, they may heavily contribute to the total noise within the system. Thus, a post-transcriptional control of these fluctuations was proposed as an effective strategy for reducing protein noise [35, 212, 226]. For instance, a post-transcriptional negative feedback was analytically predicted to have a shorter response time and provide a more robust control in comparison to transcriptional feedback networks [209]. In addition to theoretical studies, noise propagation was experimentally examined using Hfq-mediated sRNA-based synthetic circuits in *E. coli* and the use of highly abundant small RNAs was shown not to result in noise amplification [108].

Suppressed fluctuations of protein expression were experimentally detected for hundreds of microRNA-regulated genes in mouse embryonic stem cells [198]. Interestingly, those genes were among the lowly expressed. Further investigation of a coarse-grained stochastic model (applicable to both, mammalian and bacterial, cells) of a gene regulated post-transcriptionally emphasized the noise reduction in low-copy regime because of the buffering mRNA fluctuations with post-transcriptional regulators that are present in large abundance (relative to the mRNA targets) [141, 198]. A stochastic model of gene expression regulated post-transcriptionally through sRNAs was analyzed by Mehta *et al.*, and the results from the study are summarized in Figure 6.1. Based on the ratio between the values of mRNA transcription rate, α_m, and average protein number, p, three regimes were identified: repressed, crossover and expressing. Interestingly, when proteins are low abundant (repressed regimen), the distribution of protein molecules per cell was found to be narrowed down in comparison to the other cases. This noise suppression is due to the fact of highly abundant sRNAs buffering against target mRNA fluctuations by additional averaging over time. These theoretical results supported by the experimental observations of miRNA-regulated genes in mammalian cells could hint to a universal functional 'niche' for post-transcriptional regulators across the organisms. Finally, mechanisms similar to the one described in this study might underlay the expression networks of other DNA repair proteins, many of which are known to be present in small quantities.

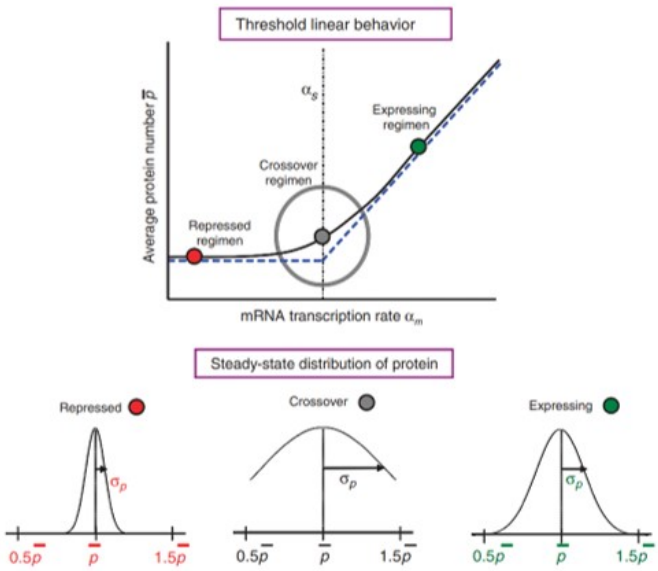

Figure 6.1: A coarse-grained stochastic model of gene expression regulated post-transcriptionally through sRNAs analyzed here [141]. Based on the ratio between the values of mRNA transcription rate, α_m, and average protein number, p, three regimes were identified: repressed, crossover and expressing. The transitioning between the regimes are defined by the sRNA transcription rate, α_s. Steady-state protein distributions for each regime are illustrated in the lower panel. The image was taken from Mehta *et al.* [141].

Chapter 7

Appendix I

7.1 Molecular biology techniques

7.1.1 Lambda red

Lambda red is a homologous recombination-based method for cloning which allows for direct modification of bacterial DNA. The technique is based on utilising the system derived from the lambda red bacteriophage [52].

According to the protocol, the integration kanamycin (Km) resistance cassette was amplified from pDK4 plasmid by PCR and linear DNA fragments were purified with DpnI restriction enzyme at 37℃ for one hour. Next, the ampicillin (Amp) resistance plasmid, pKD46, was transformed into MG1655 and MG1655 *recB165* :: *halotag E. coli* strains and positive clones were further selected on Amp at 30℃. A couple of clones were grown in LB with Amp at 30℃ until they $\mathrm{OD}_{600} \sim 0.5 - 1$ was reached and then reinoculated in SOB medium supplemented with Amp. At $\mathrm{OD}_{600} \sim 0.1$, expression of the lambda red system was induced with 10 mM L-arabinose until the cells reached $\mathrm{OD}_{600} \sim 0.5$. Next, the electrocompetent cells were electroporated with 150-200 ng of the purified PCR product and further recovered at 37℃ for one hour. The cells were plated onto LB plates supplemented with Km and incubated for $\sim$20 – 24 hours at 37℃. Large colonies were purified on LB plates with Km while the loss of pKD46 plasmid was verified by the absence of growth on LB plates with Amp. Finally, correct insertions were checked by colony PCR.

7.1.2 Plasmid-mediated gene replacement

Plasmid-mediated gene replacement (PMGR) is a cloning technique which is based on integration or deletion of a DNA fragment through homologous recombination [124]. The PMGR protocol utilizes a pTOF24 plasmid that has a temperature-sensitive replication origin, a chloramphenicol resistance gene and a sucrose gene of *Bacillus subtilis*. The key steps of the technique are shown in Figure 7.1.

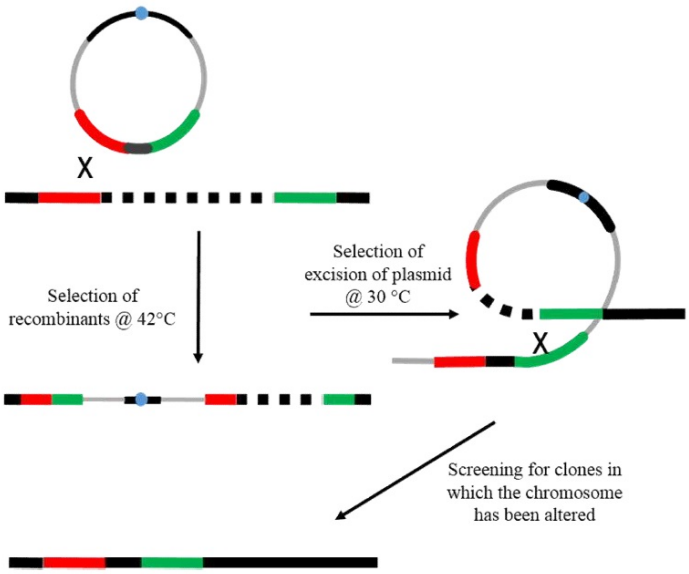

Figure 7.1: The illustration of the plasmid mediated gene replacement (PMGR) protocol. A pTOF24-derivative plasmid, which contains the desired insert surrounded by two homologous regions, is integrated into bacterial genome through recombination of one of the homology arms. Then, the plasmid DNA fragment is excised from the chromosome using one of the homology arms. Because of probabilistic nature of the homologous region used for recombination, clones are successful only in cases when one arm was used for the integration and the excision happened through the other homology. The image was borrowed from Azeroglu [22].

According to the protocol, a pTOF24 derivative plasmid was transformed into a MG1655 background and the strains were incubated overnight at 30℃ with chloramphenicol (Cm). Because of the temperature-sensitive replication origin, the plasmid is able to replicate only if it is integrated into the *E.coli* chromosome. After the incubation, Cm-resistant colonies were selected, streaked onto LB plates supplemented with Cm and grown overnight at 42℃. Then, colonies of different size were picked and purified by re-streaking on LB plates with Cm and growing overnight at 42℃. Next, the selection of cells with excision of the plasmid from the chromosome was performed by incubating cells overnight at 30℃ in liquid LB without any selection. After that, 10^{-4} and 10^{-5} dilutions were plated onto LB plates with 5% sucrose. The next morning, sucrose-resistant Cm-sensitive colonies were selected and checked by PCR for the deletion of the DNA fragment.

7.1.3 T4 assay

T4 phage is an enterobacteria phage which has double-stranded DNA and infects *E. coli*. The phage is able to undergo only a lytic lifecycle leading to formation of new virus particles, degradation of the *E. coli* membrane and phage propagation. The T4 assay utilizes the mutant of T4 phage with a deletion of *gene2*. In the wild type T4 phage, *gene2* is responsible for protection of virus DNA double-stranded ends from RecBCD action [126]. However, in the *gene2* mutants, dsDNA is not covered by the product of the *gene2* and is recognized by the RecBCD enzyme, leading to the degradation of the virus fragment [15]. In contrast, a bacterial strain which lacks RecBCD activity fails to degrade the phage DNA fragment and allows the phage to undergo its lytic lifecycle. This causes cell death and results in a plaque on a plate. Therefore, the T4 assay was used to test RecBCD activity in *E. coli* [201].

The protocol of the T4 assay is as follows: a strain was grown overnight in LB and the next day, 300 µl of the cells were incubated for 10 min with 100 µl of T4 phage. Then, this was mixed with 3 ml of warm soft agar (40% of melted LB and 60% of liquid LB), vortexed, poured onto a fresh LB plate and incubated overnight at 37°C. Finally, the plaques for each tested strain were counted.

Chapter 8

Appendix II

8.1 RecB and RecC expression plasmids

A serial dilution assay was carried out to verify that constructed *recC* and *recB* expression plasmids (pIK01 and pIK02) are capable of rescuing Δ*recC* and Δ*recB* mutant phenotypes (Fig. 8.1). The plasmids were transformed into BW27783, BW27783 Δ*recB* and BW27783 Δ*recC* strains (Table 10.5). LB plates were complemented with chloramphenicol for selection. 1% of arabinose was added to LB agar (ara^{+}) to fully induce the expression from the pBAD promoter. Nalidixic acid was used to induce DNA DSBs (Nal^{+}).

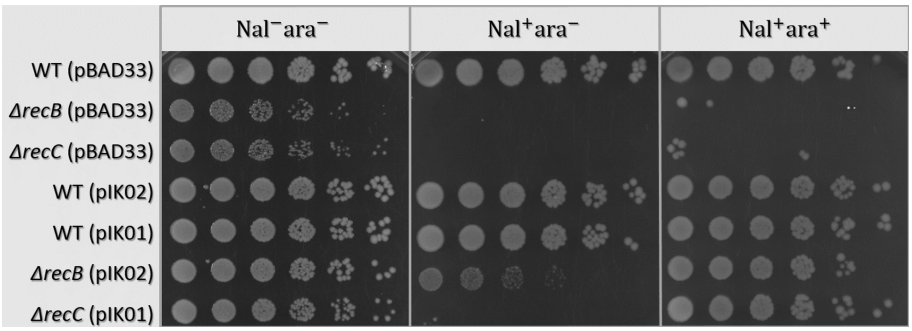

Figure 8.1: A 10-fold serial dilution assay for wild type, Δ*recB*, Δ*recC*, wild type carrying pIK01 (*recC*) or pIK02 (*recB*) expression plasmids, Δ*recB* and Δ*recC* harboring pIK01 or pIK02 expression plasmid. The BW27783 background strains were used. Three conditions are shown: LB plates without either nalidixic acid or arabinose ($\mathrm{Nal}^{-}\mathrm{ara}^{-}$); LB plates with 2.5 µg/ml of nalidixic acid but not arabinose ($\mathrm{Nal}^{+}\mathrm{ara}^{-}$); and LB plates with both reagents ($\mathrm{Nal}^{+}\mathrm{ara}^{+}$).

As seen from the results, both plasmids (pIK01 and pIK02) complement mutant phenotypes under the conditions of full expression. In the absence of arabinose, Δ*recC* with pIK01 plasmid lacks the essential RecC subunit, and thus, is unable to repair DSBs. A weak leaky expression from the pBAD promoter was detected for pIK02 plasmid what results in partial complementation of the Δ*recB* phenotype. Lastly, no toxic effect was found when RecB or RecC subunits were over-expressed in the wild-type cells.

Strain	Plasmid	# of plaques
WT	pBAD33	5
Δ*recB*	pBAD33	∼500
Δ*recC*	pBAD33	∼460
WT	pIK02	7
WT	pIK01	6
Δ*recB*	pIK02	8
Δ*recC*	pIK01	7

Table 8.1: The results of the T4 assay with pIK01 and pIK02 expression plasmids. Arabinose was added to the overnight cultures.

To test the activity of RecBCD enzyme produced from the plasmids, T4 assay was performed. Chloramphenicol was used for plasmid selection and 1% of arabinose was added to the overnight cultures. The results are shown in Table 8.1 and consistent with the previous conclusions.

Taken together, the results show that (i) pIK01 and pIK02 plasmids are capable of rescuing Δ*recC* and Δ*recB* phenotypes and (ii) no toxic effect of RecB or RecC subunit over-expression has been detected.

8.2 RecB quantification upon *ptrA* over-expression

Although Hfq is present in high abundance (10,000 molecules per cell), mRNAs and sRNAs targets are constantly competing for Hfq binding [247]. This led to the following experiment. Based on the binding between Hfq and *ptrA-recB* mRNA, sequestering Hfq molecules by over-expressing *ptrA* mRNAs was proposed as a potential way to further verify the Hfq-binding *in vivo* (Fig. 8.2). To this end, two constructs were built (Table 10.6): the pBAD33-derivative plasmids carrying either the entire *ptrA* gene (pIK06) or its truncated version containing the Hfq-binding region (pIK07) under control of an arabinose-inducible promoter.

The effect of *ptrA* overproduction on RecB protein expression was tested with HaloTag-labelling method (Fig. 8.3). First, the backbone plasmid pBAD33 did not cause any changes in RecB distribution (Fig. 8.3, left plot). Interestingly, a slight shift of the RecB distribution was detected when a truncated version of the *ptrA* gene was produced (Fig. 8.3, right plot). However, no significant change was observed for the expression of the entire *ptrA* gene (Fig. 8.3, middle plot). This may be a consequence of a modified stability of the truncated *ptrA* transcripts

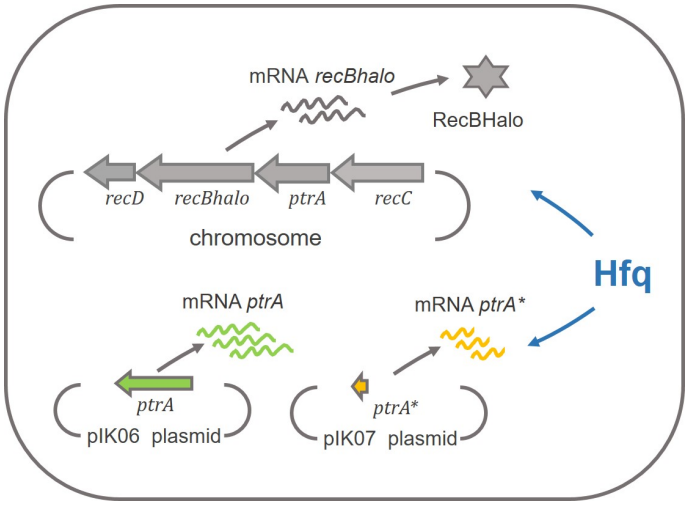

Figure 8.2: Schematic of the *ptrA* over-expression experiment. The entire or truncated *ptrA* mRNAs were produced from pBAD33-derivative plasmids in the wild-type strain. Full expression from the plasmids was induced with 1% arabinose. Presumably, Hfq molecules are titrated by *ptrA* over-expression resulting in the increased RecB protein production.

resulting in a different (possibly, higher) *ptrA* mRNA steady-state level. To test this hypothesis, the *ptrA* mRNA levels need to be quantified with RT-qPCR. In order to make any further interpretations of this preliminary experiment, the HaloTag-labelling quantification needs to be repeated.

