Quantitative examination of five stochastic cell-cycle and cell-size control models for Escherichia coli and Bacillus subtilis

Guillaume Le Treut\(^1\)*, Fangwei Si\(^2\), Dongyang Li\(^3\), and Suckjoon Jun\(^{2,4}\)*

\(^1\) Theory Group, Chan Zuckerberg Biohub, San Francisco, CA 94158, USA

\(^2\) Department of Physics, University of California, San Diego, La Jolla, CA 92093, USA

\(^3\) Howard Hughes Medical Institute, Division of Biology and Biological Engineering, Broad Center, California Institute of Technology, Pasadena, CA 91125, USA

\(^4\) Section of Molecular Biology, Division of Biology, University of California, San Diego, La Jolla, CA 92093, USA

* Correspondence: Guillaume Le Treut (guillaume.letreut@czbiohub.org) and Suckjoon Jun (suckjoon.jun@gmail.com)

Abstract

We examine five quantitative models of the cell-cycle and cell-size control in Escherichia coli and Bacillus subtilis that have been proposed over the last decade to explain single-cell experimental data generated with high-throughput methods. After presenting the statistical properties of these models, we test their predictions against experimental data. Based on simple calculations of the defining correlations in each model, we first dismiss the stochastic Helmstetter-Cooper model and the Initiation Adder model, and show that both the Replication Double Adder and the Independent Double Adder model are more consistent with the data than the other models. We then apply a recently proposed statistical analysis method and obtain that the Independent Double Adder model is the most likely model of the cell cycle. By showing that the Replication Double Adder model is fundamentally inconsistent with size convergence by the adder principle, we conclude that the Independent Double Adder model is most consistent with the data and the biology of bacterial cell-cycle and cell-size control. Mechanistically, the Independent Adder Model is equivalent to two biological principles: (i) balanced biosynthesis of the cell-cycle proteins, and (ii) their accumulation to a respective threshold number to trigger initiation and division.
1. Introduction

Quantitative microbial physiology is marked by close interactions between experiment and modeling since its birth in the mid 20th century (see (1) for a review of the history with extensive literature). In particular, bacterial cell-size and cell-cycle control has enjoyed rejuvenated interests in modeling with the advent of microfluidics techniques that allow tracking of thousands of individual cells over a hundred division cycles (see, for example, (2–5). Re-emerged from the new single-cell data is the adder principle (6–8), which states that individual cells grow by adding a fixed size from birth to division, independently from their size at birth. This principle has characteristic repercussions on cell size homeostasis. Specifically, upon perturbation, the cell size at birth relaxes toward its steady-state value according to a first-order recurrence relation with a correlation coefficient equal to 1/2 (8–10).

Although the adder principle was originally proposed and statistically tested almost three decades ago by Koppes and colleagues before its recent revival (9), its mechanistic origin has remained elusive until recently because direct experimental tests were not available for a long time (1, 11). Several models have been proposed so far (6, 8, 11–20), and we expect a consensus to emerge as more experimental data become available.

The main purpose of this article is to derive and present steady-state statistical properties of quantitative bacterial cell-cycle and cell-size control models that we are currently aware of and, where relevant, critically examine them against single-cell data from our lab's mother machine experiments accumulated over the last decade in E. coli and B. subtilis. These models are (i). the stochastic Helmstetter-Cooper model (sHC) (11), (ii). the initiation adder (IA) model (10, 13, 21), (iii). the replication double adder (RDA) model (17), (iv). the independent double adder (IDA) model (11), and (v). the concurrent cell-cycle processes (CCCP) model (22) and its stochastic version (14).

Some of these models are graphically illustrated in Figure 1. Briefly, the sHC model is a literal extension of the textbook Helmstetter Cooper model by allowing independent Gaussian fluctuations to each of the initiation mass, the \( \tau_{\text{cyc}} = C+D \) period (from initiation to division), and the cell elongation rate. The IA model assumes that replication initiation is the sole implementation point of cell-size control, and division is strictly coupled to initiation such that division is triggered after fixed \( \tau_{\text{cyc}} = C+D \) minutes have elapsed since initiation. The RDA model is similar to the IA model in that it also assumes that initiation is the reference point for cell-size control. Its main difference from the IA model is that it assumes division is triggered after the cell elongates a constant length per origin of replication \( \delta_{\text{id}} \), rather than a constant time, since initiation. In other words, \( \delta_{\text{id}} \) is the added size during the C+D period. Both the IA and the RDA models assume the initiation adder, i.e., the cell growth by a nearly fixed size per replication origin between two consecutive initiation cycles, irrespective of the cell size at initiation (initiation mass).

The IDA model also states that initiation and division are independently controlled by their own respective initiator proteins. However, the IDA model is based on mechanistic assumptions that these proteins are produced in a balanced manner (i.e., their production rate is the same as the growth rate of the cell), and initiation and division are triggered when the cell has accumulated the initiator proteins to their respective threshold numbers. The CCCP model states that replication cycle and division cycles progress independently, but checkpoints or their equivalent are activated to ensure cell division (22).

This article is structured as follows. In Section 2, we summarize the five models and derive some of their statistical properties. In Section 3, we test the predictions of these models against the data. In Section 4, we critically examine one of the recent correlation analysis methods (the \( I \)-value analysis) used to justify the RDA model. We conclude that the IDA model is as of today the model most consistent with data, which also provides a falsifiable mechanistic picture.
Figure 1: Physiological parameters that can be measured from single-cell experiments. **a.** Time-lapse images of a single *Escherichia coli* cell growing in a microfluidics channel. The cell boundaries are segmented from phase contrast images whereas the replication forks are visualized using a functional fluorescently labeled replisome protein (DnaN-YPet). **b.** Multifork replication: in most growth conditions, several replication cycles overlap. The direction of the arrows is not the direction of time, but to illustrate that the HC model’s core idea is to trace replication initiation backward in time by $C+D$ from division. **c.** Four models of *E. coli* cell cycle and their control variables, which can be measured from single-cell experiments. The sHC model describes cell size and cell cycle using three parameters: elongation rate $\lambda = \frac{d\ln(l)}{dt}$, where $l$ is the cell length (not shown), $\tau_{cyc} = C+D$, and the initiation size per origin of replication $s_i$. The IA model uses $\lambda$, $\tau_{cyc}$ and the added size per origin of replication between consecutive replication initiation events $\delta_i$. The RDA model uses $\lambda$, $\delta_i$ and the added size from initiation to division $\delta_d$. The IDA model uses $\lambda$, $\delta_i$ and the added size from birth to division $\Delta_d$. Note that both $\delta_d$ and $\tau_{cyc}$ can span multiple generations.
2. Statistical properties of five bacterial cell-size and cell-cycle control models

2.1 The stochastic Helmstetter-Cooper model (sHC)

The original HC model (23) is based on the experimental observation that the average duration of chromosome replication ("C period") can be longer than the average doubling time of the cells in fast-growing E. coli. In such growth conditions, E. coli must initiate a new round of replication before the ongoing replication cycle is completed. The core of the HC model is the recipe to trace replication initiation backward by two temporal parameters (the doubling time \( \tau \) and the duration of cell cycle \( \tau_{cyc} = C + D \)) and one spatial parameter (e.g., cell size at division or initiation). It was Donachie who showed that, if (i) \( \tau_{cyc} = C + D \) is invariant under different nutrient conditions and (ii) the average cell size increases exponentially with respect to the nutrient-imposed growth rate \( \lambda = \frac{2}{\ln n} \) as \( S = S \exp(a \lambda) \) (where \( S \) and \( a \) are constant, and \( S \) is the average cell size of a steady-state population), then the cell size at initiation per replication origin \( s_i \) (or, the "initiation mass") must be mathematically invariant in all growth conditions (24). This result was later generalized to all steady-state growth conditions with and without growth inhibition (25).

