Article

Effects of Population Dynamics on Establishment of a Restriction-Modification System in a Bacterial Host

Stefan Graovac [1,2], Andjela Rodic [1,2], Magdalena Djordjevic [3], Konstantin Severinov [4,5] and Marko Djordjevic [1,*

[1] Faculty of Biology, University of Belgrade, 11000 Belgrade, Serbia; stefan.graovac@bio.bg.ac.rs (S.G.); andjela.rodic@bio.bg.ac.rs (A.R.)
[2] Multidisciplinary PhD program in Biophysics, University of Belgrade, 11000 Belgrade, Serbia
[3] Institute of Physics Belgrade, University of Belgrade, 11080 Belgrade, Serbia; magda@ipb.ac.rs
[4] Waksman Institute of Microbiology, Rutgers University, Piscataway, NJ 08854, USA; severik@waksman.rutgers.edu
[5] Center for Life Sciences, Skolkovo Institute of Science and Technology, Skolkovo 143026, Russia
[*] Correspondence: dmarko@bio.bg.ac.rs; Tel.: +381-11-3033356; Fax: +381-11-2639-882

check for
updates

Abstract: In vivo dynamics of protein levels in bacterial cells depend on both intracellular regulation and relevant population dynamics. Such population dynamics effects, e.g., interplay between cell and plasmid division rates, are, however, often neglected in modeling gene expression regulation. Including them in a model introduces additional parameters shared by the dynamical equations, which can significantly increase dimensionality of the parameter inference. We here analyse the importance of these effects, on a case of bacterial restriction-modification (R-M) system. We redevelop our earlier minimal model of this system gene expression regulation, based on a thermodynamic and dynamic system modeling framework, to include the population dynamics effects. To resolve the problem of effective coupling of the dynamical equations, we propose a "mean-field-like" procedure, which allows determining only part of the parameters at a time, by separately fitting them to expression dynamics data of individual molecular species. We show that including the interplay between kinetics of cell division and plasmid replication is necessary to explain the experimental measurements. Moreover, neglecting population dynamics effects can lead to falsely identifying non-existent regulatory mechanisms. Our results call for advanced methods to reverse-engineer intracellular regulation from dynamical data, which would also take into account the population dynamics effects.

Keywords: restriction-modification systems; bacterial population dynamics; gene expression control; statistical thermodynamics; transcription regulation

1. Introduction

Technological developments in the past few decades enabled experimental in vivo measurements of protein levels in single cells with high temporal resolution, thus providing a good basis for studying gene expression control by mathematical modeling [1,2]. The ultimate goal of these studies is to provide accurate predictions of gene expression dynamics in a cell, which is of particular importance for the emerging synthetic biology field. In particular, advances in fluorescence microscopy and microfluidics [1,3–5] allowed measurements of protein expression levels in a clonal culture descending from the same (single) cell, which abolishes the need to synchronize the cell population. First such measurements for bacterial restriction-modification (R-M) systems have recently become available [6].

R-M systems have two main components: restriction endonuclease (R) recognizes specific DNA sequences and cuts them, while methyltransferase (M) methylates the same sequences and thereby protects them from being cut [7,8]. R-M systems carried on plasmids can spread to new (naïve) bacterial hosts by horizontal gene transfer. Consequently, the synthesis of R and M has to be tightly controlled, to ensure safe and efficient R-M establishment in a naïve bacterial host. Specifically, R has to be synthesized with a delay with respect to M, so that the host genome is protected by M, before R activity appears [7,9]. Furthermore, since the same R-M system can be established in different bacterial species, system regulation must be at least partially independent from the host transcription factors, i.e., encoded by the system itself. This property to a certain extent simplifies identifying all system regulation components and correctly describing regulatory mechanisms.

We previously showed that biophysical and dynamical system modeling can reasonably explain available experimental data on R-M transcription control [10–12]. However, in vivo protein expression dynamics reflect not only transcription regulation, but also the (often neglected) effects of global physiological factors, which can change in the culture [13–16]. It was early pointed out that to understand function of cellular networks, they should be considered in their natural environment [17]. The global physiological factors can change significantly in the environment, which impacts the growth of the population of cells. Dilution of molecule concentrations upon cell division, and the changes in gene copy number, which are especially pronounced when the gene of interest is carried on a plasmid, are prominent growth rate dependent factors [18]. These effects can significantly modulate gene expression dynamics. For example, it was previously shown in the case of the celebrated *lac* operon that population dynamics effects have to be included in modeling of its gene expression to explain the measured data [19]. For a review of interdependence between cell growth and gene expression, studied in other systems, see [18] and references therein.

Therefore, the obtained time profile of protein expression is in reality a result of coupled specific regulatory, and global physiological factors that modulate the relevant population dynamics (dynamics of cell and plasmid division). The Esp1396I R-M system which will be analysed here, provides an interesting example of an unexpected protein expression time profile [6]. In this experiment, fluorescently labeled M and R amounts were tracked in time at the level of single cells in a microcolony originating from a single cell transformed with the system on a plasmid. The obtained time profile can only in part be explained by a model of known system regulation mechanisms [6]. The minimal theoretical model (which includes the experimentally established transcription regulation mechanisms, and dilution of molecular species in accordance with empirically observed bacterial population growth) explains the main qualitative characteristics of Esp1396I system expression dynamics: massive initial synthesis of methyltransferase and a delayed start of restriction endonuclease synthesis. However, while the model provides a good description of early-time dynamics, it does not agree with experiments at later times—see Figure 4C in [6]. For example, to explain the observed increase in intracellular amounts of M at later times, one would have to invoke an otherwise non-existing transcription activation mechanism.

On the other hand, the bacterial cell division rate changes during the course of this particular experiment, so two growth regimes can be distinguished, in which the population grows exponentially, first with a higher and later, when the preferred nutrient is exhausted, with a lower cell division rate (see Figure 4A in [6]) [6,20]. In addition, the intracellular plasmid copy number dynamically increased during the experiment (unpublished observation). Consequently, there can be a highly non-trivial (and time dependent) interplay between cell and plasmid division rates, which might significantly affect the observed patterns of R and M dynamics. Such interplay (population dynamics effects) was neglected in our minimal model of Esp1396I expression dynamics, and is commonly neglected in modeling of bacterial intracellular dynamics. Therefore, before assuming that the unexplained properties of the measured dynamics are due to the action of some unknown regulatory mechanism(s) operating in the system, one should consider that they may arise from the already modelled regulation combined with the missing population dynamics effects.

Consequently, a major goal of this research is to analyse the importance of population dynamics effects for intracellular dynamics. Understanding this is necessary to accurately predict gene expression dynamics in a cell, which is also crucial for a number of practical applications, ranging from synthetic biology [21,22] to bacterial antibiotic resistance [18]. This is here addressed on the example of Esp1396I R-M system gene expression control since: (*i*) state-of-the-art measurements of protein dynamics for this system are available [6]; (*ii*) our previous work showed that theoretical modeling can accurately explain R-M experimental measurements [6,10–12], and (*iii*) R-M systems are both important experimental model systems and sufficiently simple to be realistically theoretically modeled. In particular, the number of parameters is significantly lower to what is often encountered in models of gene expression networks/dynamics [23], and all the effects introduced in the model are supported by direct experimental observations [6].