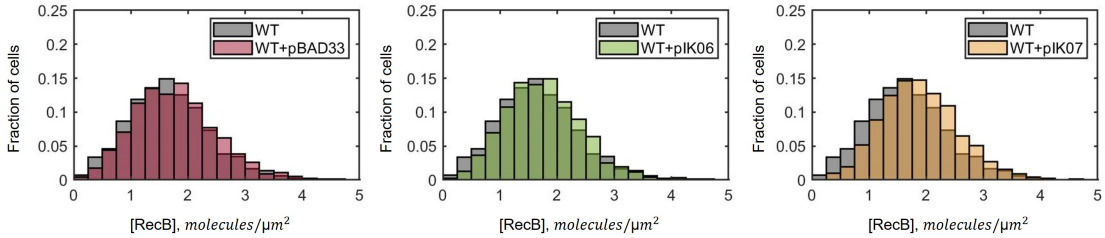

Figure 8.3: RecB protein concentration distributions quantified in wild type carrying the backbone plasmid pBAD33, the plasmid expressing the entire (pIK06) or a truncated (pIK07) *ptrA* gene. The histograms represent the results of only one experiment. Total number of cells included into distributions are 3,361 (WT+pBAD33); 1,216 (WT+pIK06); 2,271 (WT+pIK07).

Chapter 9

Appendix III

9.1 A decay process: mean and variance evolution

A decay process can be described by a simple reaction, where M is a single species in a system and γ_m is a constant degradation rate:

$$M \xrightarrow{\gamma_m} \varnothing$$

As previously, the number of mRNA in a cell at time t is defined as $m(t)$. Here, the following question has been addressed: how mRNA mean and variance depend on time? A chemical master equation for the probability function $\widetilde{P}(m,t)$ of having m mRNA molecules at time t is given by:

$$\frac{d\widetilde{P}(m,t)}{dt} = \gamma_m(m+1)\widetilde{P}(m+1,t) - \gamma_m m\widetilde{P}(m,t) \tag{9.1}$$

Then, the ordinary differential equations (ODEs) of the first and second moment evolution can be written as follows [199]:

$$\frac{d\langle m\rangle}{dt} = -\gamma_m\langle m\rangle \tag{9.2}$$

$$\frac{d\langle m^2\rangle}{dt} = \gamma_m\langle m\rangle - 2\gamma_m\langle m^2\rangle \tag{9.3}$$

Assuming the initial conditions given by $\langle m\rangle_{t=0} = \langle m\rangle_0$ and $\sigma^2_{m_{t=0}} = \sigma^2_{m_0}$, the solutions of the ODEs above can be easily obtained:

$$\langle m\rangle(t) = \langle m\rangle_0 e^{-\gamma_m t} \tag{9.4}$$

$$\sigma_m^2(t) = (\sigma_{m_0}^2 - \langle m \rangle_0) e^{-2\gamma_m t} + \langle m \rangle_0 e^{-\gamma_m t} \tag{9.5}$$

9.2 A decay process: PDF evolution

In this section, an analytical time-dependent solution $\widetilde{P}(m,t)$ for a decay process is derived for the following initial conditions on the steady-state (mRNA) probability distribution at time $t = 0$ denoted as $\widetilde{P}(m,0)$: (a) a fixed number of (mRNA) molecules m_0; (b) a Poisson distribution $Pois(\lambda)$; and (c) a negative binomial distribution $NB(r,p)$. This process models a biological experiment where a system reaches its steady state, and then, a production/synthesis reaction is inhibited. Intuitively, after the perturbation, at the limit of infinite time the system will eventually come to the state where all species (mRNAs) have been degraded: $\widetilde{P}(m,\infty) = \widetilde{P}(0,\infty) = 1$. For this reason, it is sometimes called as a non-stationary decay process with an absorbing state.

To find an analytical solution $\widetilde{P}(m,t)$ for a non-stationary decay process with an absorbing state $m = 0$ and an initial (mRNA) distribution at time $t = 0$ given by a fixed number, a Poisson or a negative binomial distribution, the method of a path integral has been applied [56, 172, 245].

For a given probability distribution $\widetilde{P}(m,t)$, its generating function $G(z,t)$ is defined by the following expression:

$$G(z,t) = \sum_{m=0}^{\infty} z^m \widetilde{P}(m,t) \tag{9.6}$$

Then, for a given generating function $G(z,t)$, the probability distribution $\widetilde{P}(m,t)$ can be recovered as follows:

$$\widetilde{P}(m,t) = \frac{1}{m!} \frac{\partial^m G(z,t)}{\partial z^m}\bigg|_{z=0} \tag{9.7}$$

As previously shown for a decay process [56, 172], evolution of a generating function $G(z,t)$ can be expressed in terms of an initial generating function at time $t = 0$ denoted as $G(z,0) = G_0$:

$$G(z,t) = G_0 \left[1 + (z-1)e^{-\gamma t}\right] \tag{9.8}$$

where γ is a constant probability per time unit of the decay reaction to happen (mRNA degradation rate). Expression 9.8 should be understood as a functional relation.

Initial condition: Fixed number m_0

For the initial condition given as a fixed number of molecules m_0, the generating function at time $t = 0$ is given as $G_{fix}(z,0) = z^{m_0}$. Thus, after applying the result from Eq. 9.8, the evolution of generating function results in:

$$G_{fix}(z,t) = \left(1 + (z-1)e^{-\gamma t}\right)^{m_0} \tag{9.9}$$

According to a property of a generating function, the probability density function can be found with Equation 9.7:

$$\widetilde{P}(m,t) = \frac{1}{m!}\frac{\partial^m}{\partial z^m}\left(1 + (z-1)e^{-\gamma t}\right)^{m_0}\Big|_{z=0} = \frac{1}{m!}\frac{m_0(m_0-1)...(m_0-m+1)}{e^{\gamma t m}}\left(1-e^{-\gamma t}\right)^{m_0-m}$$

$$= \frac{m_0!}{m!(m_0-m)!}e^{-\gamma m t}(1-e^{-\gamma t})^{m_0-m} \tag{9.10}$$

The obtained PDF is consistent with the one obtained by a different approach here [244].

Initial condition: Poisson distribution $Pois(\lambda)$

In the case when the initial condition is given by a Poisson distribution $\widetilde{P}(m,0) = Pois(\lambda)$, the generating function at time $t = 0$ can be derived as follows:

$$G_{Pois}(z,0) = \sum_{m=0}^{\infty} z^m \frac{\lambda^m}{m!}e^{-\lambda} = e^{-\lambda}\sum_{m=0}^{\infty}\frac{(z\lambda)^m}{m!} = e^{\lambda(z-1)} \tag{9.11}$$

Next, by utilizing the functional relation Eq. 9.8, the evolution of the generation function over time can be obtained:

$$G_{Pois}(z,t) = e^{\lambda(z-1)e^{-\gamma t}} \tag{9.12}$$

Further, the PDF can be calculated by applying Equation 9.7:

$$\widetilde{P}(m,t) = \frac{1}{m!}\frac{\partial^m}{\partial z^m}\left(e^{\lambda(z-1)e^{\gamma t}}\right)\Big|_{z=0} = \frac{\lambda^m e^{-\gamma m t}}{m!}e^{\lambda(z-1)e^{-\gamma t}}\Big|_{z=0}$$

$$= \frac{\lambda^m}{m!}e^{-(\gamma m t + \lambda e^{-\gamma t})} \tag{9.13}$$

Initial condition: Negative binomial distribution $NB(r,p)$

For a negative binomial distribution $NB(r,p)$ with the parameters r (number of successful events) and p (probability of success), the generating function takes the form:

$$G_{NB}(z,0) = \sum_{m=0}^{\infty} z^m \frac{(r+m-1)!}{(r-1)!m!} p^r (1-p)^m = p^r \sum_{m=0}^{\infty} \frac{(r+m-1)!}{(r-1)!m!} (z(1-p))^m$$

$$= p^r (1-z(1-p))^{-r} = \left(\frac{p}{1-z(1-p)}\right)^r \tag{9.14}$$

Here, $(1-p)$ was temporarily replaced by q to simplify further calculations. By substituting this to Eq. 9.8, $G_{NB}(z,t)$ can be explicitly expressed as follows:

$$G_{NB}(z,t) = \frac{p^r}{(1-(1+(z-1)e^{-\gamma t})q)^r} = p^r (1-q(1+(z-1)e^{-\gamma t}))^{-r} \tag{9.15}$$

Similarly to the previous cases, the PDF can be found by utilizing Equation 9.7:

$$\widetilde{P}(m,t) = \frac{1}{m!}\frac{\partial^m}{\partial z^m}\left(p^r (1-q(1+(z-1)e^{-\gamma t}))^{-r}\right)\Big|_{z=0} = \frac{p^r}{m!}\frac{\partial^m}{\partial z^m}\left(1-q+qe^{-\gamma t}-zqe^{-\gamma t}\right)\Big|_{z=0}$$

$$= \frac{(-1)^m p^r q^m}{m!}(-r)(-r-1)...(-r-m+1)\frac{e^{-\gamma m t}}{(1-q+qe^{-\gamma t})^{r+m}}$$

$$= \frac{(r+m-1)!}{m!(r-1)!} p^r q^m \frac{e^{-\gamma m t}}{(1-q+qe^{-\gamma t})^{r+m}}$$

$$= \frac{(r+m-1)!}{m!(r-1)!} p^r (1-p)^m \frac{e^{-\gamma m t}}{(p+(1-p)e^{-\gamma t})^{r+m}} \tag{9.16}$$

9.3 Bayesian and Akaike information criteria (BIC and AIC)

The probability density function (PDF) of a Poisson distribution $Pois(\lambda)$, which is expected for the one-state promoter model, is given as follows:

$$f(x,\lambda) = \frac{\lambda^x}{x!} e^{-\lambda} \tag{9.17}$$

For a negative binomial distribution, which is expected for the bursty transcription model, the PDF is represented by the expression:

$$f(x,r,p) = \frac{\Gamma(r+x)}{\Gamma(r)\Gamma(x+1)} p^r (1-p)^x \tag{9.18}$$

where Γ denotes the gamma function, while r and p are fitting parameters, known as a number of success and a probability of success, respectively.

By definition, a likelihood function for a variable X with PDF $f(x,\lambda)$ is given as:

$$L(X \mid \lambda) = \prod_{x_i \in X} f(x_i \mid \lambda) \tag{9.19}$$

Here X represents a vector of an outcome variable x_i (in the i^{th} experiment) and λ is a parameter of the density function. The notation $X \mid \lambda$ means "X given λ".

For a Poisson distribution with PDF (Eq. 9.17) the following likelihood function can be obtained:

$$L(X \mid \lambda) = \prod_{x_i \in X} f(x_i \mid \lambda) = f(x_1 \mid \lambda) f(x_2 \mid \lambda) ... f(x_n \mid \lambda) = \frac{\lambda^{x_1}}{x_1!} e^{-\lambda} ... \frac{\lambda^{x_n}}{x_n!} e^{-\lambda}$$

$$L(X \mid \lambda) = e^{-n\lambda} \frac{\lambda^{x_1+x_2+...+x_n}}{x_1! x_2! ... x_n!} = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \tag{9.20}$$

By applying a natural logarithm function to the last expression, the following log-likelihood function for the one-state model is derived:

$$\ln L = -n\lambda + \left(\sum_{i=1}^{n} x_i \right) \ln \lambda - \ln \prod_{i=1}^{n} x_i! \tag{9.21}$$

Similar derivations can be performed for a negative binomial distribution with PDF (9.18):

$$L(X \mid r,p) = \prod_{x_i \in X} f(x_i \mid r,p) \tag{9.22}$$

$$L(X \mid r,p) = \prod_{x_i \in X} f(x_i \mid r,p) = f(x_1 \mid r,p) f(x_2 \mid r,p) ... f(x_n \mid r,p) =$$

$$= \frac{\Gamma(r+x_1)}{\Gamma(r)\Gamma(x_1+1)} p^r (1-p)^{x_1} ... \frac{\Gamma(r+x_n)}{\Gamma(r)\Gamma(x_n+1)} p^r (1-p)^{x_n} =$$

$$= \left(\frac{p^r}{\Gamma(r)}\right)^n \frac{\Gamma(r+x_1)...\Gamma(r+x_n)}{\Gamma(x_1+1)...\Gamma(x_n+1)} (1-p)^{\Sigma_{i=1}^n x_i}$$

$$L(X \mid r,p) = \left(\frac{p^r}{\Gamma(r)}\right)^n \frac{\Gamma(r+x_1)...\Gamma(r+x_n)}{\Gamma(x_1+1)...\Gamma(x_n+1)} (1-p)^{\Sigma_{i=1}^n x_i} \tag{9.23}$$

$$\ln L = nr \ln p + \left(\sum_{i=1}^n x_i\right) \ln(1-p) - n \ln\Gamma(r) + \sum_{i=1}^n \ln\Gamma(r+x_i) - \sum_{i=1}^n \ln\Gamma(x_i+1) \tag{9.24}$$

Finally, BIC and AIC scores for each model can be calculated using the expressions:

$$BIC = -2 \ln L + q \ln n \tag{9.25}$$

$$AIC = -2 \ln L + 2q \tag{9.26}$$

Here, $\ln L$ represents a log-likelihood function for either a Poisson or a negative binomial distribution (Eq. 9.21 or Eq. 9.24, respectively), n is a sample size and q defines the number of fitting parameters associated with each model. In our case, q equals one for a Poisson distribution (one-state model) and two for a negative binomial distribution (bursty transcription model).

9.4 The sum of two independent negative binomial variables

The following mathematical result was used in Chapter 3: the sum of two independent variables, each of which is distributed as $NB(r,p)$, is described by a negative binomial distribution $NB(2r,p)$. Here, a proof of a more general case is reviewed: if X and Y are independent variables and $X \sim NB(r_1,p)$ and $Y \sim NB(r_2,p)$, then $(X+Y) \sim NB(r_1+r_2,p)$.