Since the original HC model is deterministic and can be defined in terms of \( \lambda \), \( \tau_{cyc} \) and \( s_i \) (or division size \( S_d \)), one possible extension to a stochastic version is by making the three physiological variables stochastic. Together, they completely determine cell sizes including the size at division, assuming perfectly symmetric division (Figure 1). For simplicity, we draw \( \lambda \), \( \tau_{cyc} \) and \( s_i \) at cell birth from a multivariate Gaussian distribution, which also encodes cross- and mother-daughter correlations in the covariances matrix (11).

The recursion relation for the cell size at division in this "stochastic" Helmstetter-Cooper (sHC) model can be written as follows:

\[
S_d^{(n)} = s_i^{(n)} e^{\lambda \tau_{cyc}^{(n)}}, \tag{Eq. 1}
\]

where \( n \) denotes the generation index. If we assume that cells elongate exponentially at the growth rate \( \lambda \) (2), the number of overlapping cell cycles \( p+1 \) is completely determined by the relation:

\[
S_d = 2 S_k \Leftrightarrow e^{\lambda \tau_{cyc}^{(n)}} = 2^{p+1} e^{-\lambda \tau_i}. \tag{Eq. 2}
\]

where \( \tau_i \) is the time duration elapsed between cell birth and replication initiation. It follows that \( p \) is the integer part of \( \tau_{cyc} / \tau \), where \( \tau = \ln 2 / \lambda \) is the generation time, so that \( p+1 \) is the number of overlapping cell cycles (unless noted otherwise we will adopt the convention that \( X^{(n)} \) denotes the value of a physiological variable in generation \( n \) whereas \( X \) is the average over the whole lineage).

In the sHC model, consecutive sizes at initiation are correlated through \( \rho_i = \rho(s^{(n)}_i, s^{(n+1)}_i) = 0 \), where \( \rho(A,B) \) stands for the Pearson correlation coefficient between variables \( A \) and \( B \). In the absence of mother-daughter correlations for all three physiological variables, the cell should behave as a sizer, \( \rho_i = \rho(S^{(n)}_d, S^{(n+1)}_d) = 0 \). However, additional cross- or auto-correlations among \( \lambda \), \( \tau_{cyc} \) and \( s_i \) (such as cross-correlations between \( s^{(n)}_i \) and \( \tau_{cyc}^{(n)} \) and/or mother-daughter correlations between \( s^{(n)}_i \) and \( s^{(n+1)}_i \)) can have a non-trivial effect on size homeostasis. Analytical expressions for \( \rho_i \), \( \rho_{ij} \) and \( \rho_{ij} = \rho(s^{(n)}_i, s^{(n+1)}_j) \) are derived in Appendix C. Importantly, \( \rho_i \) is particularly sensitive to the mother-daughter initiation-size correlation \( \rho_i = \rho(s^{(n)}_i, s^{(n+1)}_i) \) in the sHC model (11). This observation motivated an experimental study aiming at perturbing \( \rho_i \) by periodic expression of DnaA to break the adder phenotype in E. coli. This prediction was refuted in the experiments (11), since despite the perturbations to \( \rho_i \) by periodic oscillations of \( \text{dnaA} \) expression level, E. coli maintained its size homeostasis following the adder behavior (11). An important conclusion from the oscillation experiments is that replication initiation and cell division are independently controlled in steady-state conditions in both E. coli and Bacillus subtilis, thus firmly refuting the particular version of the sHC model.

2.2. The initiation adder model (IA)

The IA model is a variant of the sHC model in which the constraint on the initiation mass \( s_i = \text{constant in all growth conditions} \) is replaced by an adder mechanism running between consecutive replication initiations (21, 26). Specifically, the cell initiates replication following the adder principle, i.e., the size added per origin between two consecutive
initiation cycles, \( \delta_i \) is independent of the cell size at initiation (13). Yet, as in the sHC model, the IA model assumes that division is triggered after a fixed duration of time, \( \tau_{cyc} \), has elapsed since initiation. The three stochastic control parameters in the IA model are therefore: \( \lambda \), \( \tau_{cyc} \) and \( \delta_i \). A given cell size at replication initiation determines the next replication initiation event and one division event.

The recursion relation for cell size at division is the same as in the sHC model (Eq. 1). However, this relation is replication initiation event and one division event.

In the RDA model, the cell size per origin of replication, \( s_i \), is measured in generation \( n \). For two overlapping cell cycles (i.e. generation \( n-1 \)) as shown in Figure 1. Cells are born with 2 origins of replications, therefore we have \( S_{id}^{(n)} = 2S_i^{(n)} \). In this example, the mass synthesized between two consecutive replication initiation events must take into account one division event. Back to the RDA model, and using the same convention, the total added size from initiation to division is \( 2S_{d}^{(n)} - S_{id}^{(n)} = 4 \delta_{id}^{(n)} \). The factor of 4 accounts for the 4 origin of replications present after replication initiation. Dividing by 2, we obtain Eq. 4. This reasoning generalizes to any number of overlapping cell cycles. From Eq. 4, we also obtain that the average cell size at division is \( S_d = 2(\delta_i + \delta_d) \). An argument similar to Eq. 2 yields the number of overlapping cell cycles \( \rho + 1 \) as a function of the mean of the physiological variables: \( \rho \) is the integer part of \( \log_2(1 + \delta_d/\delta_i) \).

The IA model is not compatible with size convergence by the adder principle. While \( \rho = 1/2 \) as in the IA model, the division size mother-daughter correlation is given by

\[
\rho_{id} = \frac{1}{2} \left( 1 + 3\sigma_{id}^2 / \sigma_d^2 \right)^{-1} < \frac{1}{2},
\]  

(Eq. 5)

where \( \sigma_i^2 \) and \( \sigma_d^2 \) are the variances for \( \delta_i \) and \( \delta_{id} \) respectively (see Appendix A). Since the adder principle is equivalent to \( \rho_{id} = \gamma_2 \) (see Appendix A), the IA model converges to the adder only in the deterministic limit \( \sigma_d \to 0 \). In addition, we can also compute the correlation between initiation size per origin and division size \( \rho_{id} = \rho(S_i^{(n)}, S_d^{(n)}) \) and obtain:
\[ Q_{id} = \left( 1 + \frac{3 \ln 2}{\sqrt{n_d}} \right)^{-1/2}. \]  \hspace{1cm} \text{(Eq. 6)}

### 2.4. The independent double adder (IDA) model

The independent double adder (IDA) model states that, in steady state, initiation and division independently follow the adder principle. That is, the size added between two consecutive initiations is independent of the size at initiation (as in the IA and RDA models), whereas the size added between two division cycles is independent of the cell size at birth (or division). The recursion relation for the division size can be written as:

\[ S_d^{(n+1)} = \frac{1}{2} S_d^{(n)} + \Delta_d^{(n)}. \]  \hspace{1cm} \text{(Eq. 7)}

It follows that the average cell size at division is \( S_d = 2 \delta_d \). An argument similar to Eq. 2 yields the number of overlapping cell cycles \( p+1 \), where \( p \) is the integer part of \( \log_2(\delta_d/\delta_i) \).

We have \( \rho_i = 1/2 \) and \( \rho_d = 1/2 \) as expected by the definition of the model. Furthermore, since initiation and division follow two independent processes (Eqs. 3 and 7), division and initiation sizes are independent from each other, namely \( \rho_d = 0 \).