2. Methods

2.1. Modeling cr Operon and m Gene Transcription Activities

The C protein is a transcription factor, encoded by Esp1396I [9,24], which regulates transcription of both the operon consisting of the *c* gene itself and the *r* gene (*cr* operon), and the methyltransferase (*m*) gene, as indicated in (Figure 1A). In particular, binding of the C protein dimer to a site partially overlapping with the strong P$_M$ promoter of the *m* gene (denoted with MBS—for Methyltransferase Binding Site), represses *m* transcription (Figure 1C) [24].
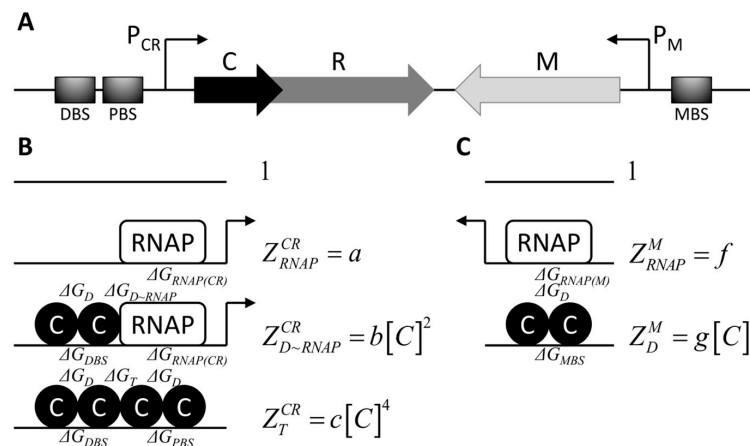

Figure 1. Modelling regulation of transcription in the R-M system Esp1396I. (**A**) Gene organization in the Esp1396I system: genes encoding control (C) protein and restriction endonuclease (R) are co-transcribed from the P$_{CR}$ promoter, while the gene encoding methyltransferase (M) is transcribed from the P$_M$ promoter. DBS (distal binding site), PBS (proximal binding site) and MBS (binding site in the P$_M$) represent three binding sites for C protein, where PBS and MBS overlap with the corresponding core promoters (RNAP binding sites); (**B**) Allowed configurations of RNA polymerase (a rectangle denoted as "RNAP") and C protein molecules (circles) on the P$_{CR}$ and (**C**) the P$_M$ promoters are schematically represented. Configurations in which the promoter is transcriptionally active are labelled with an arrow. The following free energies of protein-DNA and protein-protein interactions are indicated in the figure: $\Delta G_{RNAP(CR)}$ and $\Delta G_{RNAP(M)}$—for binding of RNAP to P$_{CR}$ and P$_M$, respectively; ΔG_{DBS}, ΔG_{PBS} and ΔG_{MBS}—for binding of a C dimer to the distal and the proximal binding sites, and to the binding site in the P$_M$ region, respectively; ΔG_D—for C protein dimerization; $\Delta G_{D\sim RNAP}$—for the interaction between the C dimer bound to the DBS and RNAP; ΔG_T—for interaction between two bound C dimers. For each configuration, a statistical weight is assigned (on the right of the configuration), expressed in terms of C protein concentration ([C]) and constants a, b, c, f and g that are independent of C protein concentration.

The mechanism of *cr* transcription regulation is more complex (Figure 1B): at low C concentrations, C dimer is bound to the strong, distal binding site (DBS), and recruits RNA polymerase (RNAP) to the P_{CR} promoter thereby activating transcription of the *cr* operon (Figure 1B); at higher C concentrations, another C dimer is recruited to the adjacent weak, proximal binding site (PBS) forming a C tetramer on DNA, which represses transcription of *cr* [10,24].

Transcription activity of Esp1396I R-M system promoters P_{CR} and P_M depicted in Figure 1A is modelled using statistical thermodynamics, i.e., by representing the relevant molecular components in a cell as a system whose probability of different microstates is described by the Boltzmann distribution. Appropriate dimensionless statistical weights are assigned to each equilibrium configuration of RNA polymerase and C protein molecules at P_{CR} (Figure 1B) and P_M (Figure 1C), reflecting energetic and entropic costs of their establishment [25,26]. One should note that the same statistical weights can be obtained by applying the law of mass action to appropriate equilibrium binding reactions, where the equilibrium dissociation constants $K_d \sim \exp(\Delta G)$ (see [26] for a brief description of both approaches).

Statistical weights of the P_{CR} promoter configurations are given by the following Equations:

$$Z_{RNAP}^{CR} = k[\text{RNAP}] \exp\left(-\Delta G_{RNAP(CR)}\right) \tag{1}$$

$$Z_{D \sim RNAP}^{CR} = k^3 [\text{C}]^2 [\text{RNAP}] \exp\left(-\Delta G_{DBS} - \Delta G_{RNAP(CR)} - \Delta G_D - \Delta G_{D \sim RNAP}\right) \tag{2}$$

$$Z_T^{CR} = k^4 [\text{C}]^4 \exp(-\Delta G_{DBS} - \Delta G_{PBS} - 2\Delta G_D - \Delta G_T) \tag{3}$$

where statistical weights in Equations (1)–(3) correspond, respectively, to: (*i*) only RNA polymerase bound to the promoter, corresponding to basal transcription of the genes encoding C protein and restriction endonuclease (Equation (1)), (*ii*) RNA polymerase recruited to the promoter by the C dimer bound to the DBS (Equation (2)), and (*iii*) a second C dimer recruited to the PBS by the C dimer bound to the DBS, forming a C tetramer on DNA which represses transcription (Equation (3)). In the upper equations, k is a proportionality constant (with units one over concentration), concentrations of molecular species are labelled with square brackets, while protein-DNA and protein-protein interaction energies that enter Equations (1)–(3), (ΔGs in the units of $k_B T$) are denoted in Figure 1B and listed in the caption of the Figure. Constants in Equations (1)–(3) can be absorbed into few parameters (*a*, *b* and *c*) that do not depend on C concentration, which results in the expressions for statistical weights denoted next to the appropriate configurations in Figure 1B.

The standard assumption in thermodynamic modeling is that transcriptional activity of the promoter is proportional to the equilibrium probability that RNA polymerase is bound to that promoter [27]. Accordingly, transcriptional activity of P_{CR} is given by the following equation (1 is the statistical weight corresponding to the empty promoter):

$$\varphi_r = \alpha \frac{Z_{RNAP}^{CR} + Z_{D \sim RNAP}^{CR}}{1 + Z_{RNAP}^{CR} + Z_{D \sim RNAP}^{CR} + Z_T^{CR}} \tag{4}$$

where α is a proportionality constant with units of transcript amount over time. Equation (4) can be rewritten introducing the derived statistical weights from Figure 1B, and defining the basal transcription activity of the P_{CR} in the absence of C protein, $\varphi_r^{basal} = \alpha \cdot a/(1 + a)$:

$$\varphi_r(\text{C}) = \varphi_r^{basal} \frac{1 + \frac{b}{a}[\text{C}]^2}{1 + \frac{b}{a+1}[\text{C}]^2 + \frac{c}{a+1}[\text{C}]^4}. \tag{5}$$

Transcription activity of P_M promoter is modelled in the same manner (Figure 1C), assuming that it can be found: (*i*) empty, (*ii*) in a transcriptionally active state when it is occupied by RNA

polymerase, or (*iii*) in a repressed state, when a C dimer is bound to the MBS. Statistical weights of the configurations described under (*ii*) and (*iii*) read:

$$Z^M_{RNAP} = k[\text{RNAP}] \exp\left(-\Delta G_{RNAP(M)}\right) \tag{6}$$

$$Z^M_D = k^2[\text{C}]^2 \exp(-\Delta G_{MBS} - \Delta G_D) \tag{7}$$

while the P_M transcription activity is given by the equation:

$$\varphi_m = \beta \frac{Z^M_{RNAP}}{1 + Z^M_{RNAP} + Z^M_D} \tag{8}$$

where β is a proportionality constant with units of transcript amount over time. As in the case of the P_{CR} activity modeling, constants in Equations (6) and (7) can be absorbed into few parameters (f and g) that do not depend on C concentration, which results in the expressions for statistical weights denoted next to the appropriate configurations in Figure 1C. Equation (8) becomes

$$\varphi_m(\text{C}) = \varphi_m^{basal} \frac{1}{1 + K_i^2[\text{C}]^2} \tag{9}$$

when $K_i^2 = g/(f+1)$ and $\varphi_m^{basal} = \beta \cdot f/(f+1)$ are introduced. Note that K_i corresponds to the equilibrium association constant for C dimer binding to DNA in the presence of a bound RNAP, which exerts an inhibiting effect on transcription.