The probability density functions for the variables X and Y are given by the following expressions:

$$P(X=k,r_1,p) = \frac{\Gamma(r_1+k)}{\Gamma(r_1)\Gamma(k+1)} p^{r_1} (1-p)^k \tag{9.27}$$

$$P(Y=k,r_2,p)=\frac{\Gamma(r_2+k)}{\Gamma(r_2)\Gamma(k+1)}p^{r_2}(1-p)^k \tag{9.28}$$

First, the moment generating function is derived for a negative binomial distribution. The moment generating function $M_X(s)$ of a random variable X is defined as follows:

$$M_X(s)=E[e^{sX}] \tag{9.29}$$

Then, for a random variable that is distributed as $NB(r,p)$, the moment generating function is derived as:

$$M_X(s)=E[e^{sX}]=\sum_{k=0}^{\infty}e^{sk}\frac{\Gamma(r+k)}{\Gamma(r)\Gamma(k+1)}p^r(1-p)^k=p^r\sum_{k=0}^{\infty}\frac{\Gamma(r+k)}{\Gamma(r)\Gamma(k+1)}(e^s(1-p))^k$$

$$=p^r\sum_{k=0}^{\infty}\frac{(r+k-1)!}{(r-1)!k!}(e^s(1-p))^k=p^r\sum_{k=0}^{\infty}\frac{r(r+1)...(r+k-1)}{k!}(e^s(1-p))^k$$

$$=p^r\sum_{k=0}^{\infty}\frac{(-1)^k(-r)(-r-1)...(-r-k+1)}{k!}(e^s(1-p))^k$$

$$=p^r\sum_{k=0}^{\infty}\frac{(-1)^k(-r)!}{(-r-k)!k!}(e^s(1-p))^k=p^r\sum_{k=0}^{\infty}\binom{-r}{k}(-e^s(1-p))^k$$

$$=\left(\frac{p}{1-(1-p)e^s}\right)^r \tag{9.30}$$

Finally, applying the property of multiplication of moment generating functions for the sum of independent variables, the proof of the statement above is finalized:

$$M_{X+Y}(s)=M_X(s)M_Y(s)=\left(\frac{p}{1-(1-p)e^s}\right)^{r_1}\left(\frac{p}{1-(1-p)e^s}\right)^{r_2}$$

$$=\left(\frac{p}{1-(1-p)e^s}\right)^{r_1+r_2} \tag{9.31}$$

Indeed, this is the moment generating function of a $NB(r_1+r_2,p)$ random variable. Thus, $(X+Y)\sim NB(r_1+r_2,p)$.

9.5 Deterministic model of RecB expression upon DSB induction

In this section, the dynamics of RecB mRNA and protein levels after a perturbation with ciprofloxacin treatment is analytically derived. Taking into account different time-scales of the response in mRNA and protein levels, the following question was addressed: how long does the DSB induction need to be applied for to see changes (if any) in both species? Average mRNA and protein concentrations instead of molecule numbers were considered because of cell elongation upon ciprofloxacin treatment. Therefore, a two-stage model of RecB expression was modified by including time-dependency of cell volume, $V(t)$.

As discussed above, the same set of reactions for two species of the system has been considered: mRNAs and proteins. While the processes of translation, active mRNA and protein degradation are described by first-order reactions with constant rates: k_p, γ_m and $\gamma_{p_{act}}$ respectively, mRNA transcription is given by a zeroth-order reaction. According to the law of mass action, a zeroth-order reaction rate shall be scaled by the volume of the system, as $k_m V(t)$ [199, 244]. As previously, the number of mRNA and protein molecules at time t in a cell of volume $V(t)$ are defined as $m(t)$ and $p(t)$, respectively. Then, the evolution of the average numbers of mRNA and protein molecules in time is described by the following rate equations:

$$\frac{d\langle m(t)\rangle}{dt} = k_m V(t) - \gamma_m \langle m(t)\rangle \tag{9.32}$$

$$\frac{d\langle p(t)\rangle}{dt} = k_p \langle m(t)\rangle - \gamma_{p_{act}} \langle p(t)\rangle \tag{9.33}$$

Further, *recB* mRNA and RecB protein concentrations are defined as $c_m(t) = m(t)/V(t)$ and $c_p(t) = p(t)/V(t)$. Thus, differentiation with respect to t results in:

$$\frac{dc_m(t)}{dt} = \frac{1}{V(t)}\frac{dm(t)}{dt} - \frac{c_m(t)}{V(t)}\frac{dV(t)}{dt} \tag{9.34}$$

$$\frac{dc_p(t)}{dt} = \frac{1}{V(t)}\frac{dp(t)}{dt} - \frac{c_p(t)}{V(t)}\frac{dV(t)}{dt} \tag{9.35}$$

In order to obtain time-dependent solutions for mRNA and protein concentrations, the rate equations (Eq. 9.32 and Eq. 9.33) were normalized by cell volume $V(t)$; $dm(t)/dt$ and $dp(t)/dt$ were replaced with the expressions from the equations: Eq. 9.34 and Eq. 9.35. Finally, an exponential dependence of cell volume from time has been assumed: $V(t) = V_0 e^{\gamma t}$, where γ is a growth rate, as that was demonstrated in several experimental studies [148, 252]. As a result, following ordinary differential equations for mRNA and protein concentrations were obtained:

$$\frac{d\langle c_m(t)\rangle}{dt} = k_m - (\gamma_m + \gamma)\langle c_m(t)\rangle \tag{9.36}$$

$$\frac{d\langle c_p(t)\rangle}{dt} = k_p\langle c_m(t)\rangle - (\gamma_{p_{act}} + \gamma)\langle c_p(t)\rangle \tag{9.37}$$

By equating the left-hand sides of the equations to zero, the steady-state mRNA and protein concentrations, c_m^{ss} and c_p^{ss}, can be found:

$$c_m^{ss} = \frac{k_m}{\gamma_m + \gamma} \qquad c_p^{ss} = \frac{k_m}{(\gamma_m + \gamma)}\frac{k_p}{(\gamma_{p_{act}} + \gamma)} \tag{9.38}$$

Next, by assuming the initial conditions for mRNA and protein concentration to be $c_{m(t=0)}$ and $c_{p(t=0)}$, the time-dependent solutions for mRNA and protein concentrations have been obtained:

$$\langle c_m(t)\rangle = \left(c_{m0} - \frac{k_m}{\gamma_m + \gamma}\right)e^{-(\gamma_m+\gamma)t} + \frac{k_m}{\gamma_m + \gamma} \tag{9.39}$$

$$\langle c_p(t)\rangle = \left(c_{p0} + \frac{c_{m0}k_p}{\gamma_m - \gamma_{p_{act}}} - \frac{k_p k_m}{(\gamma_m - \gamma_{p_{act}})(\gamma_{p_{act}} + \gamma)}\right)e^{-(\gamma_{p_{act}}+\gamma)t} + \tag{9.40}$$

$$+\left(-\frac{c_{m0}k_p}{\gamma_m - \gamma_{p_{act}}} + \frac{k_p k_m}{(\gamma_m - \gamma_{p_{act}})(\gamma_m + \gamma)}\right)e^{-(\gamma_m+\gamma)t} + \frac{k_p k_m}{(\gamma_{p_{act}} + \gamma)(\gamma_m + \gamma)}$$

The last expression can be simplified by taking into account that the RecB protein does not have an active degradation mechanism (Chapter 3). Thus, after imposing the condition of $\gamma_{p_{act}} = 0$, a time-dependent solution for RecB protein concentration is obtained as follows:

$$\langle c_p(t)\rangle = \left(c_{p0} + \frac{c_{m0}k_p}{\gamma_m} - \frac{k_p k_m}{\gamma_m \gamma}\right)e^{-\gamma t} + \left(-\frac{c_{m0}k_p}{\gamma_m} + \frac{k_p k_m}{\gamma_m(\gamma_m + \gamma)}\right)e^{-(\gamma_m+\gamma)t} + \tag{9.41}$$

$$+\frac{k_p k_m}{\gamma(\gamma_m + \gamma)}$$

Chapter 10

Appendix IV

10.1 Tables

Table 10.1: Solutions and buffers used in the study.

Solution	Composition
M9 salts (4X)	197 mM Na_2HPO_4, 88.2 mM KH_2PO_4, 34.2 mM $NaCl$, 74.8 mM NH_4Cl dissolved in dH_2O
Transformation Storage Buffer (TSB)	10% (w/v) PEG (8000), 5% (w/v) DMSO, 50 mM $MgCl_2$, dissolved in LB
GCT-Phenol	0.2 g/ml of GTC, 21.2 mM Tris-HCl (pH8) and 4.24 mM EDTA (pH8), 0.42% β-mercaptoethanol, 0.85% sarkosyl and 40% of phenol (pH7) mixed in DEPC-treated water
Phenol-Chloroform-IAA	25:24:1 (v:v:v) phenol(pH7):chloroform:isoamyl alcohol
Chloroform-IAA	24:1 (v:v) chloroform:isoamyl alcohol

Table 10.2: Growth media used in the study.

Medium	Composition
Luria broth (LB)	1% Bacto-tryptone, 0.5% yeast extract, 1% NaCl
LB agar	1% Bacto-tryptone, 0.5% yeast extract, 1% NaCl, 1.5% Bacto-agar
SOB Medium	0.5% (w/v) Yeast extract, 2% (w/v) Tryptone, 10 mM NaCl, 10 mM KCl, 10 mM $\mathrm{MgCl_2}$, 10 mM $\mathrm{MgSO_4}$, dissolved in $\mathrm{dH_2O}$
SOC Medium	0.5% (w/v) Yeast extract, 2% (w/v) Tryptone, 10 mM NaCl, 10 mM KCl, 10 mM $\mathrm{MgCl_2}$, 10 mM $\mathrm{MgSO_4}$, dissolved in $\mathrm{dH_2O}$, 20 mM glucose
Imaging Medium	25% M9 Salts (4X), 10% LB, 0.2% glucose, 2 mM $\mathrm{MgSO_4}$, 0.1 mM $\mathrm{CaCl_2}$, dissolved in $\mathrm{dH_2O}$
Generic Medium	25% M9 Salts (4X), 0.2% glucose, 2 mM $\mathrm{MgSO_4}$, 0.1 mM $\mathrm{CaCl_2}$, 2% MEM Essential (X50), 1% MEM Non-Essential (X100), dissolved in $\mathrm{dH_2O}$

Table 10.3: Nutrient supplements used in the study.

Supplement	Abbreviation	Solvent	Final concentration
Arabinose	ara	$\mathrm{dH_2O}$	$10^{-5} - 1\%$
Glucose	glu	$\mathrm{dH_2O}$	0.2%
Glycerol	gly	$\mathrm{dH_2O}$	0.2%

Table 10.4: Antibiotics used in the study.

Antibiotic	Abbreviation	Solvent	Final concentration
Ampicillin	Amp	$\mathrm{dH_2O}$	50 – 100 µg/ml
Chloramphenicol	Cm	100% ethanol	30 µg/ml
Ciprofloxacin	Cipro	$\mathrm{dH_2O}$	2 – 12 ng/ml
Kanamycin	Km	$\mathrm{dH_2O}$	50 µg/ml
Nalidixic acid	Nal	$\mathrm{dH_2O}$	1 – 3 µg/ml
Rifampicin	Rif	$\mathrm{dH_2O}$	500 µg/ml

Table 10.5: *Escherichia coli* strains used in the study.

Strain	Genotype	Source
MG1655	$F^- \lambda^-$ *ilvG*$^-$ *rfb-50 rph-1*	Lab stock
BW27783	*lacI*^{q} *rrnB*$_{T14}$ Δ*lacZ*$_{WJ16}$ *hsdR514* Δ*araBAD*$_{AH33}$ Δ*rhaBAD*$_{LD78}$ *DE araFGH*Φ(Δ*araEp*P_{CP8}*-araE*)	Lab stock
DH5α	$F^- \phi 80$*lacZ*Δ$M15$ Δ(*lacZYA-argF*)$U169$ *recA1 endA1 hsdR17*(rk^-, mk^+) *phoA supE44* λ^- *thi-1 gyrA96 relA1*	Lab stock
MEK65	MG1655 *recB165*::*halotag*	[119]
MEK1326	MG1655 Δ*recB*	[119]
MEK1345	MG1655 Δ*recC*	This work*
MEK1329	BW27783 Δ*recB*	This work*
MEK1349	BW27783 Δ*recC*	This work*
MEK1902	MG1655 Δ*hfq*	This work**
MEK1457	MG1655 *recB165*::*halotag* Δ*hfq*	This work**
MEK1888	MG1655 Δ*chiX*	This work**
MEK1449	MG1655 *recB165*::*halotag* Δ*chiX*	This work**
MEK1887	MG1655 Δ*cyaR*	This work**
MEK1453	MG1655 *recB165*::*halotag* Δ*cyaR*	This work**
MEK1919	MG1655 Δ*arcZ*	This work**
MEK1921	MG1655 *recB165*::*halotag* Δ*arcZ*	This work**

* Constructed under supervision of Dr Benura Azeroglu
** Constructed by Léna Le Quellec and Dr Ira Alexandra Iosub

Table 10.6: Plasmids used in the study.

Plasmids	Description	Source
pTOF24Δ*recB*	pTOF24-derivative plasmid used for construction of Δ*recB* strains (MEK1326, MEK1329)	[51] [1]
pTOF24Δ*recC*	pTOF24-derivative plasmid used for construction of Δ*recC* strains (MEK1345, MEK1349)	[51] [1]
pBAD33	Backbone plasmid used for construction of pIK01-02 and pIK06-07	[85]
pIK01	pBAD33-derivative plasmid carrying *recC* gene under control of arabinose-inducible promoter, P*araBAD*	This work
pIK02	pBAD33-derivative plasmid carrying *recB* gene under control of arabinose-inducible promoter, P*araBAD*	This work
pZA21	Backbone plasmid used for construction of pZA21-ChiX, pZA21-CyaR, pZA21-ArcZ and pZA21-GcvB	[238]
pZA21-ChiX	pZA21-derivative plasmid carrying *chiX*	This work*
pZA21-CyaR	pZA21-derivative plasmid carrying *cyaR*	[100]
pZA21-ArcZ	pZA21-derivative plasmid carrying *arcZ*	[100]
pZA21-GcvB	pZA21-derivative plasmid carrying *gcvB*	[100]
pDWS2	pBR322-derivative plasmid carrying native *recC-ptrA-recB-recD* chromosomal region	[180] [2]
pQE80L	Backbone plasmid used for construction of pQE-Hfq-His and pQE-Hfq(R16A/Y25D/K56A)	[157] [3]
pQE-Hfq	pQE80L-derivative plasmid carrying *hfq-*His_6	[157] [3]
pQE-Hfq(R16A)	pQE80L-derivative plasmid carrying *hfq(R16A)-*His_6	[157] [3]
pQE-Hfq(Y25D)	pQE80L-derivative plasmid carrying *hfq(Y25D)-*His_6	[157] [3]
pQE-Hfq(K56A)	pQE80L-derivative plasmid carrying *hfq(K56A)-*His_6	[157] [3]

pIK06	pBAD33-derivative plasmid carrying *ptrA* gene under control of arabinose-inducible promoter, P*araBAD*	This work
pIK07	pBAD33-derivative plasmid carrying a truncated *ptrA* gene under control of arabinose-inducible promoter, P*araBAD*	This work

[1] Gift from Prof. David Leach, The University of Edinburgh
[2] Gift from Prof. Gerald Smith, The Fred Hutchinson Cancer Research Center
[3] Gift from Teppei Morita and Hiroji Aiba, Suzuka University of Medical Sciences
[*] Constructed by Dr Ira Alexandra Iosub

Table 10.7: Primers used for strain and plasmid construction.