Mechanistically, the IDA model is based on (i) balanced biosynthesis of cell-cycle proteins and (ii) their accumulation to respective threshold numbers to trigger initiation and division (11).

### 2.5. The concurrent cell-cycle processes model (CCCP)

The CCCP model is an adaptable model with several adjustable parameters (as in the sHC model) and lies somewhere in between the IA and the IDA model. The adaptability is analogous to the presentation by Amir (10) so that the model can be continuously adjusted between sizer and timer depending on the mother-daughter size correlations between -1 and +1. To ensure 1:1 correspondence between the replication cycle and the division cycle, the model explicitly implements a constraint that division must wait until after replication termination. Biologically, the model follows the view by Boye and Nordström (22).

We discuss the specific case of the adder by fixing the mother-daughter size correlation coefficients to 1/2 as explained throughout this section 2. That is, the cell size at initiation follows the recursion relation:

\[ \ln(S_i^{(n+1)}) = \frac{1}{2} \ln(S_i^{(n)}) + A^{(n)}. \]  \hspace{1cm} \text{(Eq. 8)}

where \( A^{(n)} \) is the logarithmic added size between consecutive replication initiations. As mentioned above, the CCCP model was originally introduced in a more general form than Eq. 8, with an adjustable correlation parameter (see Appendix D). However, as explained by the authors a value of 1/2 is the most consistent with experimental data. Eq. 8 is very similar to Eq. 2: it is an adder on the logarithmic sizes rather than on the actual sizes at replication initiation.

Denoting \( C \) the time to replicate the chromosome, a candidate size for the division size is:

\[ \ln(S_R^{(n)}) = \ln(S_i^{(n)}) + \lambda C^{(n)}. \]  \hspace{1cm} \text{(Eq. 9)}

where as before \( \lambda \) is the elongation rate. If chromosome replication was the only process determining the size at division, \( S_R^{(n)} \) would be the division size. However, another process, namely the division adder, is constraining the division size, resulting in a second candidate size for division:

\[ \ln(S_H^{(n)}) = \frac{1}{3} \ln(S_H^{(n-1)}) + B^{(n)}. \]  \hspace{1cm} \text{(Eq. 10)}

where \( B \) is the added logarithmic size between consecutive division adder cycles. Eq. 10 is similar to Eq. 7 and represents the division adder. Finally, cell size at division is determined by the slowest of the two processes from Eqs. 9 and 10:

\[ \ln(S_d^{(n)}) = \max(\ln(S_R^{(n)}), \ln(S_H^{(n)})). \]  \hspace{1cm} \text{(Eq. 11)}

Equation 11 simply means that division should start only after replication termination. Denoting \( f \) as the fraction of cases in which division size is limiting (namely \( S_R < S_d = S_d \)), the average time elapsed between replication initiation and cell division can be expressed as (assuming that \(< \ln(x) >= \ln(< x >) \) (21)):
\[\tau_{\text{cyc}} = (1 - f)C + \frac{2(B-A)}{\lambda}, \quad \text{(Eq. 12)}\]

where \(A, B, C\) stand for means. Therefore, the number of overlapping cell cycles \(\rho+1\) is determined by Eq. 2. Eq. 12 has a functional dependence on growth rate compatible with experimental reports (16).

**Table 1** - Summary of a few models of the *E. coli* cell cycle. The definition column indicates the equations defining the division and replication cycles. The control parameters are summarized in the next column. In the three rightmost columns we give the three correlations \(\varrho, \rho_d\), and \(\rho_d^2\). We have used the following variables: (i) \(\sigma_{\text{id}}^2\): variance of \(\delta_{\text{id}}\), (ii) \(\sigma_{\text{id}}^2\): variance of \(\delta_{\text{id}}\), (iii) \(\mu\): mean of \(s_i\), (iv) \(\sigma_i^2\): variance of \(s_i\), (v) \(\mu_i\): mean of \(\text{exp}(\lambda \tau_{\text{cyc}})\), (vi) \(\sigma_i^2\): variance of \(\alpha\), (vii) \(\eta_i = \sigma_i/\mu_i\) is the coefficient of variation (CV) of \(s_i\), (viii) \(\eta_{\text{id}} = \sigma_{\text{id}}/\mu_{\text{id}}\) is the CV of \(\alpha\), (ix) \(\mu_{\text{id}}\): mean of \(\ln(S_{\text{id}})\), (x) \(\sigma_{\text{id}}^2\): variance of \(\ln(S_{\text{id}})\), (xi) \(\mu_{\text{id}}\): mean of \(\ln(S_{\text{id}})\), (xii) \(\sigma_{\text{id}}^2\): variance of \(\ln(S_{\text{id}})\).

### 2.6 Similarities and differences between the shC, IA, RDA, IDA, and the CCCP models.

The question of implementation point for cell size control has been controversial in the past. In the shC, IA, and RDA models, replication initiation is the implementation point of cell size control. By contrast, the IDA and CCCP models assume that the division and replication cycles are controlled by independent processes.

These models reflect a major challenge for identifying a cell-size control model that is compatible with the new plethora of high-throughput single-cell data (2, 16). Although the shC and IA models can be dismissed by experimental evidence (Section 3.1), the other models require more thorough analysis. For example, in contrast to the IDA model, the RDA model only ensures the initiation adder and it only reproduces the division adder behavior in the deterministic limit where \(\delta_{\text{id}}\) is constant. The essential difference between the IDA and RDA models comes from the correlation between the size per oriC at initiation and the added size per oriC from initiation to division. Specifically, \(\rho(s_i, \delta_{\text{id}})\) is null for the RDA model whereas it takes negative values for the IDA.

In the next section, we test these models against data in more detail.
3. Test of the models against data

3.0 Description of the experimental data used in this study

We used datasets from our previous studies for *E. coli* and *B. subtilis* (11, 27). We have also performed additional experiments for this study (see Methods). All data and numerical analysis are available (28). In total, we have 15 experimental datasets from our studies. We have also analyzed the 4 experimental datasets made available by Witz and colleagues in their work (17).

3.1 Comparison of correlations vs experimental data

We first set out to test the different cell-cycle and cell-size control models. Specifically, we computed the four correlations $\rho(S_b, \Delta d)$, $\rho(s_i, \delta_{id})$, $\rho(s_i, \delta_{ii})$, and $\rho(s_i, \tau_{cyc})$ (Figure 2). The correlation $\rho(S_b, \Delta d)$ is important because $\rho(S_b, \Delta d)=0$ defines the adder-based cell-size homeostasis. Indeed, $\rho(S_b, \Delta d)$ is zero in virtually all experimental data. The $\rho(s_i, \delta_{ii})$ is also close to zero, although deviations are seen for some experiments. These results suggest that both the IDA and RDA models are possible. By contrast, the $\rho(s_i, \tau_{cyc})$ correlation shows consistently a negative value. This refutes the sHC and IA models, which both assume that the initiation-to-division duration and the initiation size per origin are independent control parameters, and thus predict $\rho(s_i, \tau_{cyc}) = 0$. In addition, $\rho(s_i, \delta_{id})$ is also close to zero, in favor of the RDA model, but it is also slightly negative for several conditions, in agreement with the IDA model (Appendix B). We did not test the CCCP model because it is a model to be adjusted to the data. The candidate models are therefore the RDA and IDA models. Hereafter, we therefore focus on these two models.

![Figure 2](image-url) - The four correlations $\rho(S_b, \Delta d)$, $\rho(s_i, \delta_{id})$, $\rho(s_i, \delta_{ii})$ and $\rho(s_i, \tau_{cyc})$ are computed for the 4 experimental datasets by Witz and colleagues, and 15 experimental datasets that we produced (see Methods). While the first three vanish for most experimental data, the $\rho(s_i, \tau_{cyc})$ displays a consistent negative correlation, inconsistent with the HC and IA models.