2.2. Introducing the Interplay of Cell and Plasmid Division Rates

We here develop a model that includes an interplay of cell and plasmid division rates, so that full population dynamics effects due to (time dependent) division of both cells and plasmids are taken into account. Note that R-M systems typically produce thousands of RNAs and proteins of the two enzymes in the cell, so these systems can be reliably modelled deterministically.

The number of R-M system encoding plasmids per cell (n_p) is introduced as a time dependent variable, which increases due to plasmid replication, and decreases due to dilution after cell division. Specifically, the change of cell numbers (n_{cell}), and the number of plasmids per cell (n_p) with time is described by the following differential equations:

$$\frac{dn_{cell}}{dt} = \lambda_{cell} \cdot n_{cell} \tag{10}$$

$$\frac{dn_p}{dt} = (\lambda_p - \lambda_{cell}) \cdot n_p \tag{11}$$

where λ_{cell} and λ_p are division rates for cells and plasmids, respectively. Experimentally measured cell number dependence with time [6] indicates a relatively sharp transition from faster to slower cell division rate, happening at t_{cell}^{trans} ~150 min. To obtain a curve that can satisfactory describe the measured data, the area of transition is interpolated using the hyperbolic tangent function. Note that this interpolation covers a relatively short time interval ~30 min, so it does not significantly affect the predictions, but is necessary to avoid discontinuity when solving differential equations—moreover, it naturally extrapolates between the two regimes of cell division. In particular, $\lambda_{cell}(t)$ is determined as follows:

$$w_{cell} = \tfrac{1}{2} + \tfrac{1}{2} \cdot \tanh\left(\frac{t - t_{cell}^{trans}}{s_{cell}}\right)$$
$$\lambda_{cell}(t) = w_{cell} \cdot \lambda_{cell2} + (1 - w_{cell}) \cdot \lambda_{cell1} \tag{12}$$

where λ_{cell1}, λ_{cell2} are constant parameters denoting the cell division rates in the first and the second time interval, while s_{cell} is a fixed parameter defining smoothness of the interpolation (taken as

120 min). Note that Equation (12) is just an empirical fit (basically continual interpolation) to the experimentally measured data, which is an input to the model of the gene expression dynamics, rather than a model prediction.

The plasmid division rate $\lambda_p(t)$ is introduced in an analogous manner:

$$w_p = \tfrac{1}{2} + \tfrac{1}{2}\cdot\tanh\left(\frac{t - t_p^{trans}}{s_p}\right)$$
$$\lambda_p(t) = w_p\cdot\lambda_{p2} + \left(1 - w_p\right)\cdot\lambda_{p1}$$

(13)

with equivalent parameter labelling as in Equation (12). Finally, in the full dynamical model, the number of plasmids is obtained by solving Equations (11) and (13), and then multiplied by promoter transcription activities to obtain the generation rates of the corresponding RNA species. The full model is described with the following set of differential equations:

$$\frac{dr}{dt} = n_p\cdot\varphi_r(C) - \lambda_r\cdot r$$

(14)

$$\frac{dm}{dt} = n_p\cdot\varphi_m(C) - \lambda_m\cdot m$$

(15)

$$\frac{dC}{dt} = k_C\cdot r - \lambda_C\cdot C$$

(16)

$$\frac{dR}{dt} = k_R\cdot r - \lambda_R\cdot R$$

(17)

$$\frac{dM}{dt} = k_M\cdot m - \lambda_M\cdot M.$$

(18)

Equations (14) and (15) describe how the amounts of transcripts of the *cr* operon and the *m* gene change with time, while Equations (16)–(18) describe the same for the amounts of proteins, namely, of C protein (C), restriction endonuclease (R) and methyltransferase (M). For parameter notation, see Table 1. Note that each transcript and protein decay rate (λ's in the equations above) represents a sum of λ_{cell} (Equation (12)) and the corresponding molecule degradation rate (which we mark by ~ signs, e.g., $\lambda_r = \tilde{\lambda}_r + \lambda_{cell}(t)$). Since we take protein decay rates of R and C to be equal, Equation (17) can be excluded from solving, i.e., replaced by an algebraic relation $R = C\cdot k_R/k_C$ (note that C and R in Equations (16) and (17) can be rescaled with k_C and k_R, respectively, leading to the same equation).

2.3. Numerically Solving and Fitting the Model Equations

If the population dynamics effects, in particular changes of per cell plasmid copy number throughout the experiment, would be neglected [6], the equations for R (and consequently also C, see above) dynamics could be solved separately from the equation for M dynamics. Once C dynamics are solved, they could be used to determine the M dynamics (since the *m* gene is regulated by C, see Equations (15) and (18)), as was done with the minimal model provided in [6]. However, the introduction of plasmid dynamics in the model leads to nontrivial coupling of R (Equations (14), (16) and (17)) and M dynamics (Equations (15) and (18)) through the time-dependent term n_p that enters both Equations (14) and (15). Consequently, these sets of equations have to be solved simultaneously, and their parameters can no longer be separately fitted to the experimental data for R and M dynamics. This then significantly increases dimensionality of the parameter inference problem, i.e., the joint fit leads to inferring parameters in a 17-dimensional space, which is computationally complicated, since a very large number of parameter combinations have to be explored to find the best fit of the model to experimental data.

To resolve this problem, which would become even more prominent with a larger number of species in the model, we here propose an iterative, "mean field-like" approach to effectively decouple such equations. The main idea is schematically depicted in Figure 2A, and in essence corresponds to solving for dynamics of only one molecular species (W in Figure 2A), in the "field" obtained by empirically approximating the dynamics of the other species (U and V and arrow 1 in Figure 2A). This then allows inferring the population dynamics parameters (which couple the species dynamics), by fitting the model only to experimental data for W dynamics. Once these population dynamics parameters are fixed, one can then return (arrow 2 in Figure 2A) to other species, and separately solve for their dynamics, by fits to corresponding experimental data. With solved dynamics for U and V, one then goes back (arrow 3) and solves W dynamics: (*i*) with either fixed (previously inferred) parameters for population dynamics, in the case that the procedure converges, i.e., leads to a satisfactory fit to experimental data or (*ii*) by refitting (inferring again) the population dynamics parameters, if the convergence has not been achieved—in this case, the whole procedure is further iteratively repeated until convergence.

Table 1. Model parameters.