Primer ID	Sequence (5' – 3')
hfq_H1_P1	GAATCGAAAGGTTCAAAGTACAAATAAGCATATAAGGAAAAG AGAGAATGGTGTAGGCTGGAGCTGCTTC
hfq_H2_P2	CTCCCCGTGTAAAAAAACAGCCCGAAACCTTATTCGGTTTCT TCGCTGTCGGTCCATATGAATATCCTCCTTAG
ChiX_H1_P1	TCTTGCCTAAGAGTATTGGCAGGATGGTGAGATTGAGCGACA ATCGAGTTGTGTAGGCTGGAGCTGCTTC
ChiX_H2_P2	CACCTGTATGGAGAAGGGAATTTGCCGCAAATGTTGCGCTAA AAAAATGGCGGTCCATATGAATATCCTCCTTAG
CyaR_H1_P1	CTTAGAAACCGATCACATACAGCTGCATTTATTAAGGTTATCA TCCGTTTCGTGTAGGCTGGAGCTGCTTC
CyaR_H2_P2	TGTGTGGACGTGACCAGAAATAAATCCTTTTATTTCATTGTAT TACGCGTGGTCCATATGAATATCCTCCTTAG
ArcZ_H1_P1	ATTTAACTGATTCATGTAACAAATCATTTAAGTTTTGCTATCTT AACTGCGTGTAGGCTGGAGCTGCTTC
ArcZ_H2_P2	AGTGGCGTCACGACGTAGAAGTGCTGAAAGCGTGGGTGGCA AAAGCCACTGGTCCATATGAATATCCTCCTTAG
chiX_ZA21	ACACCGTCGCTTAAAGTGACGGCATAATAATAAAAAAATGAA ATTCCTCTTTGACGGGCCAATAGCGATATTGGCCATTTTTTTG GTACGCGTGCTAGAGGCATC
pZA21MCS_5P	5P-GTGCTCAGTATCTCTATCACTGA
oIK01	ACCCGTTTTTTTGGGCTAGCGAATTCGAGCTCGGTCAGGAGC CGCTATGTTAAGG
oIK02	CTTCTCTCATCCGCCAAAACAGCCAAGCTTACAGCACTCATG ACTGATTAAAG
oIK03	AAGCTTGGCTGTTTTGGCGGATGAG
oIK04	ACCGAGCTCGAATTCGCTAGCCCAAAA
oIK10	ACCCGTTTTTTTGGGCTAGCGAATTCGAGCTCGGTCCTGATG AGTGAAAAGAATGAGTG
oIK11	GAAAATCTTCTCTCATCCGCCAAAACAGCCAAGCTTCTCCAC AGCTTCCAGTAATTGC

Primer ID	Sequence (5’ – 3’)
oIK66	CCCGTTTTTTTGGGCTAGCGAATTCGAGCTCGGTGCGCACCAGTCACGGACTGATGG
oIK67	GAAAATCTTCTCTCATCCGCCAAAACAGCCAAGCTTTTAGGCGCTCACCCTGTAAGGGC
oIK68	CTGAAAATCTTCTCTCATCCGCCAAAACAGCCAAGCTTTTACCACCAGCGCCGAGAGCG

Table 10.8: Oligos used for RT-qPCR quantification.

Oligo ID	Sequence (5' – 3')	Gene
oIK30	TGGCCTGACGCGTATGTTGT	*recC*
oIK31	TGCCCCACCAGTTCTGCAAT	*recC*
oIK32	CTCTTGCGGGTTACGGACGT	*recD*
oIK33	ATTCAGTCCAGCCACGCCAA	*recD*
oIK34	TCTGGCTTCATCGCTCGCAA	*ptrA*
oIK35	TGGGGCATGGGCTTCCTTTT	*ptrA*
oIK60	CACTGTGGCGCTTCCCAAAC	*halo*
oIK61	CCAGAACAGCAGCTTCGGGA	*halo*
oIK62	ACCCGCGCATTGGCTGAGAT	*recB*
oIK63	CACCGCTTTCGCTACGCAGC	*recB*
oIK64	CGGTGGTCCCACCTGACC	*rrfD*
oIK65	CCTACTCTCGCATGGGGAGACC	*rrfD*
oIK99	CGTCTCGCCCGGTTTCTCAT	*hfq*
oIK100	GGAAGTATTCTGCGCGCTGC	*hfq*
oIK101	GTCGCTTAAAGTGACGGCAT	*chiX*
oIK102	TCGCTATTGGCCCGTCAAAGA	*chiX*
oIK103	TCGGAATGCGTGTTCTGGTGA	*gcvB*
oIK104	TCGCAGACCAATTGCAAACACA	*gcvB*
oIK105	GCTAGCTGTACCAGGAACCACC	*cyaR*
oIK106	GGGAGATTACACAGGCTAAGGAGG	*cyaR*
oIK107	ATTTCCCTGGTGTTGGCGCA	*arcZ*
oIK108	TAGACCGGGGTGCGCGAATA	*arcZ*

Table 10.9: Sequences of RNA FISH probes for *recB* gene labelled with TAMRA dye.

Probe ID	**Sequence (5' – 3')**
RecB-TAMRA_1	ctgtaagggcaagcgcaaag
RecB-TAMRA_2	gcaatcgtaaaggttttgcc
RecB-TAMRA_3	tcgtggatattgctacggat
RecB-TAMRA_4	gttcggctaacaacaaccac
RecB-TAMRA_5	aaaggcattcaggttgagca
RecB-TAMRA_6	aatcagctgctgctcaaaca
RecB-TAMRA_7	aaagacgacctgggctattt
RecB-TAMRA_8	cgccttgcagataacgatta
RecB-TAMRA_9	ctgctgttttaccgtatcaa
RecB-TAMRA_10	accagaagattcgatcagcg
RecB-TAMRA_11	cggttaaacttgcgtcgatc
RecB-TAMRA_12	atcttgtcgatccatttagc
RecB-TAMRA_13	ccggcaactgataactgttt
RecB-TAMRA_14	cttcgtgcgatcttctaaga
RecB-TAMRA_15	ttgatcgatcgcctcaaaca
RecB-TAMRA_16	cagatcgcggatcgacaatg
RecB-TAMRA_17	agccgacttaacatgtcatc
RecB-TAMRA_18	cgccaattagcaacaatgcg
RecB-TAMRA_19	tacgcgccttcatataagtg
RecB-TAMRA_20	tgtctaaagtgtagtgggcg
RecB-TAMRA_21	aagcttattcacgctgttca
RecB-TAMRA_22	tcgcgaaacatgaacgcgtc
RecB-TAMRA_23	aaacggaagggatttccagc
RecB-TAMRA_24	cctgcaacaaccaaagcatt
RecB-TAMRA_25	cagatttgccgataaccatc
RecB-TAMRA_26	gttttcagcaatgttacgcg
RecB-TAMRA_27	ttgtagcagttcgctgatat
RecB-TAMRA_28	tggatcgtgacaatctgcac
RecB-TAMRA_29	tggacgcggaaattggtgat
RecB-TAMRA_30	atcgtgataaaacgcctgct
RecB-TAMRA_31	ttaagatccagaactgcctc
RecB-TAMRA_32	aagcaaacgcagatcttccg
RecB-TAMRA_33	aatgccaaaccgaacgtgtc
RecB-TAMRA_34	aacgcttcaatacaggtgcg
RecB-TAMRA_35	ttggttatcaccagtttgtg

Probe ID	Sequence (5' – 3')
RecB-TAMRA_36	tcagctctgctgtagaaaca
RecB-TAMRA_37	aaaccagagtagctggtgac
RecB-TAMRA_38	tgatgtggtgttaacgtcgg
RecB-TAMRA_39	tcaaccggctgggtaaaatc
RecB-TAMRA_40	ttattgcgggcggaaagttg
RecB-TAMRA_41	gataaaactccatctccacc
RecB-TAMRA_42	aacgtatcaagctgactggc
RecB-TAMRA_43	gccagggtataaagctgata
RecB-TAMRA_44	caatgcgatggcgcagataa
RecB-TAMRA_45	tggtgctcatagtcgtaatc
RecB-TAMRA_46	acagataaataacgccgcca
RecB-TAMRA_47	gatgttctttatcaacgcca
RecB-TAMRA_48	cataccggcaaacatctcat

Table 10.10: Sequences of RNA FISH probes for *recC* gene labelled with TAMRA dye.

Probe ID	**Sequence (5' – 3')**
RecC-TAMRA_1	ggaatggtagacccttaaca
RecC-TAMRA_2	actttgcaccagaatcatct
RecC-TAMRA_3	ccgcaataccaaacttttgc
RecC-TAMRA_4	gctggcagcggaaaatcaat
RecC-TAMRA_5	aacacccggacgaacatatc
RecC-TAMRA_6	tgttaaaggcgctctctttg
RecC-TAMRA_7	ataatgccgcaacagggtaa
RecC-TAMRA_8	tggaacagttttcgcttgtc
RecC-TAMRA_9	gactccagcgtttcgataaa
RecC-TAMRA_10	aaagacgcgcgaaggtaacc
RecC-TAMRA_11	aggcggtaacgcggaaatac
RecC-TAMRA_12	caggggttggtaaacaggag
RecC-TAMRA_13	gaaatagcggcaattcgcga
RecC-TAMRA_14	ctccaggtcagaaaggagat
RecC-TAMRA_15	tggcgtgacatcgacaaagg
RecC-TAMRA_16	cagttccagaatgtcagact
RecC-TAMRA_17	cgctacgggaaaactcttcg
RecC-TAMRA_18	gactgtagctgtcgatatca
RecC-TAMRA_19	acggtcggaaatggcgtaag
RecC-TAMRA_20	tgataaacgcttccagtacc
RecC-TAMRA_21	agcacatcctctgacacaaa
RecC-TAMRA_22	taaataacgcagcccttctt
RecC-TAMRA_23	aacgttgtcgtcatctatgc
RecC-TAMRA_24	cgctcgattcatcataaggt
RecC-TAMRA_25	cattagcagtgaagccagat
RecC-TAMRA_26	agaaggcgttgagcatatcg
RecC-TAMRA_27	cctaaaccttcggcgataat
RecC-TAMRA_28	tggctgatacgttcttgatc
RecC-TAMRA_29	gcattggcatcagagtacaa
RecC-TAMRA_30	agcaggcaaaccactttgaa
RecC-TAMRA_31	gcggaaattaacgcttccag
RecC-TAMRA_32	ggaacgaccgatatagctga
RecC-TAMRA_33	gggaaacgctcactgttatc
RecC-TAMRA_34	gtagataatgactttgcccg
RecC-TAMRA_35	tcgctttcatcacagttgag

Probe ID	Sequence (5' – 3')
RecC-TAMRA_36	cgggtatggaggcaagtaag
RecC-TAMRA_37	gttttgtggatcaaacggca
RecC-TAMRA_38	attcacgagcatagctttgt
RecC-TAMRA_39	gtttccggtaaggtaaacgg
RecC-TAMRA_40	acggaagttcacctgcaaac
RecC-TAMRA_41	gtccttccagaataaatggc
RecC-TAMRA_42	tttcaccaaaagcgccatac
RecC-TAMRA_43	aaggcgactttcaccattac
RecC-TAMRA_44	agaggtaatgcaaagcctgt
RecC-TAMRA_45	tgcaacgtggaatcgtcatc
RecC-TAMRA_46	gcctgaaggaatttcgtacg
RecC-TAMRA_47	agagcctttgataccagata
RecC-TAMRA_48	caggaaacgttgcgactgtt

Table 10.11: Sequences of RNA FISH probes for *recD* gene labelled with TAMRA dye.

Probe ID	**Sequence (5' – 3')**
RecD-TAMRA_1	gtcagggcaaattgcacatc
RecD-TAMRA_2	atcatgacttaacagtgccg
RecD-TAMRA_3	caaacaaacgtgtccctctc
RecD-TAMRA_4	cgttattttccagtcgtgaa
RecD-TAMRA_5	gatttcactgacacaggtcg
RecD-TAMRA_6	attcttcccaattttgtagc
RecD-TAMRA_7	aaagacgatcgccacagagg
RecD-TAMRA_8	ttacaccacatgcgattcaa
RecD-TAMRA_9	cttcgttgaaaaagcgtgcc
RecD-TAMRA_10	agcttcatcaacctcaatgg
RecD-TAMRA_11	ggaaaaagtttgtccagggt
RecD-TAMRA_12	cgcaacttttgccagttaa
RecD-TAMRA_13	tgaattaacgctgccagcaa
RecD-TAMRA_14	tgccgagagattcggttaag
RecD-TAMRA_15	cagcggtaactgtcgcaaag
RecD-TAMRA_16	cggaatgcgtttcttttgtt
RecD-TAMRA_17	caatcggtgcaaagtgctgg
RecD-TAMRA_18	gatgacgtaaacgctggcta
RecD-TAMRA_19	atcaagatgcagcgggttac
RecD-TAMRA_20	gatcgatcattgacgcttca
RecD-TAMRA_21	tcgatcagtctcgacatcat
RecD-TAMRA_22	gatcgccgagaaagatcact
RecD-TAMRA_23	tggcataagcgcagatatcg
RecD-TAMRA_24	gagactgtcgcgcaaagatg
RecD-TAMRA_25	cgatagctttttgcagcag
RecD-TAMRA_26	actgaccaatgccagaatcg
RecD-TAMRA_27	tttatcaccacggttgatcg
RecD-TAMRA_28	cctgctgaaaaacggttttc
RecD-TAMRA_29	gccgtttttcgatatcagta
RecD-TAMRA_30	aatataatcttcgccgctct
RecD-TAMRA_31	caagagcttcctcaagcatc
RecD-TAMRA_32	cagcagatccagataacgtc
RecD-TAMRA_33	cctgaatgattaaatccggc
RecD-TAMRA_34	cgcacaaaagctggtactca
RecD-TAMRA_35	ctcaattcgctcattcagtc

Probe ID	Sequence (5' – 3')
RecD-TAMRA_36	tgcgcttctgttgcataaac
RecD-TAMRA_37	aacgagagtgcggatgacga
RecD-TAMRA_38	gtcattacgggcaatcatca
RecD-TAMRA_39	cattaaacaacccaagcgcg
RecD-TAMRA_40	atccagcgcaataccgatat
RecD-TAMRA_41	actcggttgcacagacttaa
RecD-TAMRA_42	tagtttcgtgctctggcaag
RecD-TAMRA_43	tgcgatttatgtaccgtcat
RecD-TAMRA_44	aaaatcaacgccgcatggtc
RecD-TAMRA_45	ataaaccagctctcgcgtta
RecD-TAMRA_46	gcacttaatatgcgctcatc
RecD-TAMRA_47	ctcagtacgagtggcgattg
RecD-TAMRA_48	cgtgaactaaacaacgccgc

Table 10.12: *recBCD* fitting parameters with 95% confidence intervals (95% CI).

Parameter	Gene	Fitting function	Value	95% CI
λ	*recB*	Poisson	0.315	[0.291, 0.338]
p	*recB*	Negative binomial	0.513	[0.456, 0.570]
r	*recB*	Negative binomial	0.335	[0.267, 0.403]
λ	*recC*	Poisson	0.319	[0.282, 0.357]
p	*recC*	Negative binomial	0.498	[0.409, 0.586]
r	*recC*	Negative binomial	0.324	[0.222, 0.426]
λ	*recD*	Poisson	0.404	[0.353, 0.455]
p	*recD*	Negative binomial	0.568	[0.466, 0.671]
r	*recD*	Negative binomial	0.544	[0.335, 0.753]

Table 10.13: The Kullback–Leibler divergence (D_{KL}) scores between distributions.