3.2 Statistical analysis and the case study of the $I$-value analysis

In their recent paper (17), Witz *et al.* tracked replication and division cycles at the single-cell level, using experimental methods similar to previous works (11, 16, 29). They computed correlations between all pairs of measured physiological parameters, and attempted to identify the set of most mutually uncorrelated physiological variables by computing the “$I$-value”, a metric to measure the statistical independence of the measured variables. They then assumed that statistically uncorrelated physiological parameters must represent biologically independent controls. Such approaches previously facilitated the discovery of the adder principle and its formal description (8). Based on this correlation
analysis or I-value analysis, Witz et al. concluded that the RDA model is the most likely model of the cell-cycle and cell-size control.

To compute the I-value for a given model, one needs to identify the control parameters of the model and their characteristic features. For the RDA model, they are (i) the absence of correlation between the cell size at initiation and the added size between initiation and division, namely $\rho(s, \delta_i) = 0$, and (ii) the absence of correlation between the cell size at initiation and the added size between consecutive initiations, namely $\rho(s, \delta_j) = 0$. In addition to these size variables, it is known that the growth rate is mostly independent of the other physiological variables, namely: $\rho(\lambda, \delta_i) = 0$ and $\rho(\lambda, \delta_j) = 0$. Witz et al. hence proposed a scalar metric that summarizes these four correlations being equal to zero, namely the determinant $I$ (or I-value) of the matrix of correlations between the 4 variables $s$, $\delta_i$, $\delta_j$ and $\lambda$ (Eq. 13). When $I < 1$, some cross-correlations exists and both $\rho(\lambda, \delta_i)$ and $\rho(\lambda, \delta_j)$ cannot vanish. On the other hand, when $I = 1$, the RDA model holds. Although since the work by Cooper and Helmstetter (23) it has been known that the progression of cell size and cell cycle can be completely described using three variables (16, 25), four variables are necessary here to encompass the correlation structure characterizing the RDA model, as explained by Witz et al. (17, 30). In summary, to measure the statistical independence of each set of parameters, the I-value analysis needs a correlation matrix of the following form (Eq. 13).

$$
\begin{pmatrix}
1 & \rho_{12} & \rho_{13} & \rho_{14} \\
\rho_{12} & 1 & \rho_{23} & \rho_{24} \\
\rho_{13} & \rho_{23} & 1 & \rho_{34} \\
\rho_{14} & \rho_{24} & \rho_{34} & 1 \\
\end{pmatrix}
$$

(Eq. 13)

The diagonal elements are 1's and the off-diagonal elements are cross-correlations between pairs of parameters. Therefore, if all parameters are statistically independent of each other, the off-diagonal elements should be 0, and the determinant $I$ of the matrix should be 1. Based on this observation, Witz et al. used the determinant $I \leq 1$ of the matrix as a metric for statistical independence of the hypothetical control parameters, with $I=1$ being the set of most independent parameters.

In Section 4, we will come back later to some of the limitations of the I-value analysis.

3.3. Test of the I-value analysis with various models

Our 4-variable I-value analysis of the models described in Section 2 (except the CCCP) for all 19 datasets is shown in Figure 3 (top). We computed I-values for the following four models: RDA, IDA, IA and HC (see Methods). The results of the analysis indicate that 7 out of our 15 experiments support the IDA rather than the RDA model, whereas 1 supports the IA model. Furthermore, when we applied the same analysis to all 4 datasets from Witz et al., we found that all 4 experiments support the IDA model. Note that Witz et al. had only analyzed one dataset (17) The sHC and IA models were included for completeness, and they show systematically a lower I-value (except in one B. subtilis condition in which the IA models had the largest score for reasons we do not understand). Overall, the results of this analysis suggest that the IDA model is most consistent with the 19 datasets we have analyzed.

4. Discussion

4.1 Limitations of the I-value analysis

We noticed that the I-values for the RDA and IDA models were in general very close, although these two models point to two fundamentally different mechanisms of the cell cycle. We therefore asked to what extent the I-value analysis could be used to effectively identify meaningful models of the cell cycle. In the spirit of the ranking performed in the study by Witz et al., we considered all possible combinations of 4 among 18 physiological variables (see Methods; see Figure S1), and computed the I-values for each of Witz and colleagues’ 4 datasets and for each of our 15 datasets. Although the IDA and RDA models have high scores, we found that many other combinations have higher I-values, including combinations that do not correspond to any meaningful model of the cell cycle (Figure 3 bottom). Since I-values cannot be used to distinguish sound from unsound models of the cell-cycle, we conclude that this analysis lacks predictive power.
Figure 3 - We applied the I-value analysis proposed by Witz and colleagues to 4 models (RDA, IDA, IA and HC) using 15 experimental datasets that we produced (see Methods), and the 4 datasets published by Witz and colleagues. The length of the arrows indicate how much IDA or RDA model is favored by the data.

Furthermore, the I-value analysis can only be employed to compare models with the same number of variables in defining correlations. To see this, let us consider the RDA and IDA models in Figure 4. The RDA model can be defined by the 3 parameters \(\delta_{ii}, \delta_{id}, \lambda\) (Figure 4a). Indeed, from an initial condition consisting of an initiation size, only those 3 parameters need to be known at each generation to construct a whole lineage. Yet the defining correlations of the RDA model are given by \(\rho(\lambda, \delta_{id}) = 0\) and \(\rho(\lambda, \delta_{ii}) = 0\). Thus, we need a total of 4 variables \(\{s, \delta_{ii}, \delta_{id}, \lambda\}\) to characterize the RDA model, which leads to the 4x4 correlation matrix shown in Figure 4c.

The problem with the above procedure is that the size of the correlation matrix becomes model dependent, and thus the I-value analysis cannot compare different models. For example, following the same reasoning, the IDA model would require 5 variables for the I-value analysis, because of the two defining correlations \(\{s, \delta_{ii}\}\) and \(\{S_b, \Delta_d\}\) (Figure 4b). Therefore, in addition to the three independent control parameters \(\{\delta_{ii}, \Delta_d, \lambda\}\), the I-value analysis would require two additional parameters \(\{s, S_s\}\) from the defining correlations. The resulting correlation matrices would then be 5x5 from \(\{s, S_s, \delta_{ii}, \Delta_d, \lambda\}\) instead of 4x4 (Figure 4c). Since I-values obtained from correlation matrices of different sizes cannot be meaningfully compared, the I-value analysis is fundamentally limited to compare a specific class of models.

Finally, it is worthwhile mentioning that the I-value analysis employed by Witz et al. is only valid for non-overlapping cell cycles (30).
Figure 4 - Each model (a) is characterized by a set of two correlations (b). Note that these defining correlations require 1 additional parameter $S_b$ for the sHC model, 1 additional parameter $s_i$ for the RDA and the IA models, whereas the IDA model requires 2 additional parameters, $s_i$ and $S_b$. As a result, the covariance matrix for the $I$-value analysis of these models, according to Witz et al., would be 4x4 for the sHC, IA and RDA models and 5x5 for the IDA model. Therefore, the $I$-values of these models cannot be meaningfully compared.

4.2 The RDA model does not produce the adder principle

From Eq. 5, size convergence according to the RDA model is incompatible with the adder principle. More specifically, in the presence of fluctuations, the RDA model is skewed toward a sizer behavior. Using experimentally measured values for the variance of $\delta_{ii}$ and $\delta_{id}$, we computed $\rho_i$ according to Eq. 5 (Figure 5). The RDA model would predict a deviation from the adder principle, in contradiction to several experimental results (6, 8, 31, 32).