Notation	Value	Description
Population Dynamics		
λ_{cell1}	2.3×10^{-2}	Cell division rate in the first time interval
λ_{cell2}	3.1×10^{-3}	Cell division rate in the second time interval
t_{cell}^{trans}	1.5×10^{2}	Time of transition between the cell division rates
λ_{p1}	2.3×10^{-2}	Plasmid division rate in the first time interval
λ_{p2}	7.5×10^{-3}	Plasmid division rate in the second time interval
t_{p}^{trans}	4.2×10^{2}	Time of transition between the plasmid division rates
Restriction endonuclease dynamics		
φ_r^{basal}	5.3×10^{-1}	Basal transcription activity of the P_{CR} promoter
a	2.7×10^{-1}	Constants which absorb the relevant interaction energies and RNA polymerase concentration
b	4.3×10^{-3}	
c	2.7×10^{-9}	
k_C	5.1×10^{-1}	Translation rate for control protein
k_R	$= k_C \cdot 2/3$	Translation rate for restriction endonuclease
λ_r	3.8×10^{-2}	Decay rate for the operon transcript
λ_C	1.2×10^{-6}	Decay rate for control protein
λ_R	$= \lambda_C$	Decay rate for restriction endonuclease
Methyltransferase dynamics		
φ_m^{basal}	4.6×10^{1}	Basal transcription activity of the P_M promoter
K_i	3.0×10^{-5}	Constants which absorb the relevant interaction energies and RNA polymerase concentration
k_M	1.5	Translation rate for methyltransferase
λ_m	2.0×10^{-1}	Decay rate for the m gene transcript
λ_M	8.0×10^{-4}	Decay rate for methyltransferase

Time is measured in minutes, while the rates are given in 1/min.

The application of this procedure to R-M dynamics is depicted in Figure 2B. Note that the time dependence of R directly determines the time dependence of C, as they are proportional to each other (see the previous subsection). The procedure is started by empirically fitting the time dependence of R, as it is simpler compared to M. In fact, it can be well approximated by quadratic dependence on time. On the other hand, a more complex time dependence of M data is more suitable (i.e., likely more sensitive) to inferring the population dynamics parameters, in addition to fit being of smaller dimensionality compared to R. Once C dynamics is empirically fitted, the Equations (15) and (18) for M dynamics are solved and the solution is fitted to corresponding experimental data, from which the plasmid dynamics parameters are inferred. Then, one can use these parameters to solve for R dynamics, and thereby determine the rest of the parameters in R dynamics. One then goes back to M dynamics and solves it by using the newly inferred C(*t*). If the inferred (fixed) parameters of M dynamics lead to a satisfactory agreement with experimental data (converge) the procedure is stopped. If not, the plasmid dynamics parameters are again estimated (with C(*t*) from the previous step). Note that after each full cycle, the solutions for both M(*t*) and R(*t*) (i.e., C(*t*)) exactly solve the full system of the dynamical equations, though the parameters inferred after each cycle may not provide an optimal fit to data—thus, the fit is, if needed, improved through iterative cycles.
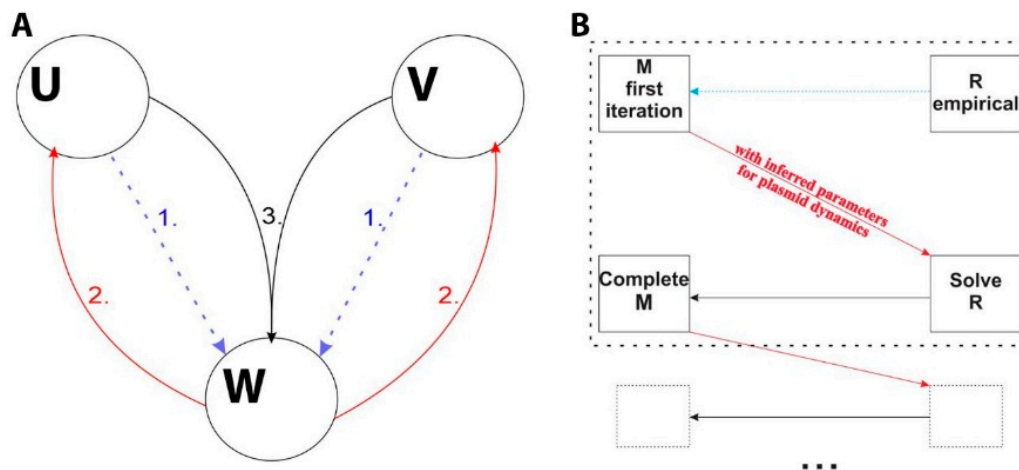

Figure 2. Schematic representation of "mean field-like" procedure for reducing dimensionality of the parameter inference problem for the model with population dynamics effects included. (**A**) Generalized procedure. Circles represent sets of ODEs describing dynamics of particular species (here labeled by U, V and W). The blue dashed arrows denote that in the first step, dynamics of species U and V are empirically approximated and used to determine the parameters of W dynamics by solving only W related ODEs. In the next fitting step (indicated by the red solid arrows and the number 2), the shared parameters (those describing population dynamics) inferred by fitting in the previous step are used in fitting the solution of U and V related ODEs and thereby inferring their corresponding parameters. In the step 3, the fitted (and not the empirically approximated) solution of the U and V related ODEs is used in inferring the parameters of W dynamics. The steps 2 and 3 repeat iteratively until the best fit of the whole model to the data is obtained; (**B**) Applying the procedure to the case of R-M dynamics: boxes represent consecutive fitting steps, and arrows indicate the order of solving the sets of equations corresponding to R and M dynamics. Dashed arrows and squares indicate that the steps in the upper box can be iteratively repeated until convergence is reached. The arrow style and labeling is equivalent as in part A of the Figure.

To implement the above procedure, the system of differential Equations (11), (14)–(16) and (18) that represents the full model of Esp1396I R-M system is solved numerically using the Runge-Kutta method [28]. The initial conditions are set to zero for all species except for plasmids, for which one plasmid per cell at the beginning of the experiment was set. The system parameters are varied in the ranges that correspond to biochemically realistic values [25]. The set of parameters that best fits

experimental data was determined by minimizing R^2 (the sum of the squared errors). Note that this numerical procedure is quite more complicated than standard fits to experimental data, as in our case there is no closed form expression that can be fitted to the data points. That is, the closed form expressions for transcription activities (Equations (1)–(9)) serve as an input for non-linear differential equations (Equations (14)–(18)), which cannot be integrated in a closed form expression, but have to be solved numerically, and these solutions then compared with experimental data.

To quantify the model comparison with experimental data, adjusted R^2 was calculated for each fit. To quantitatively compare different fits (e.g., for "complex" and "simple" models), F values and corresponding P values used [29]:

$$\text{F} = \left(\frac{\chi_1^2 - \chi_2^2}{p_2 - p_1} \right) \Big/ \left(\frac{\chi_2^2}{n - p_2} \right) \tag{19}$$

where χ_1^2 and χ_2^2 are the sum of residuals squared for the first ("simple"), and the second ("complex") model; p_1 and p_2 are the corresponding numbers of parameters, while n is the number of data points. From this, we can estimate the corresponding P-values from cumulative distribution function of F statistics.

3. Results and Discussion

3.1. Including the Population Dynamics Effects

The minimal model from [6] was redeveloped to include the effects of simultaneous cell and plasmid division. However, the associated coupling of R and M dynamics brought a technical challenge. In particular, since the system genes are located together on a plasmid (see Equations (14)–(18)), the corresponding equations for R and M can no longer be solved (and fitted to experimental data) separately. This in turn resulted in a significant increase in the dimensionality of the parameter inference problem (to 17). This technical difficulty would become even more pronounced in a model with a larger number of species, where their dynamics would become coupled due to introducing population dynamics effects. To solve this problem, we here proposed a "mean field-like" iterative approach in which the dynamics of two enzymes is effectively decoupled by first empirically approximating dynamics of species 1 and using it to solve for the parameters related to the dynamics of species 2, including those describing the population effects.

Next, the parameters inferred for species 2 are used to solve the equations for species 1 and estimate the appropriate parameters of its regulation. The new species 1 parameters are then used to solve again for the species 2 parameters, and the process is repeated iteratively until the best fit to both species dynamics data is obtained. Note that this "mean field-like" procedure can be generalized to multiple molecular species, as described in Methods (see Figure 1A). In addition to a much simpler parameter inference, note that the decreased dimensionality of the parameter space, also leads to generically smaller chance to overfit the data (due to a smaller number of parameters involved in the fit).