Gene	**Distribution #1**	**Distribution #2**	D_{KL} **score**
recB	Experimental (newborns)	Poisson fit	0.0770
recB	Experimental (newborns)	negative binomial fit	0.0021
recB	Experimental (adults)	Poisson prediction	0.1635
recB	Experimental (adults)	negative binomial prediction	0.0079
recC	Experimental (newborns)	Poisson fit	0.0836
recC	Experimental (newborns)	negative binomial fit	0.0040
recC	Experimental (adults)	Poisson prediction	0.1379
recC	Experimental (adults)	negative binomial prediction	0.0115
recD	Experimental (newborns)	Poisson fit	0.0825
recD	Experimental (newborns)	negative binomial fit	0.0282
recD	Experimental (adults)	Poisson prediction	0.1516
recD	Experimental (adults)	negative binomial prediction	0.0308

Table 10.14: Bayesian (BIC) and Akaike (AIC) values for comparison of transcriptional models.

Gene	Criterion	Model	Value
recB	AIC	constitutive	3475
recB	AIC	bursty transcription	3149
recB	BIC	constitutive	3481
recB	BIC	bursty transcription	3160
recC	AIC	constitutive	1187
recC	AIC	bursty transcription	1066
recC	BIC	constitutive	1192
recC	BIC	bursty transcription	1075
recD	AIC	constitutive	851
recD	AIC	bursty transcription	798
recD	BIC	constitutive	855
recD	BIC	bursty transcription	807

Table 10.15: Parameters of the two-stage model of RecB expression.

Parameter	Meaning	Value	95% CI	Source
k_m	transcription rate	0.21 min^{-1}	[0.13 0.33] min^{-1}	Estimated
γ_m	mRNA degradation rate	0.62 min^{-1}	[0.48 0.75] min^{-1}	Measured
b	mRNA burst size	0.95 molec	[0.76 1.19] molec	Estimated
k_p	translation rate	0.15 min^{-1}	[0.12 0.18] min^{-1}	Estimated
γ_p	protein removal rate	0.015 min^{-1}	[0.011 0.019] min^{-1}	Measured

Bibliography

[1] URL: https://edbrownlab.shinyapps.io/brown_lab_srna_phenobrowser/.

[2] URL: https://gitlab.com/ikalita/cellsegmentation_mek.

[3] URL: https://code.google.com/archive/p/spatzcells/.

[4] URL: https://gitlab.com/ikalita/spotdetection_spatzcells.

[5] URL: https://www.mathworks.com/matlabcentral/fileexchange/20688-kullback-leibler-divergence.

[6] URL: https://github.com/nvictus/Gillespie.

[7] URL: https://git.ecdf.ed.ac.uk/sgrannem.

[8] URL: https://pypi.org/user/g_ronimo/.

[9] URL: https://salislab.net/software/predict_rbs_calculator.

[10] SE Ades, IL Grigorova and CA Gross. "Regulation of the alternative sigma factor sigma(E) during initiation, adaptation, and shutoff of the extracytoplasmic heat shock response in *Escherichia coli*". In: *J Bacteriol* 185.8 (2003), pp. 2512–9. DOI: 10.1128/JB.185.8.2512-2519.2003.

[11] MR Adhin and J van Duin. "Scanning model for translational reinitiation in eubacteria". In: *J Mol Biol* 213.4 (1990), pp. 811–8. DOI: 10.1016/S0022-2836(05)80265-7.

[12] T Ali Azam et al. "Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid". In: *J Bacteriol* 181.20 (1999), pp. 6361–70. DOI: 10.1128/JB.181.20.6361-6370.1999.

[13] U Alon. "Network motifs: theory and experimental approaches". In: *Nat Rev Genet* 8.6 (2007), pp. 450–61. DOI: 10.1038/nrg2102.

[14] S Altuvia et al. "A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator". In: *Cell* 90.1 (1997), pp. 43–53. DOI: 10.1016/s0092-8674(00)80312-8.

[15] S Amundsen et al. "Genetic Dissection of the Biochemical Activities of RecBCD Enzyme". In: *Genetics* 126.1 (1990), pp. 25–40.

[16] S Amundsen et al. "*recD*: the gene for an essential third subunit of Exonuclease V". In: *PNAS* 83.15 (1986), pp. 5558–5562. DOI: 10.1073/pnas.83.15.5558.

[17] T Andriuskevicius, O Kotenko and S Makovets. "Putting together and taking apart: assembly and disassembly of the Rad51 nucleoprotein filament in DNA repair and genome stability". In: *Cell Stress* 2.5 (2018), pp. 96–112. DOI: 10.15698/cst2018.05.134.

[18] V Arluison et al. "Spectroscopic observation of RNA chaperone activities of Hfq in post-transcriptional regulation by a small non-coding RNA". In: *Nucleic Acids Res* 35.3 (2007), pp. 999–1006. DOI: 10.1093/nar/gkl1124.

[19] D Arnold and S Kowalczykowski. “Facilitated Loading of RecA Protein is Essential to Recombination by RecBCD Enzyme”. In: *J Biol Chem* 275.16 (2000), pp. 12261–12265. DOI: 10.1074/jbc.275.16.12261.

[20] F Arsene, T Tomoyasu and B Bukau. “The heat shock response of *Escherichia coli*”. In: *Int J Food Microbiol* 55.1-3 (2000), pp. 3–9. DOI: 10.1016/s0168-1605(00)00206-3.

[21] MS Azam and CK Vanderpool. “Translational regulation by bacterial small RNAs via an unusual Hfq-dependent mechanism”. In: *Nucleic Acids Res* 46.5 (2018), pp. 2585–99. DOI: 10.1093/nar/gkx1286.

[22] B Azeroglu. “DNA Synthesis during Double-Strand Break Repair in *Escherichia coli*”. PhD thesis. 2014.

[23] W Bae et al. “*Escherichia coli* CspA-family RNA Chaperones are Transcription Antiterminators”. In: *PNAS* 97.14 (2000), pp. 7784–9. DOI: 10.1073/pnas.97.14.7784.

[24] E Bakker and PS Swain. “Estimating numbers of intracellular molecules through analysing fluctuations in photobleaching”. In: *Sci Rep* 9.1 (2019), p. 15238. DOI: 10.1038/s41598-019-50921-7.

[25] N Balaban et al. “Bacterial Persistence as a Phenotypic Switch”. In: *Science* 305.5690 (2004), pp. 1622–1625. DOI: 10.1126/science.1099390.

[26] B Barreto et al. “The Small RNA GcvB Promotes Mutagenic Break Repair by Opposing the Membrane Stress Response”. In: *J Bacteriol* 198.24 (2016), pp. 3296–308. DOI: 10.1128/JB.00555-16.

[27] A Battesti, N Majdalani and S Gottesman. “The RpoS-mediated general stress response in *Escherichia coli*”. In: *Annu Rev Microbiol* 65 (2011), pp. 189–213. DOI: 10.1146/annurev-micro-090110-102946.

[28] C Beentjes, R Perez-Carrasco and R Grima. “Exact solution of stochastic gene expression models with bursting, cell cycle and replication dynamics”. In: *Phys Rev E* 101.3 (2020), p. 032403. DOI: 10.1103/PhysRevE.101.032403.

[29] Otto G Berg. “A model for the statistical fluctuations of protein numbers in a microbial population”. In: *J Theor Biol* 71.4 (1978), pp. 587–603. DOI: 10.1016/0022-5193(78)90326-0.

[30] J Bernstein et al. “Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays”. In: *PNAS* 99.15 (2002), pp. 9697–9702. DOI: 10.1073/pnas.112318199.

[31] P Bianco et al. “Processive translocation and DNA unwinding by individual RecBCD enzyme molecules”. In: *Nature* 409 (2001), pp. 374–378. DOI: 10.1038/35053131.

[32] I Bilusic et al. “Revisiting the coding potential of the *E. coli* genome through Hfq co-immunoprecipitation”. In: *RNA Biol* 11.5 (2014), pp. 641–54. DOI: 10.4161/rna.29299.

[33] WJ Blake et al. “Phenotypic consequences of promoter-mediated transcriptional noise”. In: *Mol Cell* 24.6 (2006), pp. 853–65. DOI: 10.1016/j.molcel.2006.11.003.

[34] Small RNA induced mRNA degradation achieved through both translation block and activated cleavage. “Prévost, K and Desnoyers, G and Jacques, JF and Lavoie, F and Massé, E”. In: *Genes Dev* 25.4 (2011), pp. 385–96. DOI: 10.1101/gad.2001711.

[35] P Bokes, M Hojcka and A Singh. “MicroRNA Based Feedforward Control of Intrinsic Gene Expression Noise”. In: *IEEE* 18.1 (2021), pp. 272–282. DOI: 10.1109/TCBB.2019.2938502.

[36] IR Booth and P Louis. “Managing hypoosmotic stress: aquaporins and mechanosensitive channels in *Escherichia coli*”. In: *Curr. Opin. Microbiol.* 2.2 (1999), pp. 166–169. DOI: 10.1016/S1369-5274(99)80029-0.

[37] CA Brackley et al. “Stochastic Model of Supercoiling-Dependent Transcription”. In: *Phys Rev Lett* 117.1 (2016), p. 018101. DOI: 10.1103/PhysRevLett.117.018101.

[38] H Bremer and P Dennis. “Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates”. In: *EcoSal Plus* 3.1 (2008). DOI: 10.1128/ecosal.5.2.3.

[39] J Bryant et al. “Chromosome position effect on gene expression in *Escherichia coli* K12”. In: *NAR* 42.18 (2014), pp. 11383–92. DOI: 10.1093/nar/gku828.

[40] R Buda et al. “Dynamics of *Escherichia coli*’s passive response to a sudden decrease in external osmolarity”. In: *PNAS* 113.40 (2016), E5838–E5846. DOI: 10.1073/pnas.1522185113.

[41] Z Cao et al. “A Stochastic Model of Gene Expression with Polymerase Recruitment and Pause Release”. In: *Biophys J* 119.5 (2020), pp. 1002–14. DOI: 10.1016/j.bpj.2020.07.020.

[42] Y Chao and J Vogel. “The role of Hfq in bacterial pathogens”. In: *Curr Opin Microbiol* 13.1 (2010), pp. 24–33. DOI: 10.1016/j.mib.2010.01.001.

[43] Y Chao et al. “An atlas of Hfq-bound transcripts reveals 3’ UTRs as a genomic reservoir of regulatory small RNAs”. In: *EMBO J* 31.20 (2012), pp. 4005–19. DOI: 10.1038/emboj.2012.229.

[44] H Chen et al. “Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*”. In: *Mol Syst Biol* 11.781 (2015). DOI: https://doi.org/10.15252/msb.20145794.

[45] J Chen and S Gottesman. “Hfq links translation repression to stress-induced mutagenesis in *E. coli*”. In: *Genes Dev* 31.13 (2017), pp. 1382–95. DOI: 10.1101/gad.302547.117.

[46] SM Chiang and HE Schellhorn. “Regulators of oxidative stress response genes in *Escherichia coli* and their functional conservation in bacteria”. In: *Archives of Biochemistry and Biophysics* 525.2 (2012), 161–169. DOI: 10.1016/j.abb.2012.02.007.

[47] N Christopoulou and S Granneman. “The role of RNA-binding proteins in mediating adaptive responses in Gram-positive bacteria”. In: *FEBS J* (2021). DOI: 10.1111/febs.15810.

[48] F Claverie-Martin, M Diaz-Torres and S Kushner. “Analysis of the regulatory region of the protease III (*ptrA*) gene of *Escherichia coli* K-12”. In: *Gene* 54 (1987), pp. 185–95. DOI: 10.1016/0378-1119(87)90486-0.

[49] G Cromie. “Phylogenetic Ubiquity and Shuffling of the Bacterial RecBCD and AddAB Recombination Complexes”. In: *J Bacteriol* 191.16 (2009), pp. 5076–84. DOI: 10.1128/JB.00254-09.

[50] GA Cromie, JC Connelly and DR Leach. “Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans”. In: *Mol Cell* 8.6 (2001), pp. 1163–74. DOI: 10.1016/s1097-2765(01)00419-1.

[51] E Darmon et al. “*E. coli* SbcCD and RecA Control Chromosomal Rearrangement Induced by an Interrupted Palindrome”. In: *Mol Cell* 39.1-2 (2010), pp. 59–70. DOI: 10.1016/j.molcel.2010.06.011.

[52] K Datsenko and B Wanner. “One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products”. In: *PNAS* 97.12 (2000), pp. 6640–6645. DOI: 10.1073/pnas.120163297.

[53] D De Biase and PA Lund. “The *Escherichia coli* Acid Stress Response and Its Significance for Pathogenesis”. In: *Adv Appl Microbiol* 92 (2015), pp. 49–88. DOI: 10.1016/bs.aambs.2015.03.002.

[54] A Delhaye, JF Collet and G Laloux. “Fine-tuning of the Cpx envelope stress response is required for cell wall homeostasis in *Escherichia coli*”. In: *mBio* 7.1 (2016), e00047–16. DOI: 10.1128/mBio.00047-16.

[55] D Dermic et al. “RecBCD enzyme overproduction impairs DNA repair and homologous recombination in *Escherichia coli*”. In: *Res Microbiol* 156.3 (2005), pp. 304–11. DOI: 10.1016/j.resmic.2004.10.005.

[56] R Dickman and R Vidigal. “Path Integrals and Perturbation Theory for Stochastic Processes”. In: *Braz. J. Phys* 33.1 (2003).

[57] M Dillingham and S Kowalczykowski. “RecBCD Enzyme and the Repair of Double-Stranded DNA Breaks”. In: *Microbiol Mol Biol Rev* 72.4 (2008), pp. 642–671. DOI: 10.1128/MMBR.00020-08.

[58] M Dillingham, M Spies and S Kowalczykowski. “RecBCD enzyme is a bipolar DNA helicase”. In: *Nature* 423 (2003). DOI: 10.1038/nature01673.

[59] M Doherty et al. “Turnover of the human proteome: determination of protein intracellular stability by dynamic SILAC”. In: *J Proteome Res* 8.1 (2009), pp. 104–12. DOI: 10.1021/pr800641v.

[60] C Dykstra and S Kushner. “Physical characterization of the cloned protease III gene from *Escherichia coli* K-12”. In: *J Bacteriol* 163 (1985), pp. 1055–59.

[61] C Dykstra, D Prasher and S Kushner. “Physical and biochemical analysis of the cloned *recB* and *recC* genes of *Escherichia coli* K-12”. In: *J Bacteriol* 157 (1984), pp. 21–7. DOI: 10.1128/jb.157.1.21-27.1984.

[62] Sauer E. “Structure and RNA-binding properties of the bacterial LSm protein Hfq”. In: *RNA Biol* 10.4 (2013), pp. 610–618. DOI: 10.4161/rna.24201.

[63] M Eldeeb et al. “A molecular toolbox for studying protein degradation in mammalian cells”. In: *J Neurochem* 151.4 (2019), pp. 520–533. DOI: 10.1111/jnc.14838.

[64] J Elf and M Ehrenberg. “Fast evaluation of fluctuations in biochemical networks with the linear noise approximation”. In: *Genome Res.* 13.11 (2003), pp. 2475–2484. DOI: 10.1101/gr.1196503.