Witz et al.’s simulation of the RDA showed a good agreement with the experimental data (17). Yet further investigation showed that the agreement was a direct consequence of introducing yet another adjustable parameter, namely the variance of the septum position (31, 32). Indeed, for perfectly symmetric division, the simulation results also show deviations from experimental adder behavior (Figure S2), in agreement with Eq. 5.
Figure 5: The theory based on the initiation-centric model predicts a more sizer-like behavior. We used Eq. 1 and the experimental values of $\sigma_{id}/\sigma_{ii}$ from (17) and (11).

4.3 Mechanistic origin of the IDA model

The IDA model is a mechanistic model based on two experimentally verified hypotheses. First, the cell cycle proteins are produced in a balanced manner (i.e., the synthesis rate of each protein is the same as the growth rate of the cell):

$$\frac{dN}{dt} = c^* \frac{dV}{dt}, \quad \text{(Eq. 14)}$$

where $N$ is the protein copy number in the cell, $V$ is the cell volume and $c^*$ is the steady-state protein concentration. Second, initiation or division is triggered when the respective initiator protein reaches a threshold, namely:

$$N = N_0. \quad \text{(Eq. 15)}$$

The adder phenotype is a natural consequence of these two assumptions, provided that the initiator proteins are equally partitioned at division between daughter cells (11). The requirement that $N_0/2$ proteins must be synthesized between birth and division results in the added volume from birth to division to be:

$$\Delta_d = \frac{N_0}{2c^*}. \quad \text{(Eq. 16)}$$

This model was substantiated in an experimental study showing that perturbing the first condition, namely balanced biosynthesis, was enough to break the adder phenotype (11). Balanced biosynthesis was perturbed in two orthogonal ways: (i) by oscillating the production rate of the FtsZ protein through periodical induction and (ii) by relieving FtsZ degradation through ClpX inhibition.

A similar mechanism is thought to apply to the initiation process, through the initiator protein DnaA, which accumulates at the origin of replication to trigger replication initiation. An important difference with the division mechanism however is that this results in a threshold to be reached at each origin of replication, thus (Eq. 14) is modified to:

$$M = (#oriC) \times M_0. \quad \text{(Eq. 17)}$$

Provided that once again proteins are equally partitioned at division, Equations (2) and (4) result in a fixed added volume per origin of replication between consecutive initiation events, hence the initiation-to-initiation adder $\delta_{ir}$.

Perspective and concluding remarks

While applying a recently proposed correlation analysis, namely the I-value analysis, we became aware of some of its limitations, as explained above. In our view, this illustrates some of the caveats one may encounter when applying correlation analysis. While valuable in various contexts, correlation analysis can lead to erroneous conclusions when
additional sources of variability such as experimental and measurement errors are not properly taken into account in the analysis. For example, adder correlations can emerge from non-adder mechanisms due to measurement errors in the cell radius (33). Therefore, while correlation analysis is useful to confront models to experimental data, we believe it is important to seek a molecular understanding of a model of the cell cycle.

The question of whether the implementation point of the cell cycle is birth or replication initiation has a long history. Although cell-size control was initially thought to be division centric because the CV of the division size was smaller than that of the doubling time (34), many interpreted the HC model (23) and Donachie’s theoretical observation (24) as an initiation-centric view for long. The rediscovery of the adder principle, not explained within the HC model, has revealed the need to revisit models of the bacterial cell cycle. In this article, we have reviewed one of the latest controversies that has emerged in the field of quantitative bacterial physiology, namely the question of the implementation point of the cell cycle. Based on recent results and single-cell experimental data that we have generated over the last decade, we favor a mixed implementation strategy with two independent adders namely the IDA, as the most likely mechanism ruling the *E. coli* and *B. subtilis* cell cycle. Furthermore, the fact that both initiation and division share the same adder phenotype suggest to us that they also must share the same mechanistic principles. That is, initiation and division must require (i) balanced biosynthesis of their initiator proteins such as DnaA for initiation and FtsZ for division, and (ii) their accumulation to a respective threshold number to trigger initiation and division (11). Ultimately, these predictions should be tested experimentally to gain mechanistic understanding and their generality beyond correlation analysis.
References

1. S. Jun, F. Si, R. Pugatch, M. Scott, Fundamental principles in bacterial physiology-history, recent progress, and the future with focus on cell size control: a review. *Rep. Prog. Phys.* 81, 056601 (2018).

2. P. Wang, *et al.*, Robust Growth of Escherichia coli. *Current Biology* 20, 1099–1103 (2010).

3. J. R. Moffitt, J. B. Lee, P. Cluzel, The single-cell chemostat: an agarose-based, microfluidic device for high-throughput, single-cell studies of bacteria and bacterial communities. *Lab Chip* 12, 1487–1494 (2012).

4. Z. Long, *et al.*, Microfluidic chemostat for measuring single cell dynamics in bacteria. *Lab Chip* 13, 947–954 (2013).

5. H. Vashistha, M. Kohram, H. Salman, Non-genetic inheritance restraint of cell-to-cell variation. *Elife* 10, e64779 (2021).

6. M. Campos, *et al.*, A Constant Size Extension Drives Bacterial Cell Size Homeostasis. *Cell* 159, 1433–1446 (2014).

7. S. Jun, S. Taheri-Araghi, Cell-size maintenance: universal strategy revealed. *Trends Microbiol.* 23, 4–6 (2015).

8. S. Taheri-Araghi, *et al.*, Cell-Size Control and Homeostasis in Bacteria. *Current Biology* 25, 385–391 (2015).

9. W. J. Voorn, L. J. H. Koppes, N. B. Grover, Mathematics of cell division in Escherichia coli: comparison between sloppy-size and incremental-size kinetics. *Curr. Top. Mol. Gen* 1, 187–194 (1993).

10. A. Amir, Cell size regulation in bacteria. *Physical Review Letters* 112, 208102 (2014).

11. F. Si, *et al.*, Mechanistic Origin of Cell-Size Control and Homeostasis in Bacteria. *Curr. Biol.* 29, 1760–1770.e7 (2019).

12. S. Iyer-Biswas, *et al.*, Scaling laws governing stochastic growth and division of single bacterial cells. *Proc. Natl. Acad. Sci. U. S. A.* 111, 15912–15917 (2014).

13. A. Amir, Point of view: Is cell size a spandrel? *Elife* 6, e22186 (2017).

14. G. Micali, J. Grilli, M. Osella, M. C. Lagomarsino, Concurrent processes set E. coli cell division. *Sci Adv* 4, eaau3324 (2018).

15. L. K. Harris, J. A. Theriot, Relative Rates of Surface and Volume Synthesis Set Bacterial Cell Size. *Cell* 165, 1479–1492 (2016).

16. M. Wallden, D. Fange, E. G. Lundius, Ö. Baltekin, J. Elf, The Synchronization of Replication and Division Cycles in Individual E. coli Cells. *Cell* 166, 729–739 (2016).

17. G. Witz, E. van Nimwegen, T. Julou, Initiation of chromosome replication controls both division and replication cycles in through a double-adder mechanism. *Elife* 8 (2019).

18. H. Zheng, *et al.*, General quantitative relations linking cell growth and the cell cycle in Escherichia coli. *Nat Microbiol* 5, 995–1001 (2020).

19. F. Bertaux, J. von Kügelgen, S. Marguerat, V. Shahrezaei, A bacterial size law revealed by a
coarse-grained model of cell physiology. *PLoS Comput. Biol.* **16**, e1008245 (2020).