Applying this procedure to Esp1396I system data leads to a much better agreement of the model with experiments (Figure 3A), compared to the fit of the minimal model (Figure 4C in [6]). In particular, the adjusted R^2 for the two fits in Figures 3A and 4C from [6], are respectively, 0.97 and 0.25, where F value (160) for the fit comparison is statistically highly significant ($P \sim 10^{-15}$). The increase of M later in the experiment, which is now accounted for, is a consequence of slower cell division compared to plasmid replication later in the experiment and not due to overfitting (Figure 3A). These results also demonstrate the suitability of "the mean field–like" procedure for reducing the dimensionality of the parameter inference problem. The inferred parameter values (see Table 1) are also in general agreement with experimental observations: high stability of R-M proteins predicted by the model is in a direct agreement with experimental observations [6]; the inferred decay rates are also in accordance with the standard expectation that mRNAs are more rapidly degraded than proteins [30]; the obtained

three times lower translation rate of *cr* compared to *m* is consistent with providing a robust delay of R with respect to M, and such mechanism was also found in AhdI R-M system [10]. The errors for the parameter fit values could be estimated through either Monte Carlo or bootstrapping procedure. While this is in principle straightforward, it is in practice highly computationally intensive, as the system has to be simulated for all the parameter combinations and for all of the many generated synthetic datasets [31].
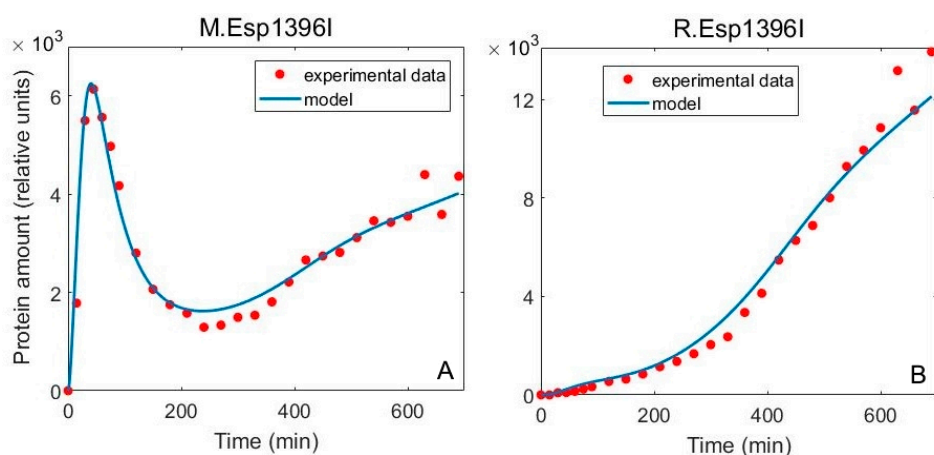

Figure 3. Comparison of the model with included population dynamics with experimental data. Comparison of predictions for Esp1396I R-M system with experimentally measured time-dependence of protein amounts in single cells. The model (blue curve) is fitted to experimental data (red dots) for (**A**) methyltransferase (adjusted R^2 = 0.97) and (**B**) restriction endonuclease dynamics (adjusted R^2 = 0.98).

3.2. Including Population Dynamics Improves Agreement with the Expression Measurements

We next address if the number of parameters in the model is minimal, i.e., if experimental data could be explained with a less complex model. Namely, plasmid dynamics were modelled with three parameters representing the plasmid division rate in the first time interval (λ_{p1}), the plasmid division rate in the second time interval (λ_{p2}), and the time of transition between the two intervals (t_p^{trans}). Such dependence is analogous to experimentally measured cell division rate, where two intervals with almost constant division rates (transition at ~150 min; see Figure 4A) are separated by a transition period. By comparing Figure 4A,B, one can see that plasmid dynamics providing the best fit of the model to the data comprise a late (at ~400 min; see Figure 4B) transition between the two intervals with constant plasmid division rates. This inferred transition of the plasmid division rate from higher to lower value is consistent with general notion of positive correlation between the cell division rate and gene copy number increase [18], and with the fact that the cell division rate becomes slower later in the experiment (see Figure 4A).

We next test if the data can be explained by a simpler model of plasmid dynamics, with a smaller number of parameters. To that end, we try to describe plasmid division with only one parameter (instead of three), specifically, with a constant plasmid division rate throughout the course of the experiment. In the three parameter model, the plasmid division dynamics are dominated by the first division rate (see the full curve in Figure 4B). Accordingly, we test if the late decrease in the plasmid division rate is dispensable in terms of explaining the data, by assuming that the obtained value of the first plasmid division rate (describing the early plasmid dynamics) remains constant during the whole course of the experiment (see the dashed line in Figure 4B).

This one-parameter assumption clearly leads to a poorer agreement of the model with the data points in the late phase of the experiment (Figure S1A,B in the Supplementary Material), emphasizing that a finer description of plasmid division dynamics is important for explaining the experiment. Further, we model plasmid division dynamics with one parameter which now becomes free, and

fit the whole gene expression model to the experimental data. However, such a model also cannot provide a good fit to the data (Figure S1C,D), where the disagreement is particularly pronounced for M expression dynamics (Figure S1C)—note that F (and corresponding P) values for comparison with fits in Figure 3) is provided in caption of Figure S1. Overall, neither of alternative possibilities can provide good agreement of the model with the data, implying that the original (three-parameter) model is indeed necessary (minimal) to explain the data.
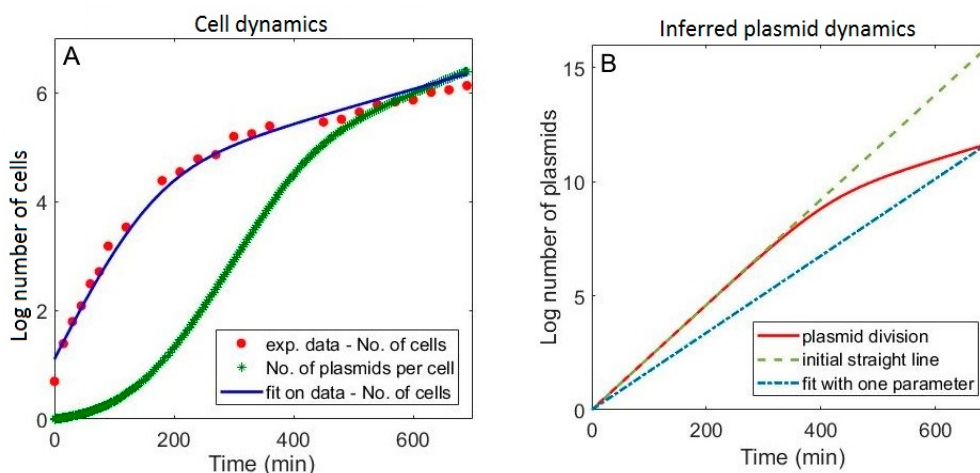
Figure 4. Population (cell and plasmid) dynamics. (**A**) Experimentally measured log number of cells vs. time (red dots) fitted by equation (12) (blue curve). The average number of plasmids per cell in time (green crosses), is shown on the same graph to allow comparison; (**B**) Solid red curve corresponds to the plasmid dynamics from Equation (13), described by three parameters (in analogy to the experimentally observed cell dynamics). Green dashed line represents the plasmid dynamics generated by the constant plasmid division rate, which corresponds to extrapolating the early part of the red curve to the end of the experiment. Blue dash-dotted line also corresponds to the assumption of constant plasmid division rate, where the division rate is taken as a free parameter obtained by fitting the model to R and M dynamics data.