[65] M Ellis, R Trussler and D Haniford. “Hfq binds directly to the ribosome-binding site of IS10 transposase mRNA to inhibit translation”. In: *Mol Microbiol* 96.3 (2015), pp. 633–50. DOI: 10.1111/mmi.12961.

[66] M Elowitz et al. “Stochastic Gene Expression in a Single Cell”. In: *Science* 297.5584 (2002), pp. 1183–1186. DOI: 10.1126/science.1070919.

[67] P Emmerson. “Recombination deficient mutants of *Escherichia coli* K12 that map between *thyA* and *argA*”. In: *Genetics* 60.1 (1968), pp. 19–30. DOI: 10.1093/genetics/60.1.19.

[68] J Engel and M Zijlstra. “A Characterization of the Gamma Distribution by the Negative Binomial Distribution.” In: *J Appl Probab* 17.4 (1980), 1138–44. DOI: 10.2307/3213228.

[69] I Erill, S Campoy and J Barbé. “Aeons of distress: an evolutionary perspective on the bacterial SOS response”. In: *FEMS Microbiol Rev* 31.5 (2007), pp. 637–56. DOI: 0.1111/j.1574-6976.2007.00082.x.

[70] R Faigenbaum-Romm et al. “Hierarchy in Hfq Chaperon Occupancy of Small RNA Targets Plays a Major Role in Their Regulation”. In: *Cell Rep* 30.9 (2020), pp. 3127–3138. DOI: 10.1016/j.celrep.2020.02.016.

[71] M Fasnacht and N Polacek. “Oxidative stress in bacteria and the Central Dogma of Molecular Biology”. In: *Front Mol Biosci* 8.671037 (2021). DOI: 10.3389/fmolb.2021.671037.

[72] M Franze de Fernandez, L Eoyange and J August. “Factor Fraction required for the Synthesis of Bacteriophage Qβ-RNA”. In: *Nature* 219 (1968), 588–590. DOI: 10.1038/219588a0.

[73] T Filatova, N Popovic and R Grima. “Statistics of Nascent and Mature RNA Fluctuations in a Stochastic Model of Transcriptional Initiation, Elongation, Pausing, and Termination”. In: *Bull Math Biol* 83.3 (2021). DOI: 10.1007/s11538-020-00827-7.

[74] P Finch et al. “Complete nucleotide sequence of the *Escherichia coli recB* gene”. In: *Nucleic Acids Res* 14.21 (1986), pp. 8573–82.

[75] M Folichon et al. “The poly(A) binding protein Hfq protects RNA from RNase E and exoribonucleolytic degradation”. In: *Nucleic Acids Res* 31.24 (2003), pp. 7302–10. DOI: 10.1093/nar/gkg915.

[76] N Friedman, L Cai and XS Xie. “Linking stochastic dynamics to population distribution: an analytical framework of gene expression”. In: *Phys Rev Lett* 97.16 (2006), p. 168302. DOI: 10.1103/PhysRevLett.97.168302.

[77] CW Gardiner. *Handbook of stochastic methods for physics, chemistry and the natural sciences*. 3rd ed. Berlin/New York: Springer-Verlag, 2004.

[78] D Gillespie. “Exact stochastic simulation of coupled chemical reactions”. In: *J. Phys. Chem.* 81.25 (1977), 2340–2361. DOI: 10.1021/j100540a008.

[79] I Golding. “Infection by bacteriophage lambda: an evolving paradigm for cellular individuality”. In: *Curr Opin Microbiol* 43 (2018), pp. 9–13. DOI: 10.1016/j.mib.2017.09.014.

[80] I Golding. “Revisiting Replication-Induced Transcription in *Escherichia coli*”. In: *Bioessays* 42.1 (2020), e1900193. DOI: 10.1002/bies.201900193.

[81] I Golding et al. “Real-Time Kinetics of Gene Activity in Individual Bacteria”. In: *Cell* 123.6 (2005), pp. 1025–36. DOI: 10.1016/j.cell.2005.09.031.

[82] J Gorden and PL Small. “Acid resistance in enteric bacteria”. In: *Infect Immun* 61.1 (1993), pp. 364–7. DOI: 10.1128/iai.61.1.364-367.1993.

[83] S Granneman et al. “Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs”. In: *PNAS* 106.24 (2009), pp. 9613–8. DOI: 10.1073/pnas.0901997106.

[84] P Guptasarma. “Does replication-induced transcription regulate synthesis of the myriad low copy number proteins of *Escherichia coli*?” In: *Bioessays* 17.11 (1995), pp. 987–97. DOI: 10.1002/bies.950171112.

[85] L Guzman et al. “Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter”. In: *J Bacteriol* 177.14 (1995), pp. 4121–30. DOI: 10.1128/jb.177.14.4121-4130.1995.

[86] K Han, B Tjaden and S Lory. “GRIL-seq provides a method for identifying direct targets of bacterial small regulatory RNA by *in vivo* proximity ligation”. In: *Nat. Microbiol.* 2.16239 (2016). DOI: 10.1038/nmicrobiol.2016.239.

[87] N Handa et al. “$\chi*$, a χ-related 11-mer sequence partially active in an *E. coli recC** strain”. In: *Genes to Cells* 2.8 (1997), pp. 525–536. DOI: 10.1046/j.1365-2443.1997.1410339.x.

[88] M Hashimoto et al. “Noise-driven growth rate gain in clonal cellular populations”. In: *PNAS* 113.12 (2016), pp. 3251–6. DOI: 10.1073/pnas.1519412113.

[89] D Hawley and W McClure. “Compilation and analysis of *Escherichia coli* promoter DNA sequences”. In: *Nucl Acids Res* 11 (1983), pp. 2237–55. DOI: 10.1093/nar/11.8.2237.

[90] I Hickson et al. “Molecular amplification and purification of the *E. coli recC* gene product”. In: *Nucl Acids Res* 12 (1984), pp. 3807–19. DOI: 10.1093/nar/12.9.3807.

[91] I Hickson et al. “Reconstitution of RecBC DNase activity from purified *Escherichia coli* RecB and RecC proteins”. In: *J Biol Chem* 260.2 (1985), pp. 1224–29. DOI: 10.1016/S0021-9258(20)71232-9.

[92] A Hilfinger, M Thomas and J Paulsson. “Exploiting Natural Fluctuations to Identify Kinetic Mechanisms in Sparsely Characterized Systems”. In: *Cell Systems* 2.4 (2016), pp. 251–259. DOI: 10.1016/j.cels.2016.04.002.

[93] TM Hill et al. “sfi-independent filamentation in *Escherichia coli* is lexA dependent and requires DNA damage for induction”. In: *J Bacteriol* 179.6 (1997), pp. 1931–9. DOI: 10.1128/jb.179.6.1931-1939.1997.

[94] J Holehouse, Z Cao and R Grima. “Stochastic Modeling of Autoregulatory Genetic Feedback Loops: A Review and Comparative Study”. In: *Biophys J* 118.7 (2020), pp. 1517–25. DOI: 10.1016/j.bpj.2020.02.016.

[95] E Holmqvist, S Berggren and A Rizvanovic. “RNA-binding activity and regulatory functions of the emerging sRNA-binding protein ProQ”. In: *Biochim Biophys Acta Gene Regul Mech* 1863.9 (2020), p. 194596. DOI: 10.1016/j.bbagrm.2020.194596.

[96] E Holmqvist et al. “Global Maps of ProQ Binding *In Vivo* Reveal Target Recognition via RNA Structure and Stability Control at mRNA 3 Ends”. In: *Mol Cell* 70.5 (2018), pp. 971–82. DOI: 10.1016/j.molcel.2018.04.017.

[97] E Holmqvist et al. “Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo”. In: *EMBO J* 35.9 (2016), pp. 991–1011. DOI: 10.15252/embj.201593360.

[98] D Huh and J Paulsson. “Non-genetic heterogeneity from stochastic partitioning at cell division”. In: *Nat Genet* 43.2 (2011), pp. 95–100. DOI: 10.1038/ng.729.

[99] I Iosub. “The role of small regulatory RNA networks in controlling adaptive responses in *Escherichia coli*”. PhD thesis. 2017.

[100] I Iosub et al. “Hfq CLASH uncovers sRNA-target interaction networks enhancing adaptation to nutrient availability”. In: *eLife* 9 (2020). DOI: 10.7554/eLife.54655.

[101] I Iosub et al. “The mRNA derived MalH sRNA contributes to alternative carbon source utilization by tuning maltoporin expression in *E. coli*”. In: *RNA Biology* 18.6 (2021), pp. 914–31. DOI: 10.1080/15476286.2020.1827784.

[102] DW Jackson et al. “Biofilm formation and dispersal under the influence of the global regulator CsrA of *Escherichia coli*”. In: *J Bacteriol* 184.1 (2002), pp. 290–301. DOI: 10.1128/JB.184.1.290-301.2002.

[103] C Janion. “Inducible SOS response system of DNA repair and mutagenesis in *Escherichia coli*”. In: *Int. J. Biol. Sci* 4.6 (2008), pp. 338–44. DOI: 10.7150/ijbs.4.338.

[104] S Jaramillo-Riveri et al. “Growth-dependent heterogeneity in the DNA damage response in *Escherichia coli*”. In: *bioRxiv* (2021). DOI: 10.1101/2021.05.06.442821.

[105] D Jones and J Elf. “Bursting onto the scene? Exploring stochastic mRNA production in bacteria”. In: *Curr Opin Microbiol* 45 (2018), pp. 124–130. DOI: 10.1016/j.mib.2018.04.001.

[106] M Kajitani et al. “Regulation of the *Escherichia coli hfq* gene encoding the host factor for phage Q beta”. In: *J Bacteriol* 176.2 (1994), pp. 531–4. DOI: 10.1128/jb.176.2.531-534.1994.

[107] K Kavita, F Mets and S Gottesman. “New aspects of RNA-based regulation by Hfq and its partner sRNAs”. In: *Curr Opin Microbiol* 42 (2018), pp. 53–61. DOI: 10.1016/j.mib.2017.10.014.

[108] C Kelly et al. “Synthetic negative feedback circuits using engineered small RNAs”. In: *NAR* 46.18 (2018), 9875–9889. DOI: 10.1093/nar/gky828.

[109] T Kepler and T Elston. “Stochasticity in Transcriptional Regulation: Origins, Consequences, and Mathematical Representations”. In: *Biophys J* 81.6 (2001), pp. 3116–3136. DOI: 10.1016/S0006-3495(01)75949-8.

[110] A Khlebnikov et al. “Homogeneous expression of the PBAD promoter in *Escherichia coli* by constitutive expression of the low-affinity high-capacity AraE transporter”. In: *Microbiology* 147.12 (2001), pp. 3241–3247. DOI: 10.1099/00221287-147-12-3241.

[111] AM Kierzek, J Zaim and P Zielenkiewicz. “The effect of transcription and translation initiation frequencies on the stochastic fluctuations in prokaryotic gene expression”. In: *J Biol Chem* 276.11 (2001), pp. 8165–72. DOI: 10.1074/jbc.M006264200.

[112] AL Koch and HR Levy. “Protein turnover in growing cultures of *Escherichia coli*”. In: *J Biol Chem* 217.2 (1955), pp. 947–57.

[113] MA Kohanski, DJ Dwyer and JJ Collins. “How antibiotics kill bacteria: from targets to networks”. In: *Nat Rev Microbiol* 8.6 (2010), pp. 423–435. DOI: 10.1038/nrmicro2333.

[114] S Kosuri et al. “Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*”. In: *PNAS* 110.34 (2013), pp. 14024–9. DOI: 10.1073/pnas.1301301110.

[115] KN Kreuzer. “Interplay between DNA replication and recombination in prokaryotes”. In: *Annu Rev Microbiol* 59 (2005), pp. 43–67. DOI: 10.1146/annurev.micro.59.030804.121255.

[116] N Kumar, A Singh and RV Kulkarni. “Transcriptional Bursting in Gene Expression: Analytical Results for General Stochastic Models”. In: *PLoS Comput. Biol.* 11.10 (2015). DOI: 10.1371/journal.pcbi.1004292.

[117] A Kuzminov. “Collapse and repair of replication forks in *Escherichia coli*”. In: *Mol Microbiol* 16.3 (1995), pp. 373–84. DOI: 10.1093/nar/gky463.

[118] J Le Derout et al. “Hfq affects the length and the frequency of short oligo(A) tails at the 3’ end of *Escherichia coli rpsO* mRNAs”. In: *Nucleic Acids Res* 31.14 (2003), 4017–23. DOI: 10.1093/nar/gkg456.

[119] A Lepore et al. “Quantification of very low-abundant proteins in bacteria using the HaloTag and epi-fluorescence microscopy”. In: *Sci Rep* 9.7902 (2019). DOI: 10.1038/s41598-019-44278-0.

[120] 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.

[121] GM Li. “Mechanisms and functions of DNA mismatch repair”. In: *Cell Res* 18.1 (2008), pp. 85–98. DOI: 10.1038/cr.2007.115.

[122] W Li et al. “Systematically integrated metabonomic-proteomic studies of *Escherichia coli* under ciprofloxacin stress”. In: *J Proteomics* 179 (2018), pp. 61–70. DOI: 10.1016/j.jprot.2018.03.002.

[123] J Lin and A Amir. “Homeostasis of protein and mRNA concentrations in growing cells”. In: *Nat Commun* 9.1 (2018), p. 4496. DOI: 10.1038/s41467-018-06714-z.

[124] A Link, D Phillips and G Church. “Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli*: application to open reading frame characterization”. In: *J Bacteriol* 179.20 (1997), pp. 6228–37. DOI: 10.1128/jb.179.20.6228-6237.1997.

[125] T Link, P Valentin-Hansen and R Brennan. “Structure of *Escherichia coli* Hfq bound to polyriboadenylate RNA”. In: *PNAS* 106.46 (2009), pp. 19292–19297. DOI: 10.1073/pnas.0908744106.

[126] B Lipinska et al. “Cloning and identification of bacteriophage T4 gene 2 product gp2 and action of gp2 on infecting DNA *in vivo*”. In: *J Bacteriol* 171.1 (1989), pp. 488–97.

[127] JW Little and DW Mount. “The SOS regulatory system of *Escherichia coli*”. In: *Cell* 29.1 (1982), pp. 11–22. DOI: 10.1016/0092-8674(82)90085-x.

[128] S Little et al. “Translational coupling in the threonine operon of *Escherichia coli* K-12”. In: *J Bacteriol* 171.6 (1989), pp. 3518–22. DOI: 10.1128/jb.171.6.3518-3522.1989.

[129] B Liu, R Baskin and S Kowalczykowski. “DNA unwinding heterogeneity by RecBCD results from static molecules able to equilibrate”. In: *Nature* 500 (2013), 482–485. DOI: 10.1038/nature12333.

[130] L Løvdok et al. “Role of translational coupling in robustness of bacterial chemotaxis pathway”. In: *PLoS Biol* 8 (2009), e1000171. DOI: 10.1371/journal.pbio.1000171.

[131] H Maamar, A Raj and D Dubnau. “Noise in gene expression determines cell fate in *Bacillus subtilis*”. In: *Science* 317.5837 (2007), pp. 526–9. DOI: 10.1126/science.1140818.