20. D. Serbanescu, N. Ojkic, S. Banerjee, Nutrient-Dependent Trade-Offs between Ribosomes and Division Protein Synthesis Control Bacterial Cell Size and Growth. *Cell Rep.* **32**, 108183 (2020).

21. P.-Y. Ho, A. Amir, Simultaneous regulation of cell size and chromosome replication in bacteria. *Front. Microbiol.* **6**, 662 (2015).

22. E. Boye, K. Nordström, Coupling the cell cycle to cell growth. *EMBO reports* **4**, 757–760 (2003).

23. S. Cooper, C. E. Helmstetter, Chromosome replication and the division cycle of Escherichia coli. *Journal of Molecular Biology* **31**, 519–540 (1968).

24. W. D. Donachie, Relationship between cell size and time of initiation of DNA replication. *Nature* **219**, 1077–1079 (1968).

25. F. Si, *et al.*, Invariance of Initiation Mass and Predictability of Cell Size in Escherichia coli. *Curr. Biol.* **27**, 1278–1287 (2017).

26. L. Sompayrac, O. Maaloe, Autorepressor model for control of DNA replication. *Nat. New Biol.* **241**, 133–135 (1973).

27. J. T. Sauls, *et al.*, Control of Bacillus subtilis Replication Initiation during Physiological Transitions and Perturbations. *MBio* **10** (2019).

28. Supplementary Information: numerical methods. https://github.com/junlabucsd/DoubleAdderArticle/tree/frontiers (commit 9e3899ce55f0e6c60a38b37f420796cbbf0cf893).

29. A. Adiciptaningrum, M. Osella, M. C. Moolman, M. Cosentino Lagomarsino, S. J. Tans, Stochasticity and homeostasis in the E. coli replication and division cycle. *Sci. Rep.* **5**, 18261 (2015).

30. G. Witz, T. Julou, E. van Nimwegen, Response to comment on “Initiation of chromosome replication controls both division and replication cycles in E. coli through a double-adder mechanism.” *bioRxiv* (2020) https://doi.org/10.1101/2020.08.04.227694.

31. G. Le Treut, F. Si, D. Li, S. Jun, Comment on “Initiation of chromosome replication controls both division and replication cycles in E. coli through a double-adder mechanism.” *bioRxiv* (2020) https://doi.org/10.1101/2020.05.08.084376.

32. G. L. Treut, G. Le Treut, F. Si, D. Li, S. Jun, Single-cell data and correlation analysis support the independent double adder model in both Escherichia coli and Bacillus subtilis. *bioRxiv* (2020) https://doi.org/10.1101/2020.10.06.315820.

33. G. Facchetti, B. Knapp, F. Chang, M. Howard, Reassessment of the Basis of Cell Size Control Based on Analysis of Cell-to-Cell Variability. *Biophys. J.* **117**, 1728–1738 (2019).

34. M. Schaechter, J. P. Williamson, J. R. Hood Jr, A. L. Koch, Growth, cell and nuclear divisions in some bacteria. *J. Gen. Microbiol.* **29**, 421–434 (1962).
Materials and Methods

Strains and growth conditions

The following strains were used in this study.

- **BW25113**: F- DE(araD-araB)567 lacZ4787(∆rrnB-3) LAM- rph-1 DE(rhaD-rhaB)568 hsdR514
- **SJ_FS130**: To construct this strain, we introduced ΔdnaN::[dnaN-ypet] into BW25113 using P1 transduction.
- **SJ_DL188**: MG1655 F- λ- rph-1 dnaA msfGFP kan mCherry-dnaN
- **SJ_FS130**: MG1655 F- DE(araD-araB)567 lacZ4787(del)::rrnB-3 LAM- rph-1 DE(rhaD-rhaB)568 hsdR514

(Note that the genotype of BW27378 is F- DE(araD-araB)567 lacZ4787(del)::rrnB-3 LAM- rph-1 DE(rhaD-rhaB)568 hsdR514 (::FRT).)

We used minimal MOPS media for the MG1655 experiments and minimal M9 glucose media with and without uracil for the SJ_FS130 (BW25113 based) experiments.

Microfluidics, Microscopy and Image processing

We used the same method as described in (11).

Appendix A: Cell size homeostasis in the RDA model

In the model proposed by Witz and colleagues, the cell size per origin at division is determined by the cell size at initiation per origin \(s_i\), the added size per origin between consecutive replication initiation events \(\delta_{id}\), and the added size per origin from replication initiation to cell division \(\delta_d\). The following relation holds:

\[
S_d^{(n)} = s_i^{(n)} + 2 \delta_d^{(n)} + \delta_{id}^{(n)}
\]

where the index as \(n\) denotes the generation (or division cycle). Under the assumption that \(\delta_{id}^{(n)}\) (resp. \(\delta_d^{(n)}\)) are independently and identically distributed Gaussian stochastic variables with mean \(\mu_i\) (resp. \(\mu_d\)) and standard deviation \(\sigma_i\) (resp. \(\sigma_d\)), it follows that \(s_i^{(n)}\) and \(S_d^{(n)}\) are also Gaussian stochastic variables. At large \(n\), they converge to the limiting distributions \(S_i^* = N(\mu_i, \sigma_i^2)\) and \(S_d^* = N(\mu_d, \sigma_d^2)\), where:

\[
\mu_i = 2 \mu_i, \quad \sigma_i^2 = 4 \sigma_i^2
\]

\[
\mu_d = 2 (\mu_i + \mu_d), \quad \sigma_d^2 = 4 (4 \sigma_i^2 + \sigma_d^2)
\]

The mother-daughter correlation for division size is a central quantity in cell size homeostasis, which can be derived in this model. As a first step, let us define the centered variables: \(dS_i^{(n)} = s_i^{(n)} - \mu_i\) and \(dS_d^{(n)} = S_d^{(n)} - \mu_d\). We then obtain the relations:

\[
< dS_i^{(n+1)} \cdot dS_i^{(n)} > = \frac{1}{2} \sigma_i^2,
\]

\[
< dS_d^{(n+1)} \cdot dS_d^{(n)} > = \frac{1}{2} < ds_i^{(n+1)} \cdot ds_i^{(n)} >,
\]

where the brackets denote averages. We therefore obtain the mother-daughter Pearson correlation coefficients (in the large \(n\) limit):

\[
Q_i = \frac{<ds_i^{(n+1)}ds_i^{(n)}>}{\sigma_i^2} = \frac{1}{2},
\]

\[
Q_d = \frac{<dS_d^{(n+1)}dS_d^{(n)}>}{\sigma_d^2} = \frac{1}{2} \left(1 + 3 \frac{\sigma_d^2}{\sigma_i^2}\right)^{-1}.
\]

It also follows that the Pearson correlation coefficient between initiation size per origin and division size is:

\[
Q_{id} = \frac{<ds_i^{(n)}dS_d^{(n)}>}{\sigma_d \sigma_i} = \left(1 + 3 \frac{\sigma_d^2}{\sigma_i^2}\right)^{-1/2}.
\]
In this model the joint distribution \((S_d^{(n)}, S_d^{(n+1)})\) is a bivariate Gaussian, therefore we can write the conditional expectation of \(S_d^{(n)}\) as:

\[
<S_d^{(n)} | S_d^{(n-1)} > = \varphi_d S_d^{(n-1)} + (1 - \varphi_d) \mu_d.
\]

With the hypothesis of symmetrical division, namely \(S_b^{(n)} = 2 S_d^{(n+1)}\), we obtain for the conditional expectation of the added size from birth to division:

\[
<S_d - S_b | S_b > = (2 \varphi_d - 1) S_b + (1 - \varphi_d) \mu_d.
\]

Therefore, the “adder” principle is equivalent to having \(\rho_d = 1/2\). Therefore, the RDA model always results in \(2 \rho_d - 1 < 0\). In fact, it only reproduces the “adder” principle in the deterministic limit \(\sigma_{id} \to 0\).