To further test to what extent the included population dynamics are necessary to explain the data, we assess their contribution in establishing the final protein expression pattern. To this end, cell and plasmid division rates were set to zero in the full model, so that the model only describes specific gene regulation in the system. Simulation of the model in this case results in notably qualitatively changed protein dynamics for both M and R, which poorly fit the data (Figure 5). As M in the model is very stable, its amount decreases very slowly, governed by transcription repression by the C protein (Figure 5A). As expected, an increase in the amount of M later in the experiment cannot be recovered by the model which includes transcription regulation alone.

If one would attempt to explain the increase in the amount of M later in the experiment through transcription regulation alone, a non-existent activation of the P_M promoter at higher C protein concentrations would have to be invoked. Furthermore, one may misinterpret the observed rapid M decrease after the peak, as an indication of highly cooperative repression of P_M, since a trademark of high cooperativity is a sharp transition of the system from the ON to the OFF (or vice-verse) state [30]. However, this dynamical property is clearly a result of the population effects, since the Figure 5A suggests that without the population dynamics the peak in M dynamics would be much broader. The same (even more drastic) result is obtained if the model without population dynamics would be refitted to the experimental data (see Figure S2B).

With regards to the R dynamics, if the population dynamics effects are neglected, R slowly increases to some saturation value (Figure 5B), while the shape of the curve is concave instead of the convex form observed in the experiment. Namely, the rapid increase of R late in the experiment is due to two population dynamics effects, which overcome repression by C exhibited at later times.

In particular, (i) the rate of cell division slows down at later times (Figure 4B), lowering the effective R decay rate, and (ii) the number of plasmids per cell keeps increasing, leading to increased generation of R transcripts. Both of these effects clearly promote the increase of R at later times counteracting the repression by C, which provides an explanation for the experimentally observed time dependence.

To further check if the model describing only internal system regulation can explain the observed dynamics, it was re-fitted to the experimental data of measured protein expression (Figure S2 in Supplementary material). It can be analytically shown that the observed R dynamics can be explained by such a model only under conditions that transcripts and proteins are very stable and that there is no regulation of the P_{CR} promoter by C protein (Figure S2A). Clearly, this good fit does not provide a correct picture of experimentally inferred regulation in the system. In addition, this model gives a very poor fit to the experimental data for M expression (Figure S2B)—for M the adjusted R^2 corresponding to 0.22, with F value for the fit comparison with Figure 3A of $P \sim 10^{-13}$. Together, these results strongly suggest that the observed Esp1396I system expression dynamics cannot be explained solely by the system specific regulation, and provide an additional argument that the data do not become over fitted by including the population dynamics effects.
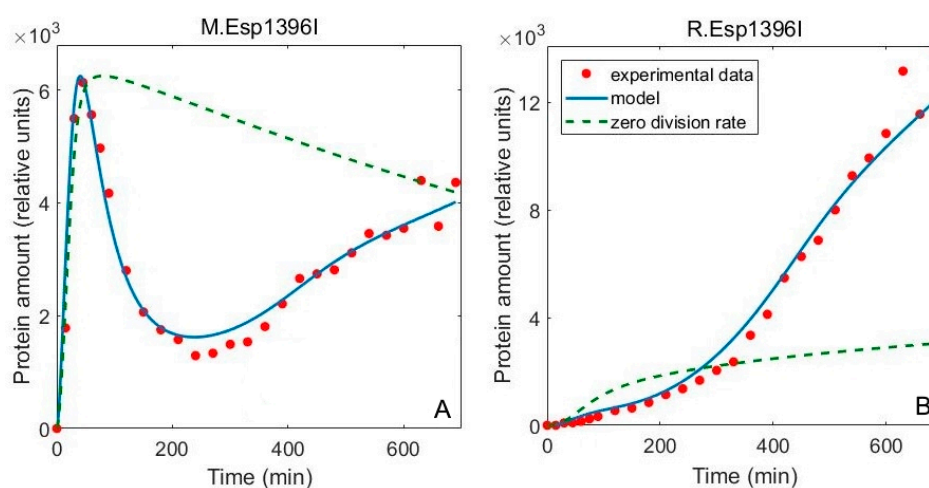

Figure 5. Excluding population dynamics from the full model. The blue solid line represents the full model fitted to experimental data (red dots) for (**A**) methyltransferase and (**B**) restriction endonuclease, while the green dashed line is obtained by excluding the population dynamics from the model and running the simulation with unchanged values for the rest of the parameters. To better compare the shape of the curves, the model without the population dynamics is rescaled, so that the maximal value reached by M is the same as in the full model.

Overall, it can be argued that population dynamics, while often neglected, significantly influences the expression dynamics of the system, and is crucial for understanding the system dynamics. Ignoring such dynamics can lead to identification of non-existent regulatory mechanisms of gene transcription in an effort to intuitively interpret experimental data.

3.3. Falsely Identifying Regulation when Plasmid Dynamics is not Included

We next analyse how neglecting only the plasmid division dynamics affects the observed kinetics of protein synthesis. Instead of focusing on quantitative comparison of the model predictions with the data, we will focus on how neglecting the plasmid dynamics can lead to false identification of regulatory mechanisms that in reality do not exist.

We start by observing how modeling agrees with M dynamics data when plasmid dynamics are included (full line), and when they are excluded from the model (dashed line), which is shown in Figure 6. After ~200 min the M amount in the model without plasmid dynamics slowly decreases over time, while the experimental data show that the amount of this enzyme starts to increase.

Consequently, one can speculate that a single, major effect behind the increase of M in the second interval of the experiment is plasmid dynamics, i.e., increase in the plasmid numbers late in the experiment. Neglecting plasmid dynamics in modeling can thus lead to (qualitative) misinterpretation of experimental results. For example, an increase in the amount of M later in the experiment can be interpreted as a consequence of non-existent P_M activation at higher concentrations of C protein, while in reality there is only repression [24].
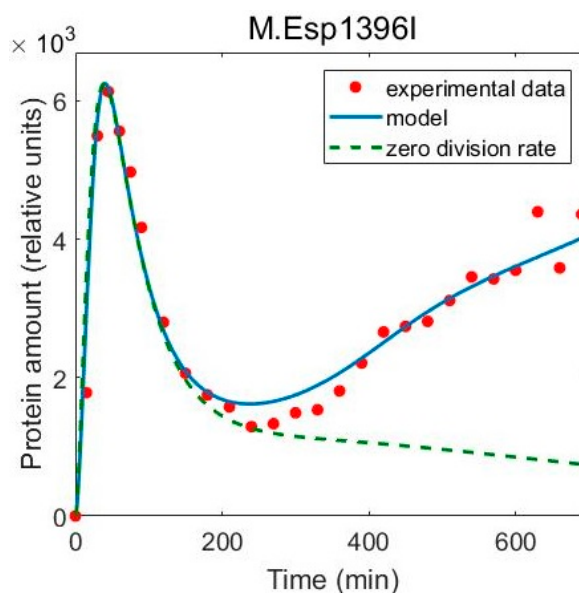
Figure 6. Effect of plasmid dynamics on intracellular M amounts. Experimentally measured intracellular M.Esp1396I amounts at different time points after transformation are denoted by red dots. The solid blue curve is the full model prediction, while the dashed green curve is obtained by setting the plasmid division rate in the full model to zero.