[132] R Maharjan et al. “The basis of antagonistic pleiotropy in *hfq* mutations that have opposite effects on fitness at slow and fast growth rates”. In: *Heredity* 110 (2013), pp. 10–18. DOI: 10.1038/hdy.2012.46.

[133] P Mandin and S Gottesman. “A genetic approach for finding small RNAs regulators of genes of interest identifies RybC as regulating the DpiA/DpiB two-component system”. In: *Mol Microbiol* 72.3 (2009), pp. 551–65. DOI: 10.1111/j.1365-2958.2009.06665.

[134] P Mandin and S Gottesman. “Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA”. In: *EMBO J* 29.18 (2010), pp. 3094–107. DOI: 10.1038/emboj.2010.179.

[135] KH Maslowska2019, K Makiela-Dzbenska and IJ Fijalkowska. “The SOS system: A complex and tightly regulated response to DNA damage”. In: *Environ Mol Mutagen* 60.4 (2019), pp. 368–84. DOI: 10.1002/em.22267.

[136] E Massé, FE Escorcia and S Gottesman. “Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*”. In: *Genes Dev* 17.19 (2003), pp. 2374–83. DOI: 10.1101/gad.1127103.

[137] E Małecka et al. “Structure of bacterial regulatory RNAs determines their performance in competition for the chaperone protein Hfq”. In: *Biochemistry* 54.5 (2015), pp. 1157–70. DOI: 10.1021/bi500741d.

[138] HH McAdams and Arkin A. “Stochastic mechanisms in gene expression”. In: *PNAS* 94.3 (1997), pp. 814–9. DOI: 10.1073/pnas.94.3.814.

[139] JE McCarthy. “Post-transcriptional control in the polycistronic operon environment: studies of the *atp* operon of *Escherichia coli*”. In: *Mol Microbiol* 4.8 (1990), pp. 1233–40. DOI: 10.1111/j.1365-2958.1990.tb00702.x.

[140] A Mehta and J Haber. “Sources of DNA Double-Strand Breaks and Models of Recombinational DNA Repair”. In: *Cold Spring Harb Perspect Biol* 6.9 (2014). DOI: 10.1101/cshperspect.a016428.

[141] P Mehta, S Goyal and NS 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.

[142] S Melamed et al. “Global mapping of small RNA-target interactions in bacteria”. In: *Mol Cell* 63 (5 2016), pp. 884–897. DOI: 10.1016/j.molcel.2016.07.026.

[143] S Melamed et al. “RNA-RNA Interactomes of ProQ and Hfq Reveal Overlapping and Competing Roles”. In: *Mol Cell* 77.2 (2020), pp. 411–25. DOI: 10.1016/j.molcel.2019.10.022.

[144] R Merrill et al. “A robust and economical pulse-chase protocol to measure the turnover of HaloTag fusion proteins”. In: *J Biol Chem* 294.44 (2019), pp. 16164–16171. DOI: 10.1074/jbc.RA119.010596.

[145] E Metzl-Raz et al. “Principles of cellular resource allocation revealed by condition-dependent proteome profiling”. In: *Elife* 6.e28034 (2017). DOI: 10.7554/eLife.28034.

[146] AS Meyer and TA Baker. “Proteolysis in the *Escherichia coli* heat shock response: a player at many levels”. In: *Curr Opin Microbiol* 14.2 (2011), pp. 194–9. DOI: 10.1016/j.mib.2011.02.001.

[147] B Michel and D Leach. “Homologous Recombination-Enzymes and Pathways”. In: *EcoSal Plus* 5.1 (2012). DOI: 10.1128/ecosalplus.7.2.7.

[148] M Mir et al. “Optical measurement of cycle-dependent cell growth”. In: *PNAS* 108.32 (2011), pp. 13124–9. DOI: 10.1073/pnas.1100506108.

[149] S Modi et al. “Functional characterization of bacterial sRNAs using a network biology approach”. In: *PNAS* 108.37 (2011), pp. 15522–15527. DOI: 10.1073/pnas.1104318108.

[150] P Modrich and R Lahue. “Mismatch repair in replication fidelity, genetic recombination, and cancer biology”. In: *Annu Rev Biochem* 65 (1996), pp. 101–33. DOI: 10.1146/annurev.bi.65.070196.000533.

[151] J Moffitt et al. “Spatial organization shapes the turnover of a bacterial transcriptome”. In: *eLife* 5 (2016). DOI: 10.7554/eLife.13065.

[152] A Mogk, D Huber and B Bukau. “Integrating protein homeostasis strategies in prokaryotes”. In: *Cold Spring Harb Perspect Biol* 3.4 (2011), a004366. DOI: 10.1101/cshperspect.a004366.

[153] I Moll et al. “Coincident Hfq binding and RNase E cleavage sites on mRNA and small regulatory RNAs”. In: *RNA* 9.11 (2003), pp. 1308–14. DOI: 10.1261/rna.5850703.

[154] GF Moolenaar, C Moorman and N Goosen. “Role of the *Escherichia coli* Nucleotide Excision Repair Proteins in DNA Replication”. In: *J Bacteriol* 182.20 (2000). DOI: 10.1128/JB.182.20.5706-5714.2000.

[155] K Moon and S Gottesman. “Competition among Hfq-binding small RNAs in *Escherichia coli*”. In: *Mol Microbiol* 82.6 (2011), pp. 1545–62. DOI: 10.1111/j.1365-2958.2011.07907.x.

[156] K Morimatsu and SC Kowalczykowski. “RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: a universal step of recombinational repair”. In: *Mol Cell* 11.5 (2003), pp. 1337–47. DOI: 10.1016/s1097-2765(03)00188-6.

[157] T Morita and H Aiba. “Mechanism and physiological significance of autoregulation of the *Escherichia coli hfq* gene”. In: *RNA* 25.2 (2019), pp. 264–76. DOI: 10.1261/rna.068106.118.

[158] R Mosteller, J Rose and C Yanofsky. “Transcription Initiation and Degradation of *trp* mRNA”. In: *Cold Spring Harb Symp Quant Biol* 35 (1970), pp. 461–6. DOI: 10.1101/SQB.1970.035.01.060.

[159] A Muffler et al. “The RNA-binding protein HF-I plays a global regulatory role which is largely, but not exclusively, due to its role in expression of the sigmaS subunit of RNA polymerase in *Escherichia coli*”. In: *J Bacteriol* 179.1 (1997), pp. 297–300. DOI: 10.1128/jb.179.1.297-300.1997.

[160] T Møller et al. “Hfq: A Bacterial Sm-like Protein that Mediates RNA-RNA Interaction”. In: *Mol Cell* 9.1 (2002), pp. 23–30. DOI: 10.1016/S1097-2765(01)00436-1.

[161] H Nagai, H Yuzawa and T Yura. “Regulation of the heat shock response in *E. coli*: involvement of positive and negative cis-acting elements in translation control of σ^{32} synthesis”. In: *Biochimie* 73.12 (1991), pp. 1473–9. DOI: 10.1016/0300-9084(91)90180-9.

[162] T Nogueira and M Springer. “Post-transcriptional control by global regulators of gene expression in bacteria”. In: *Curr Opin Microbiol* 3.2 (2000), pp. 154–8. DOI: 10.1016/s1369-5274(00)00068-0.

[163] R van Nues et al. “Kinetic CRAC uncovers a role for Nab3 in determining gene expression profiles during stress”. In: *Nat Commun* 8.12 (2017). DOI: https://doi.org/10.1038/s41467-017-00025-5.

[164] B Okumus et al. “Mechanical slowing-down of cytoplasmic diffusion allows *in vivo* counting of proteins in individual cells”. In: *Nat Commun* 7.11641 (2016). DOI: 10.1038/ncomms11641.

[165] D Oppenheim and C Yanofsky. “Translational coupling during expression of the tryptophan operon of *Escherichia coli*”. In: *Genetics* 95.4 (1980), pp. 785–95. DOI: 10.1093/genetics/95.4.785.

[166] EM Ozbudak et al. “Regulation of noise in the expression of a single gene”. In: *Nat Genet* 31.1 (2002), pp. 69–73. DOI: 10.1038/ng869.

[167] J Paulsson. “Models of stochastic gene expression”. In: *Phys Life Rev* 2.2 (2005), pp. 157 –175. DOI: 10.1016/j.plrev.2005.03.003.

[168] J Paulsson. “Summing up the noise in gene networks”. In: *Nature* 427 (2004), pp. 415–8. DOI: 10.1038/nature02257.

[169] J Paulsson and M Ehrenberg. “Random Signal Fluctuations can Reduce Random Fluctuations in Regulated Components of Chemical Regulatory Networks”. In: *Phys Rev Lett* 84 (23 2000), pp. 5447–5450. DOI: 10.1103/PhysRevLett.84.5447.

[170] J Peccoud and B Ycart. “Markovian Modeling of Gene-Product Synthesis”. In: *Theor Popul Biol* 48.2 (1995), pp. 222–234. DOI: 10.1006/tpbi.1995.1027.

[171] JM Pedraza and J Paulsson. “Effects of molecular memory and bursting on fluctuations in gene expression”. In: *Science* 319.5861 (2008), pp. 339–43. DOI: 10.1126/science.1144331.

[172] L Peliti. “Path integral approach to birth-death processes on a lattice”. In: *Journal de Physique* 46.9 (1985), pp. 1469–83. DOI: 10.1051/jphys:019850046090146900.

[173] R Perez-Carrasco, C Beentjes and R Grima. “Effects of cell cycle variability on lineage and population measurements of messenger RNA abundance”. In: *J. R. Soc. Interface* 17.168 (2020). DOI: 10.1098/rsif.2020.0360.

[174] J Peterson et al. “Effects of DNA replication on mRNA noise”. In: *PNAS* 112.52 (2015), pp. 15886–15891. DOI: 10.1073/pnas.1516246112.

[175] M Pigliucci, CJ Murren and CD Schlichting. “Phenotypic plasticity and evolution by genetic assimilation”. In: *J Exp Biol* 209.Pt 12 (2006), pp. 2362–7. DOI: 10.1242/jeb.02070.

[176] T Pilizota and JW Shaevitz. “Fast, Multiphase Volume Adaptation to Hyperosmotic Shock by *Escherichia coli*”. In: *PLOS ONE* 7.4 (2012), e35205. DOI: 10.1371/journal.pone.0035205.

[177] T Pilizota and JW Shaevitz. “Origins of *Escherichia coli* growth rate and cell shape changes at high external osmolality”. In: *Biophys J* 107.8 (2014), pp. 1962–9. DOI: 10.1016/j.bpj.2014.08.025.

[178] Z Podlesek and D Žgur Bertok. “The DNA Damage Inducible SOS Response Is a Key Player in the Generation of Bacterial Persister Cells and Population Wide Tolerance”. In: *Front Microbiol* 11.1785 (2020). DOI: 10.3389/fmicb.2020.01785.

[179] Y Pommier et al. “DNA Topoisomerases and their Poisoning by Anticancer and Antibacterial Drugs”. In: *Chem&Biol* 17.5 (2010), pp. 421–433. DOI: https://doi.org/10.1016/j.chembiol.2010.04.012.

[180] A Ponticelli et al. “Chi-dependent DNA strand cleavage by RecBC enzyme”. In: *Cell* 41.1 (1985), pp. 145–51. DOI: 10.1016/0092-8674(85)90069-8.

[181] K Potrykus and M Cashel. “(p)ppGpp: still magical?” In: *Annu Rev Microbiol* 62 (2008), pp. 35–51. DOI: 10.1146/annurev.micro.62.081307.162903.

[182] C Pourciau et al. “Diverse Mechanisms and Circuitry for Global Regulation by the RNA-Binding Protein CsrA”. In: *Front Microbiol* 11.601352 (2020). DOI: 10.3389/fmicb.2020.601352.

[183] J Pribis et al. “Gamblers: An Antibiotic-Induced Evolvable Cell Subpopulation Differentiated by Reactive-Oxygen-Induced General Stress Response”. In: *Mol Cell* 74.4 (2019), pp. 785–800. DOI: 10.1016/j.molcel.2019.02.037.

[184] S Proshkin et al. “Cooperation between translating ribosomes and RNA polymerase in transcription elongation”. In: *Science* 328.5977 (2010), pp. 504–8. DOI: 10.1126/science.1184939.

[185] TL Raivio. “Envelope stress responses and Gram-negative bacterial pathogenesis”. In: *Mol Microbiol* 56.5 (2005), pp. 1119–28. DOI: 10.1111/j.1365-2958.2005.04625.x.

[186] A Raj et al. “Stochastic mRNA synthesis in Mammalian Cells”. In: *PLOS Biol* 4.10 (2006), e309. DOI: 10.1371/journal.pbio.0040309.

[187] L Ranjha, SM Howard and P Cejka. “Main steps in DNA double-strand break repair: an introduction to homologous recombination and related processes”. In: *Chromosoma* 127.2 (2018), pp. 187–214. DOI: 10.1007/s00412-017-0658-1.

[188] JM Raser and EK O’Shea. “Noise in gene expression: origins, consequences, and control”. In: *Science* 309.5743 (2005), pp. 2010–3. DOI: 10.1126/science.1105891.

[189] MT Jr Record et al. “Responses of *E. coli* to osmotic stress: large changes in amounts of cytoplasmic solutes and water”. In: *Trends Biochem Sci* 23.4 (1998), pp. 143–8. DOI: 10.1016/s0968-0004(98)01196-7.

[190] D Rigney and W Schieve. “Stochastic models of linear, continuous protein synthesis in bacterial populations”. In: *J Theor Biol* 69.4 (1977), pp. 761–6.

[191] R Roemhild, T Bollenbach and DI Andersson. “The physiology and genetics of bacterial responses to antibiotic combinations”. In: *Nat Rev Microbiol* (2022). DOI: 10.1038/s41579-022-00700-5.

[192] L Roman, Eggleston A and S Kowalczykowski. “Processivity of the DNA helicase activity of *Escherichia coli* recBCD enzyme”. In: *J Biol Chem* 267.6 (1992), pp. 4207–14.

[193] F Sagues, JM Sancho and J García-Ojalvo. “Spatiotemporal order out of noise”. In: *Rev. Mod. Phys.* 79 (2007), 829–82. DOI: 10.1103/RevModPhys.79.829.

[194] H Salvail et al. “Antagonistic functions between the RNA chaperone Hfq and an sRNA regulate sensitivity to the antibiotic colicin”. In: *EMBO J.* 32.20 (2013), pp. 2764–78. DOI: 10.1038/emboj.2013.205.

[195] A Sanchez, S Choubey and J Kondev. “Stochastic models of transcription: From single molecules to single cells”. In: *Methods* 62.1 (2013), pp. 13 –25. DOI: 10.1016/j.ymeth.2013.03.026.

[196] M Sasaki et al. “Fine structure of the *recB* and *recC* gene region of *Escherichia coli*”. In: *Biochem* 109.2 (1982), pp. 414–422.

[197] E Sauer, S Schmidt and O Weichenrieder. “Small RNA binding to the lateral surface of Hfq hexamers and structural rearrangements upon mRNA target recognition”. In: *PNAS* 109.24 (2012), pp. 9396–401. DOI: 10.1073/pnas.1202521109.

[198] J Schmiedel et al. “MicroRNA control of protein expression noise”. In: *Science* 348.6230 (2015), pp. 128–132. DOI: 10.1126/science.aaa1738.