**Appendix B: Calculation of \(\rho(s, \delta_d)\) for the IDA model**

In the IDA model, the cell size at division is determined by the cell size at birth \(S_b\) and the added size from birth to division \(\Delta_d\); and the cell size per origin at initiation in the next cell cycle \(s^{(n+1)}\) is determined by the added size per origin between consecutive replication initiation events \(\delta_{id}\) and the cell size per origin at initiation \(s_i\). The following relations hold:

\[
\begin{align*}
S_d^{(n)} &= \frac{1}{2} S_d^{(n-1)} + \Delta_d^{(n)}, \\
S_i^{(n+1)} &= \frac{1}{2} s_i^{(n)} + \delta_{i}^{(n)},
\end{align*}
\]

where the index \(n\) denotes the generation (or division cycle), and \(\{\Delta_d^{(n)}\}\) and \(\{\delta_{i}^{(n)}\}\) are independently distributed random variables. Denoting \(\mu_d = < \Delta_d >\) and \(\mu_i = < \delta_{i} >\), we have \(\mu_d = < S_d > = 2 \mu_{dd}\) and \(\mu_i = < s_i > = 2 \mu_{ii}\). We also define the centered variables: \(d\Delta_d^{(n)} = \Delta_d^{(n)} - \mu_{dd}\), \(dS_d^{(n)} = S_d^{(n)} - \mu_d\), \(d\delta_{i}^{(n)} = \delta_{i}^{(n)} - \mu_i\) and \(ds_i^{(n)} = s_i^{(n)} - \mu_i\). Denoting \(\sigma_{dd}^2 = < d\Delta_d^2 >\) and \(\sigma_{ii}^2 = < d\delta_{i}^2 >\), we have \(\sigma_d^2 = < dS_d^2 > = \frac{1}{2} \sigma_{dd}^2\) and \(\sigma_i^2 = < ds_i^2 > = \frac{1}{2} \sigma_{ii}^2\).

The added size between initiation and division reads \(\delta_{id}^{(n)} = (S_d^{(n)} - s_i^{(n)})/2\), where the factor of 2 reflects the fact that origins of replication double at initiation. Therefore:

\[
< ds_i^{(n)} \cdot d\delta_{id}^{(n)} > = - \frac{1}{2} < (ds_i^{(n)})^2 > ,
\]

by independence of \(S_d^{(n)}\) and \(s_i^{(n)}\), and we obtain the Pearson correlation coefficient:

\[
\frac{< ds_i^{(n)} \cdot d\delta_{id}^{(n)} >}{\sigma_d \sigma_i} = - \left( 1 + \frac{\sigma_{d_{id}^2}}{\sigma_{d}^2} \right)^{-1/2} .
\]

**Appendix C: Calculation of \(\rho(s, S_d)\) and \(\rho(S_d^{(n)}, S_d^{(n+1)})\) for the sHC and IA models**

Let us rewrite Eq. 1 as:

\[
S_d^{(n)} = s_i^{(n)} \alpha^{(n)},
\]

where \(\alpha^{(n)} = \exp(\lambda^{(n)} \tau_{cyc}^{(n)}).\) To derive \(\rho(s, S_d)\) and \(\rho(S_d^{(n)}, S_d^{(n+1)})\), we will assume that \(\{s_i^{(n)}\}\) and \(\{\alpha^{(n)}\}\) are independent random vectors. This is true in the IA model (because \(\delta_{i}\) is independent of the other physiological variables), but it is an additional constraint for the sHC model. Using this assumption, we can derive the mean and variance of the cell size at division. The mean is given by:

\[
\mu_d = < S_d > = \mu_i \mu_{\alpha},
\]

where \(\mu_i = < s_i >\) and \(\mu_{\alpha} = < \alpha >\). Let us introduce the centered variables: \(ds_i^{(n)} = s_i^{(n)} - \mu_i\), \(d\alpha^{(n)} = \alpha^{(n)} - \mu_{\alpha}\) and \(dS_d^{(n)} = S_d^{(n)} - \mu_d\), and the variances \(\sigma_i^2 = < ds_i^2 >\) and \(\sigma_{\alpha}^2 = < d\alpha^2 >\). The variance of the cell size at division is given by:

\[
\sigma_d^2 = < dS_d^2 > = < s_i^2 > < \alpha^2 > + \mu_d^2 = \mu_d^2 \left( \left( \frac{\sigma_i}{\mu_d} \right)^2 + \left( \frac{\sigma_{\alpha}}{\mu_{\alpha}} \right)^2 + \left( \frac{\mu_i}{\mu_d} \right)^2 \left( \frac{\mu_{\alpha}}{\mu_d} \right)^2 \right).\]

We now compute \(\rho(s, S_d)\):
where we have introduced the coefficient of variations (CVs): \( \eta_i = \sigma_i / \mu_i \) and \( \eta_{d_i} = \sigma_{d_i} / \mu_{d_i} \). As can be seen, the value of \( \rho_{d_i} \) depends on the CVs of the initiation size and of \( \alpha = \exp(\lambda \tau_{\text{cyc}}) \). When \( \alpha \) is deterministic (i.e. \( \eta_i = 0 \)), division is “slaved” to replication initiation and \( \rho_{d_i} = 1 \).

To compute \( \rho(S_d^{(0)}), S_d^{(n+1)} \), we introduce the mother/daughter correlations:

\[
\rho_q = \frac{<dS_d^{(0)}dS_d^{(n+1)}>}{\alpha_q^2} \quad \text{and} \quad \rho_\alpha = \frac{<d\alpha^{(0)}d\alpha^{(n+1)}>}{\alpha_\alpha^2}.
\]

In particular, we have \( <S_i^{(n)}S_i^{(n+1)}> = \mu_i^2 + \sigma_i \sigma_i^2 \) and \( <\alpha^{(n)}\alpha^{(n+1)}> = \mu_\alpha^2 + \sigma_\alpha \sigma_\alpha^2 \). We thus have:

\[
\rho_q = \frac{<dS_d^{(n)}dS_d^{(n+1)}>}{\alpha_q^2} = \frac{<q_{d_i}^{(n)}q_{d_i}^{(n+1)}>}{\sigma_{d_i}^2} = \frac{\mu_i^2 \eta_i^2 + \sigma_i \sigma_i \eta_i^2 + \sigma_i \sigma_i \eta_i^2 \eta_i^2}{\eta_i^2 + \sigma_i \sigma_i \eta_i^2 + \sigma_i \sigma_i \eta_i^2}.
\]

When \( \alpha \) is deterministic (i.e. \( \eta_i = 0 \)), we have \( \rho_\alpha = \rho_d = \rho_c \); cell size homeostasis is determined by the initiation size homeostasis. On the other hand, when \( q \) is deterministic (i.e. \( \eta_q = 0 \)), we have \( \rho_c = \rho_{d_i} \). In general, the level sets of \( \rho_d \) are determined by a quadratic equation of the variables \( \rho_i \) and \( \rho_{d_i} \). Therefore each level set is a conic section:

In the absence of mother/daughter correlations, namely \( \rho_i = \rho_{d_i} = 0 \), then \( \rho_d = 0 \), which is a sizer regime. If we only have \( \rho_i = 0 \), then:

\[
\rho_d = \frac{\sigma_{d_i} \eta_i^2}{\eta_i^2 + \sigma_i \sigma_i \eta_i^2}.
\]

For the IA model, we have \( \rho_i = 1/2 \). Furthermore, there are no mother/daughter correlations among the other physiological variables, therefore \( \rho_{d_i} = 0 \). We thus obtain:

\[
\rho_d^{IA} = \frac{\sigma_{d_i} \eta_i^2}{2(\eta_i^2 + \sigma_i \sigma_i \eta_i^2)} < \frac{1}{2}.
\]

Therefore, the IA model can only reproduce the adder correlation in the deterministic limit where \( \alpha = \exp(\lambda \tau_{\text{cyc}}) \) is a deterministic variable, namely \( \eta_q = 0 \).