Similar holds when explaining the properties of R dynamics (Figure 7A): excluding plasmid replication from the model describing R expression results in a visibly different dynamics curve from that observed experimentally. Notably, experimental data for R dynamics can be empirically modelled with a quadratic equation. This time dependence can be derived analytically in the absence of population dynamics effects, if proteins and transcripts are assumed entirely stable, and if there is no regulation by C protein (see the pink curve in Figure 7B). Since a specific regulatory mechanism exploiting C proteins was confirmed in a number of previous experiments on the Esp1396I system [24,32,33], such a derivation obviously leads to a wrong conclusion about system expression control. Furthermore, the measured R dynamics naively look as arising from only transcription activation by C protein operating in the P_{CR} promoter control (as R amounts monotonically increase with time), but from a quantitative viewpoint, modeling such a case provides a significantly worse fit to the data (purple curve in Figure 7B)—F value for the comparison with the fit in Figure 3A is 25 ($P \sim 10^{-7}$). Therefore, intuitive reasoning may be in disagreement with a true situation in a cell.

R dynamics are, therefore, a nice example for apparently simple (quadratic) time dependence generated by a complex interplay between very different opposing effects, in particular, those arising from intracellular regulation (repression at higher C concentrations) and population dynamics (the increase in plasmid numbers later in the experiment). This interplay could not be inferred intuitively, since an intuitive interpretation would suggest only activation of the *cr* operon by C protein. Moreover, it could not be inferred even through analytical derivations, since they imply a constitutive (non-regulated) gene expression, and stable RNA and protein amounts, which is very different from reality. Therefore, an accurate understanding of the system regulation and the resulting dynamics requires taking into account all relevant effects, and their careful computational modeling.
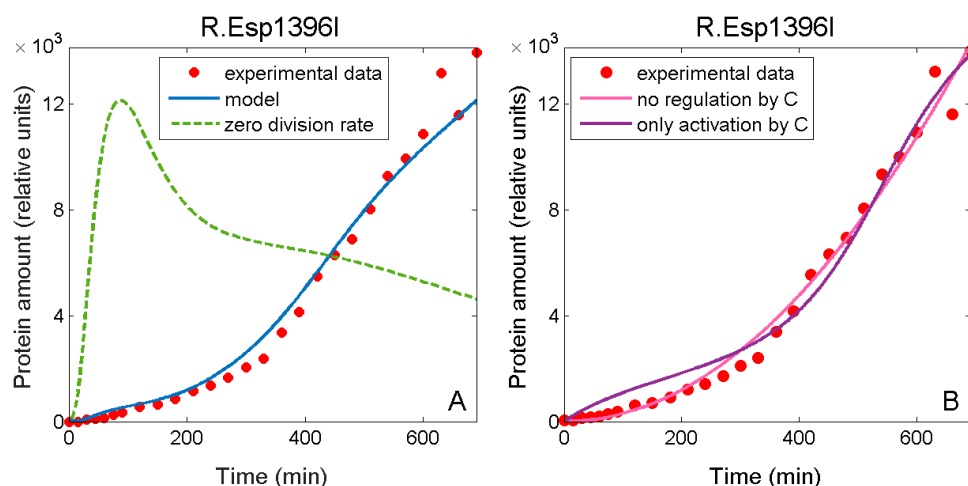
Figure 7. Effects of neglecting plasmid dynamics on intracellular R amounts. (**A**) Fit of the full model, taking into account plasmid division dynamics, to R data (red dots) is given by a solid blue curve, while the green dashed curve is obtained by setting the plasmid division rate to zero—the green curve is rescaled so that it can be compared with the blue one; (**B**) Two attempts to fit the R.Esp1396I data relying on intuitive expectations based on the form of the experimentally measured trajectory (all population dynamics effects are here neglected): (i) the model implied by the analytical derivation, which completely ignores P_{CR} promoter regulation by C protein, generates the solid pink curve, (ii) the best fit of the model assuming standard rates of protein and transcript degradation and neglecting transcription repression by C protein is given by the solid purple curve.

4. Conclusions

An earlier minimal model of the Esp1396I R-M system expression predicted that the methyltransferase amount changes oppositely in time to what was experimentally observed in microcolonies grown from single transformed cells. The disagreement could have been interpreted as a consequence of unknown regulatory mechanism(s) operating in the system, that were not accounted for in the model. However, from our analysis it follows that the reason behind this disagreement are the commonly neglected population dynamics effects—namely, kinetics of cell division and plasmid replication—which, when included, lead to a very good agreement with the data. Consequently, neglecting global physiological effects on gene expression can lead to falsely identified regulation by significantly impacting qualitative properties of intracellular protein expression dynamics.

From a computational perspective, we note that including population dynamics in the model, which is necessary to explain the experimental data, is a highly nontrivial task, as it inevitably increases dimensionality of the parameter inference. Still, we showed that this problem can be approached through a procedure in which the shared population dynamics parameters are initially estimated by considering dynamics of only one of the multiple species coupled by the regulation mechanisms, relying on approximated dynamics of the rest of the species. Such a "mean field-like" procedure can become even more necessary when considering larger gene regulatory networks, as dimensionality of the parameter inference problem would further increase from the one considered here. While the procedure is here introduced in the form that can be directly applied to any number of molecular species, it remains to be tested in practice when applied to more complex regulatory networks.

As an outlook, we have seen here that population dynamics effects, which are modulated by changes in global physiological factors, can have a significant effect on gene expression dynamics. Consequently, expression patterns of molecular species within a cell can result from a complex interplay of intracellular regulation and population dynamics effects, as demonstrated here in the case of the Esp1396I system. This then calls for advanced methods to reverse-engineer intracellular regulation from dynamical data, which would take into account both intracellular regulation and effects due to changing external conditions [13,18,20,34]. One approach used to filter out the effects of the global

physiological state of the cell, which directly impact population dynamics parameters, involves experimentally tracking expression from a constitutive and from a regulated promoter in parallel [16]. In addition, varying of gene copy number with time (e.g., due to kinetics of plasmid division) should be accounted for by both experiments and theoretical models. A particular challenge would be to reconstruct regulatory networks from sole protein dynamics data, that are becoming more and more available [35]. Development of advanced theoretical methods for such reconstruction, analogous to those for reverse engineering of gene networks from gene expression data, may thus become a necessity in the future.

Supplementary Materials: The following are available online: Figures S1 and S2.

Author Contributions: Conceptualization, M.D. (Marko Djordjevi), M.D. (Magdalena Djordjevic) and K.S.; Methodology, M.D. (Marko Djordjevi), S.G. and A.R.; Writing code, S.G. and A.R.; Validation, M.D. (Marko Djordjevi), M.D. (Magdalena Djordjevic). and K.S.; Formal Analysis, S.G., A.R. and M.D. (Marko Djordjevi); Investigation, S.G. and A.R.; Writing—Original Draft Preparation, A.R. and S.G.; Writing—Review & Editing, all authors; Visualization, S.G., A.R. and M.D. (Magdalena Djordjevic); Supervision, M.D. (Marko Djordjevi); Project Administration, M.D. (Marko Djordjevi) and M.D. (Magdalena Djordjevic); Funding Acquisition, M.D. (Marko Djordjevi) and M.D. (Magdalena Djordjevic).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Longo, D.; Hasty, J. Dynamics of single-cell gene expression. *Mol. Syst. Biol.* **2006**, *2*, 64–73. [CrossRef] [PubMed]

2. Ohno, M.; Karagiannis, P.; Taniguchi, Y. Protein expression analyses at the single cell level. *Molecules* **2014**, *19*, 13932–13947. [CrossRef]

3. Elowitz, M.B.; Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **2000**, *403*, 335–338. [CrossRef] [PubMed]

4. Rosenfeld, N.; Young, J.W.; Alon, U.; Swain, P.S.; Elowitz, M.B. Gene regulation at the single-cell level. *Science* **2005**, *307*, 1962–1965. [CrossRef] [PubMed]