[199] D Schnoerr, G Sanguinetti and R Grima. “Approximation and inference methods for stochastic biochemical kinetics — a tutorial review”. In: *J. Phys. A: Math. Theor.* 50.9 (2017). DOI: 10.1088/1751-8121/aa54d9.

[200] DJ Schu et al. “Alternative Hfq-sRNA interaction modes dictate alternative mRNA recognition”. In: *EMBO J* 34.20 (2015), pp. 2557–73. DOI: 10.15252/embj.201591569.

[201] D Schultz, A Taylor and G Smith. “*Escherichia coli* RecBC Pseudorevertants Lacking Chi Recombinational Hotspot Activity”. In: *J Bacteriol* 155.2 (1983), pp. 664–80.

[202] D Schümperli et al. “Translational coupling at an intercistronic boundary of the *Escherichia coli* galactose operon”. In: *Cell* 30.3 (1982), pp. 865–71. DOI: 10.1016/0092-8674(82)90291-4.

[203] T Schwartz, E Craig and D Kennell. “Inactivation and degradation of messenger ribonucleic acid from the lactose operon of *Escherichia coli*”. In: *JMB* 54.2 (1970), pp. 299–311. DOI: 10.1016/0022-2836(70)90431-6.

[204] M Scott et al. “Interdependence of cell growth and gene expression: origins and consequences”. In: *Science* 330.6007 (2010), pp. 1099–102. DOI: 10.1126/science.1192588.

[205] D Selinger et al. “Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation”. In: *Genome Res* 13.2 (2003), pp. 216–23. DOI: 10.1101/gr.912603.

[206] LA Sepúlveda et al. “Measurement of gene regulation in individual cells reveals rapid switching between promoter states”. In: *Science* 351.6278 (2016), pp. 1218–22. DOI: 10.1126/science.aad0635.

[207] V Shahrezaei and P Swain. “Analytical distributions for stochastic gene expression”. In: *PNAS* 105.45 (2008), pp. 17256–17261. DOI: 10.1073/pnas.0803850105.

[208] DT Sheidy and RA Zielke. “Analysis and Expansion of the Role of the *Escherichia coli* Protein ProQ”. In: *PLoS ONE* 8.10 (2013). DOI: 10.1371/journal.pone.0079656.

[209] Y Shimoni et al. “Regulation of gene expression by small non-coding RNAs: a quantitative view”. In: *Mol Syst Biol* 3.138 (2007). DOI: 10.1038/msb4100181.

[210] J Shine and L Dalgarno. “The 3’-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites”. In: *PNAS* 71.4 (1974), pp. 1342–6. DOI: 10.1073/pnas.71.4.1342.

[211] F Si et al. “Mechanistic Origin of Cell-Size Control and Homeostasis in Bacteria”. In: *Curr Biol* 29.11 (2019), pp. 1760–70. DOI: 10.1016/j.cub.2019.04.062.

[212] A Singh. “Negative feedback through mRNA provides the best control of gene-expression noise”. In: *IEEE Trans Nanobioscience* 10.3 (2011), pp. 194–200. DOI: 10.1109/TNB.2011.2168826.

[213] A Singh and J Hespanha. “Reducing noise through translational control in an autoregulatory gene network”. In: *IEEE* (2009), pp. 1712–1717. DOI: 10.1109/ACC.2009.5160206.

[214] M Singleton et al. “Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks”. In: *Nature* 432.187 (2004). DOI: 10.1038/nature02988.

[215] A Sinha et al. “Broken replication forks trigger heritable DNA breaks in the terminus of a circular chromosome”. In: *PLOS Genetics* 14.3 (2018), pp. 1–28. DOI: 10.1371/journal.pgen.1007256.

[216] K Skarstad and E Boye. “Degradation of individual chromosomes in *recA* mutants of *Escherichia coli*”. In: *J Bacteriol* 175.17 (1993), 5505–9. DOI: 10.1128/jb.175.17.5505-5509.1993.

[217] S Skinner et al. “Measuring mRNA copy-number in individual *Escherichia coli* cells using single-molecule fluorescent *in situ* hybridization (smFISH).” In: *Nat Protoc* 8.6 (2013), pp. 1100–1113. DOI: 10.1038/nprot.2013.066.

[218] S Smith and R Grima. “Single-cell variability in multicellular organisms”. In: *Nat Commun* 9.345 (2018). DOI: 10.1038/s41467-017-02710-x.

[219] L So et al. “General properties of transcriptional time series in *Escherichia coli*”. In: *Nat Genet* 43 (2011), 554–60. DOI: 10.1038/ng.821.

[220] E Sonnleitner and U Bläsi. “Regulation of Hfq by the RNA CrcZ in *Pseudomonas aeruginosa* Carbon Catabolite Repression”. In: *PLOS Genetics* 10.6 (2014). DOI: 10.1371/journal.pgen.1004440.

[221] T Soper et al. “Positive regulation by small RNAs and the role of Hfq”. In: *PNAS* 107.21 (2010), pp. 9602–9607. DOI: 10.1073/pnas.1004435107.

[222] M Spies et al. “RecBCD enzyme switches lead motor subunits in response to chi recognition”. In: *Cell* 131.4 (2007), pp. 694–705. DOI: 10.1016/j.cell.2007.09.023.

[223] K Stevenson et al. “General calibration of microbial growth in microplate readers”. In: *Sci Rep* 6.38828 (2016). DOI: 10.1038/srep38828.

[224] N Sudo et al. “RNA-binding protein Hfq downregulates locus of enterocyte effacement-encoded regulators independent of small regulatory RNA in enterohemorrhagic *Escherichia coli*”. In: *Mol. Microbiol.* 00 (2021), pp. 1–16. DOI: 10.1111/mmi.14799.

[225] AL Svitil, M Cashel and JW Zyskind. “Guanosine tetraphosphate inhibits protein synthesis *in vivo*. A possible protective mechanism for starvation stress in *Escherichia coli*.” In: *J Biol Chem* 5.268 (1993), pp. 2307–11.

[226] P Swain. “Efficient attenuation of stochasticity in gene expression through post-transcriptional control”. In: *J Mol Biol* 344.4 (2004), pp. 965–76. DOI: 10.1016/j.jmb.2004.09.073.

[227] P Swain, M Elowitz and E Siggia. “Intrinsic and extrinsic contributions to stochasticity in gene expression”. In: *PNAS* 99.20 (2002), pp. 12795–12800. DOI: 10.1073/pnas.162041399.

[228] K Swamy and A Goldberg. “Subcellular distribution of various proteases in *Escherichia coli*”. In: *J Bacteriol* 149 (1982), pp. 1027–33.

[229] B Sy et al. “High-Resolution, High-Throughput Analysis of Hfq-Binding Sites Using UV Crosslinking and Analysis of cDNA (CRAC)”. In: *Methods Mol Biol* 1737 (2018), pp. 251–72. DOI: 10.1007/978-1-4939-7634-8_15.

[230] L Symington. “End Resection at Double-Strand Breaks: Mechanism and Regulation”. In: *Cold Spring Harb Perspect Biol* 6.8 (2014). DOI: 10.1101/cshperspect.a016436.

[231] A Sánchez and J Kondev. “Transcriptional control of noise in gene expression”. In: *PNAS* 105.13 (2008), pp. 5081–6. DOI: 10.1073/pnas.0707904105.

[232] S Taheri-Araghi et al. “Cell-size control and homeostasis in bacteria”. In: *Curr Biol* 25.3 (2017), pp. 385–91. DOI: 10.1016/j.cub.2014.12.009.

[233] Y Taniguchi et al. “Quantifying *E. coli* Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells”. In: *Science* 329.5991 (2010), pp. 533–538. DOI: 10.1126/science.1188308.

[234] A Taylor and G Smith. “Unwinding and Rewinding of DNA by the RecBC Enzyme”. In: *Cell* 22 (1980), pp. 447–57. DOI: 10.1016/0092-8674(80)90355-4.

[235] H Taylor. “Quantitative detection of low abundance gene expression products in individual *E. coli* cells”. PhD thesis. 2018.

[236] M Thattai and A van Oudenaarden. “Intrinsic noise in gene regulatory networks”. In: *PNAS* 98.15 (2001), pp. 8614–9. DOI: 10.1073/pnas.151588598.

[237] P Thomas. “Making sense of snapshot data: ergodic principle for clonal cell populations”. In: *J. R. Soc. Interface* 14.136 (2017). DOI: 10.1098/rsif.2017.0467.

[238] J Tree et al. “Identification of Bacteriophage-Encoded Anti-sRNAs in Pathogenic *Escherichia coli*”. In: *Mol Cell* 55.2 (2014), pp. 199–213. DOI: 10.1016/j.molcel.2014.05.006.

[239] H Tsui, H Leung and M Winkler. “Characterization of broadly pleiotropic phenotypes caused by an *hfq* insertion mutation in *Escherichia coli* K-12”. In: *Mol Microbiol* 13.1 (1994), pp. 35–49. DOI: 10.1111/j.1365-2958.1994.tb00400.x.

[240] J Ule et al. “CLIP Identifies Nova-Regulated RNA Networks in the Brain”. In: *Science* 302.5648 (2003), pp. 1212–1215. DOI: 10.1126/science.1090095.

[241] T Updegrove, A Zhang and G Storz. “Hfq: the flexible RNA matchmaker”. In: *Curr Opin Microbiol* 30 (2016), pp. 133–138. DOI: 10.1016/j.mib.2016.02.003.

[242] S Uphoff et al. “Stochastic activation of a DNA damage response causes cell-to-cell mutation rate variation”. In: *Science* 351.6277 (2016), pp. 1094–1097. DOI: 10.1126/science.aac9786.

[243] E Van Assche et al. “RNA-binding proteins involved in post-transcriptional regulation in bacteria”. In: *Front Microbiol* 6.141 (2015), pp. 1–12. DOI: 10.3389/fmicb.2015.00141.

[244] NG Van Kampen. *Stochastic processes in physics and chemistry*. Elsevier, 1992.

[245] J Vastola. “Solving the chemical master equation for monomolecular reaction systems and beyond: a Doi-Peliti path integral view”. In: *J Math Biol* 83.48 (2021). DOI: 10.1007/s00285-021-01670-7.

[246] B Vecerek, I Moll and U Bläsi. “Translational autocontrol of the *Escherichia coli hfq* RNA chaperone gene”. In: *RNA* 11.6 (2005), pp. 976–84. DOI: 10.1261/rna.2360205.

[247] J Vogel and B Luisi. “Hfq and its constellation of RNA”. In: *Nat Rev Microbiol* 9.8 (2011), 578–89. DOI: 10.1038/nrmicro2615.

[248] U Vogel and K Jensen. “The RNA chain elongation rate in *Escherichia coli* depends on the growth rate”. In: *J Bacteriol.* 176.10 (1994), pp. 2807–13. DOI: 10.1128/jb.176.10.2807-2813.1994.

[249] J Wade et al. “Genomic analysis of LexA binding reveals the permissive nature of the *Escherichia coli* genome and identifies unconventional target sites”. In: *Genes Dev* 19.21 (2005), pp. 2619–30. DOI: 10.1101/gad.1355605.

[250] M Wallden et al. “The synchronization of replication and division cycles in individual *E.coli* cells”. In: *Cell* 166 (2016), 729–739. DOI: 10.1016/j.cell.2016.06.052.

[251] M Wang et al. “Measuring transcription at a single gene copy reveals hidden drivers of bacterial individuality”. In: *Nat. Microbiol* 4 (2019), 2118–2127. DOI: 10.1038/s41564-019-0553-z.

[252] P Wang et al. “Robust growth of *Escherichia coli*”. In: *Curr Biol.* 22.12 (2010), pp. 1099–103. DOI: 10.1016/j.cub.2010.04.045.

[253] SA Waters et al. “Small RNA interactome of pathogenic *E. coli* revealed through crosslinking of RNase E”. In: *EMBO J* 36.3 (2017), pp. 374–87. DOI: 10.15252/embj.201694639.

[254] H Weber et al. “Genome-wide analysis of the general stress response network in *Escherichia coli*: sigmaS-dependent genes, promoters, and sigma factor selectivity”. In: *J Bacteriol* 187.5 (2005), pp. 1591–603. DOI: 10.1128/JB.187.5.1591-1603.2005.

[255] MH Weber and MA Marahiel. “Bacterial cold shock responses”. In: *Sci Prog* 86.1-2 (2003), pp. 9–75. DOI: 10.3184/003685003783238707.

[256] M White et al. “RecBCD coordinates repair of two ends at a DNA double-strand break, preventing aberrant chromosome amplification”. In: *Nucleic Acids Res* (June 2018). DOI: 10.1093/nar/gky463.

[257] J Wiktor et al. “Direct observation of end resection by RecBCD during double-stranded DNA break repair *in vivo*”. In: *Nucleic Acids Res* 46.4 (2018), pp. 1821–1833. DOI: 10.1093/nar/gkx1290.

[258] JM Wood. “Bacterial osmoregulation: a paradigm for the study of cellular homeostasis”. In: *Annu Rev Microbiol* 65 (2011), pp. 215–38. DOI: 10.1146/annurev-micro-090110-102815.

[259] C Wyman and R Kanaar. “DNA double-strand break repair: all’s well that ends well”. In: *Annu Rev Genet* 40 (2006), pp. 363–83. DOI: 10.1146/annurev.genet.40.110405.090451.

[260] H Xu et al. “Stochastic Kinetics of Nascent RNA”. In: *Phys Rev Lett* 117.12 (2016), p. 128101. DOI: 10.1103/PhysRevLett.117.128101.

[261] K Yamaguchi et al. “Pulse-chase experiment for the analysis of protein stability in cultured mammalian cells by covalent fluorescent labeling of fusion proteins”. In: *Methods Mol Biol* 577 (2009), pp. 121–31. DOI: 10.1007/978-1-60761-232-2_10.

[262] J Yewdell et al. “Out with the old, in with the new? Comparing methods for measuring protein degradation”. In: *Cell Biol Int* 35.5 (2011), pp. 457–62. DOI: 10.1042/CBI20110055.

[263] K Zahradka et al. “Roles of ExoI and SbcCD nucleases in "reckless" DNA degradation in *recA* mutants of *Escherichia coli*”. In: *J Bacteriol* 191.5 (2009), 1677–87. DOI: 10.1128/JB.01877-07.

[264] A Zhang et al. “Global analysis of small RNA and mRNA targets of Hfq”. In: *Mol Microbiol* 50.4 (2003), pp. 1111–24. DOI: 10.1046/j.1365-2958.2003.03734.x.

[265] A Zhang et al. “Mutations in interaction surfaces differentially impact *E. coli* Hfq association with small RNAs and their mRNA targets”. In: *J Mol Biol* 425.19 (2013), pp. 3678–97. DOI: 10.1016/j.jmb.2013.01.006.

[266] J Zhang, L Chen and T Zhou. “Analytical distribution and tunability of noise in a model of promoter progress”. In: *Biophys J* 102.6 (2012), pp. 1247–57. DOI: 10.1016/j.bpj.2012.02.001.

[267] Y Zhang and CA Gross. “Cold Shock Response in Bacteria”. In: *Annu Rev Genet* 55 (2021), pp. 377–400. DOI: 10.1146/annurev-genet-071819-031654.