5. Young, J.W.; Locke, J.C.; Altinok, A.; Rosenfeld, N.; Bacarian, T.; Swain, P.S.; Mjolsness, E.; Elowitz, M.B. Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. *Nat. Protoc.* **2012**, *7*, 80–97. [CrossRef] [PubMed]

6. Morozova, N.; Sabantsev, A.; Bogdanova, E.; Fedorova, Y.; Maikova, A.; Vedyaykin, A.; Rodic, A.; Djordjevic, M.; Khodorkovskii, M.; Severinov, K. Temporal dynamics of methyltransferase and restriction endonuclease accumulation in individual cells after introducing a restriction-modification system. *Nucleic Acids Res.* **2016**, *44*, 790–800. [CrossRef]

7. Kobayashi, I. Behavior of restriction–modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res.* **2001**, *29*, 3742–3756. [CrossRef]

8. Mruk, I.; Kobayashi, I. To be or not to be: Regulation of restriction–modification systems and other toxin–antitoxin systems. *Nucleic Acids Res.* **2013**, *42*, 70–86. [CrossRef]

9. Nagornykh, M.; Bogdanova, E.; Protsenko, A.; Solonin, A.; Zakharova, M.; Severinov, K. Regulation of gene expression in a type II restriction-modification system. *Russ. J. Genet.* **2008**, *44*, 523–532. [CrossRef]

10. Bogdanova, E.; Djordjevic, M.; Papapanagiotou, I.; Heyduk, T.; Kneale, G.; Severinov, K. Transcription regulation of the type II restriction-modification system AhdI. *Nucleic Acids Res.* **2008**, *36*, 1429–1442. [CrossRef]

11. Rodic, A.; Blagojevic, B.; Zdobnov, E.; Djordjevic, M. Understanding key features of bacterial restriction-modification systems through quantitative modeling. *BMC Syst. Biol.* **2017**, *11*, 377–391. [CrossRef] [PubMed]

12. Klimuk, E.; Bogdanova, E.; Nagornykh, M.; Rodic, A.; Djordjevic, M.; Medvedeva, S.; Pavlova, O.; Severinov, K. Controller protein of restriction–modification system Kpn2I affects transcription of its gene by acting as a transcription elongation roadblock. *Nucleic Acids Res.* **2018**, *46*, 10810–10826. [CrossRef] [PubMed]

13. Stefan, D.; Pinel, C.; Pinhal, S.; Cinquemani, E.; Geiselmann, J.; De Jong, H. Inference of quantitative models of bacterial promoters from time-series reporter gene data. *PLoS Comput. Biol.* **2015**, *11*, e1004028. [CrossRef] [PubMed]

14. Klumpp, S.; Zhang, Z.; Hwa, T. Growth rate-dependent global effects on gene expression in bacteria. *Cell* **2009**, *139*, 1366–1375. [CrossRef] [PubMed]

15. Gerosa, L.; Kochanowski, K.; Heinemann, M.; Sauer, U. Dissecting specific and global transcriptional regulation of bacterial gene expression. *Mol. Syst. Biol.* **2013**, *9*, 658–668. [CrossRef] [PubMed]

16. Berthoumieux, S.; De Jong, H.; Baptist, G.; Pinel, C.; Ranquet, C.; Ropers, D.; Geiselmann, J. Shared control of gene expression in bacteria by transcription factors and global physiology of the cell. *Mol. Syst. Biol.* **2013**, *9*, 634–644. [CrossRef] [PubMed]

17. Savageau, M.A. Design of molecular control mechanisms and the demand for gene expression. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5647–5651. [CrossRef]

18. Klumpp, S.; Hwa, T. Bacterial growth: Global effects on gene expression, growth feedback and proteome partition. *Curr. Opin. Biotechnol.* **2014**, *28*, 96–102. [CrossRef]

19. Vilar, J.M.G.; Guet, C.C.; Leibler, S. Modeling network dynamics: The lac operon, a case study. *J. Cell Biol.* **2003**, *161*, 471–476. [CrossRef]

20. De Jong, H.; Geiselmann, J. Fluorescent Reporter Genes and the Analysis of Bacterial Regulatory Networks. In *Hybrid Systems Biology*; Maler, O., Halász, Á., Dang, T., Piazza, C., Eds.; Springer: Cham, Switzerland, 2015; pp. 27–50.

21. Wang, L.-Z.; Wu, F.; Flores, K.; Lai, Y.-C.; Wang, X. Build to understand: Synthetic approaches to biology. *Integr. Biol.* **2015**, *8*, 394–408. [CrossRef]

22. Zhang, C.; Tsoi, R.; You, L. Addressing biological uncertainties in engineering gene circuits. *Integr. Biol.* **2015**, *8*, 456–464. [CrossRef] [PubMed]

23. De Jong, H. Modeling and simulation of genetic regulatory systems: A literature review. *J. Comput. Biol.* **2002**, *9*, 67–103. [CrossRef] [PubMed]

24. Bogdanova, E.; Zakharova, M.; Streeter, S.; Taylor, J.; Heyduk, T.; Kneale, G.; Severinov, K. Transcription regulation of restriction-modification system Esp1396I. *Nucleic Acids Res.* **2009**, *37*, 3354–3366. [CrossRef] [PubMed]

25. Sneppen, K.; Zocchi, G. *Physics in Molecular Biology*; Cambridge University Press: Cambridge, UK, 2005.

26. Rodic, A.; Blagojevic, B.; Djordjevic, M. Systems Biology of Bacterial Immune Systems: Regulation of Restriction-Modification and CRISPR-Cas Systems. In *Systems Biology*; Rajewsky, N., Jurga, S., Barciszewski, J., Eds.; Springer Nature: Cham, Switzerland, 2018; pp. 37–58.

27. Shea, M.A.; Ackers, G.K. The OR control system of bacteriophage lambda: A physical-chemical model for gene regulation. *J. Mol. Biol.* **1985**, *181*, 211–230. [CrossRef]

28. Butcher, J.C. *Numerical Methods for Ordinary Differential Equations*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2016.

29. Lomax, R.G.; Hahs-Vaughn, D.L. *Statistical Concepts: A Second Course*, 4th ed.; Routledge: Abingdon, UK, 2013.

30. Phillips, R.; Kondev, J.; Theriot, J.; Garcia, H. *Physical Biology of the Cell*; Garland Science: New York, NY, USA, 2012.

31. Press, W.H.; Teukolsky, S.A.; Vetterling, W.T.; Flannery, B.P. *Numerical Recipes: The Art of Scientific Computing*, 3rd ed.; Cambridge University Press: Cambridge, UK, 2007.

32. McGeehan, J.; Ball, N.J.; Streeter, S.; Thresh, S.-J.; Kneale, G. Recognition of dual symmetry by the controller protein C. Esp1396I based on the structure of the transcriptional activation complex. *Nucleic Acids Res.* **2011**, *40*, 4158–4167. [CrossRef] [PubMed]

33. Martin, R.N.; McGeehan, J.E.; Ball, N.J.; Streeter, S.D.; Thresh, S.-J.; Kneale, G. Structural analysis of DNA–protein complexes regulating the restriction–modification system Esp1396I. *Acta Cryst.* **2013**, *F69*, 962–966. [CrossRef]

34.　Barzel, B.; Liu, Y.-Y.; Barabási, A.-L. Constructing minimal models for complex system dynamics. *Nat. Commun.* **2015**, *6*, 7186–7193. [CrossRef]

35.　Porreca, R.; Cinquemani, E.; Lygeros, J.; Ferrari-Trecate, G. Structural identification of unate-like genetic network models from time-lapse protein concentration measurements. In Proceedings of the 49th IEEE Conference on Decision and Control, Atlanta, GA, USA, 15–17 December 2010; pp. 2529–2534.

Sample Availability: Not applicable.