**Appendix D: Calculation of \( \rho(S_n, S_d) \) and \( \rho(S_d^{(n)}, S_d^{(n+1)}) \) for the CCCP model**

For simplicity, and following the conventions used by the authors of this model (14), we rewrite Eq. (8-11) as:

\[
\ln(s_i^{(n+1)}) = q_i^{(n+1)} = q_i q_i^{(n)} + A^{(n)} ,
\]

\[
\ln(S_R^{(n)}) = q_R^{(n)} = q_i^{(n)} + \lambda C^{(n)} ,
\]

\[
\ln(S_H^{(n)}) = q_H^{(n)} = q_H q_H^{(n-1)} + B^{(n)} ,
\]

\[
\ln(S_d^{(n)}) = q_d^{(n)} = \max(q_R^{(n)}, q_H^{(n)}) ,
\]

where we have introduced \( \rho_i \) and \( \rho_{d_i} \) to be more general. Assuming that \( A^{(n)}, B^{(n)} \) and \( C^{(n)} \) are Gaussian variables with means \( \mu_A, \mu_B \) and \( \mu_C \), and variances \( \sigma_A^2, \sigma_B^2 \) and \( \sigma_C^2 \), then \( q_i^{(n)}, q_R^{(n)} \) and \( q_H^{(n)} \) are also Gaussian variables. We first compute their means and variances. We have:

\[
\begin{align*}
q_i &= N(\mu_i, \sigma_i) \quad \text{with} \quad \mu_i = \frac{\mu_i}{1 - \sigma_i^2} \quad \text{and} \quad \sigma_i^2 = \frac{\sigma_i^2}{1 - \sigma_i^2} ; \\
q_R &= N(\mu_R, \sigma_R) \quad \text{with} \quad \mu_R = \frac{\mu_i + \lambda \mu_C}{\sigma_R^2} \quad \text{and} \quad \sigma_R^2 = \sigma_i^2 + \lambda^2 \sigma_C^2 ; \\
q_H &= N(\mu_H, \sigma_H) \quad \text{with} \quad \mu_H = \frac{\mu_H}{1 - \sigma_H^2} \quad \text{and} \quad \sigma_H^2 = \frac{\sigma_H^2}{1 - \sigma_H^2} .
\end{align*}
\]

Note that \( q_i \) is independent from \( q_R \) and \( q_H \). We now compute the autocovariances for these 3 variables. We obtain:

\[
<q_i^{(n+1)}q_i^{(n)}> = q_i \sigma_i^2 ;
\]
\begin{itemize}
\item \( dq_R^{(n+1)} dq_R^{(n)} = q_i \sigma_i^2 \) since \( C^{(n)} \) is an independently distributed random variable;
\item \( dq_H^{(n+1)} dq_H^{(n)} = \sigma_H \sigma_H^2 \),
\end{itemize}

where as before \( dX^{(n)} = \chi^{(n)} - \langle X \rangle \). We furthermore compute for subsequent use the following cross-correlation:

\[ < dq_R^{(n)} dq_i^{(n)} > = \sigma_i^2 \] since \( C^{(n)} \) is an independently distributed random variable. These preliminary results will be useful to compute \( \rho_d, \rho \) and \( \rho_{id} \).

We now follow the approach of the authors, and replace the expression of \( q_d \) in Eq. (11) by the effective equation:

\[ q_d^{(n)} = q_H^{(n)} u^{(n)} + q_R^{(n)} (1 - u^{(n)}) , \]

where \( \{ \rho^{(k)} \} \) are independent Bernoulli variables such that: \( Pr(\rho^{(k)} = 1) = f \). We note that \( Pr(\rho^{(k)} = 0) = 1 - f \) and \( Pr((\rho^{(k)})^2 = 1) = f \). The previous equation is an effective approach in which \( f \) represents the fraction of the cases in which the division process is limiting, namely \( q_R^{(n)} < q_H^{(n)} \). With this effective expression, one can compute the mean \( \mu_d \) and the variance \( \sigma_d^2 \) of \( q_d \). We obtain:

\begin{itemize}
\item \( \mu_d = \mu_H f + \mu_R (1 - f) ; \)
\item \( \sigma_d^2 = (q_H^{(n)})^2 < (u^{(n)})^2 > + (q_R^{(n)})^2 < (1 - u^{(n)})^2 > + 2 < q_H^{(n)} > q_R^{(n)} < u^{(n)} (1 - u^{(n)}) > - \mu_d^2 , \)
\end{itemize}

We now turn our attention to the computation of \( \rho_{id} \). For that purpose, we compute the average:

\[ < q_i q_d >= < q_i q_H > < u > + < q_i q_R > < 1 - u > = \mu_i \mu_d + < dq_d dq_R > + \mu_i \mu_R (1 - f) = \mu_i \mu_d + \sigma_i^2 (1 - f) . \]

We thus obtain:

\[ < dq_d dq_d >= q_i q_d - \mu_i \mu_d = (1 - f) \sigma_i^2 \]

and therefore:

\[ Q_{id} = \frac{< dq_d dq_d >}{\sigma_{id}} = (1 - f) \frac{\sigma_i}{\sqrt{\sigma_H^2 f^2 + \sigma_R^2 (1 - f)^2 + (1 - f) (\mu_d - \mu_H)^2}} . \]

We now turn our attention to the computation of \( \rho_{id} \). We start by computing the average:

\[ < q_d^{(n)} q_d^{(n+1)} >= q_H^{(n)} q_H^{(n+1)} < u^{(n)} > < u^{(n+1)} > + q_R^{(n)} q_R^{(n+1)} < 1 - u^{(n)} > < 1 - u^{(n+1)} > + q_H^{(n)} q_H^{(n+1)} < u^{(n)} > < 1 - u^{(n+1)} > + q_R^{(n)} q_R^{(n+1)} < u^{(n)} > < u^{(n+1)} > < 1 - u^{(n+1)} > , \]

\[ = \mu_H^2 f^2 + \mu_R^2 (1 - f)^2 + 2 \mu_H \mu_R (1 - f) + Q_H \sigma_H^2 f^2 + Q_i \sigma_i^2 (1 - f)^2 . \]

We therefore obtain:

\[ < dq_d^{(n)} dq_d^{(n+1)} >= q_i q_d - \mu_d^2 = Q_H \sigma_H^2 f^2 + Q_i \sigma_i^2 (1 - f)^2 . \]

Finally, we obtain:

\[ Q_d = \frac{< dq_d^{(n)} dq_d^{(n+1)} >}{\sigma_d^2} = \frac{Q_H \sigma_H^2 f^2 + Q_i \sigma_i^2 (1 - f)^2}{\sigma_H^2 f^2 + \sigma_R^2 (1 - f)^2 + (1 - f) (\mu_d - \mu_H)^2} . \]

When the division process is limiting, namely \( f = 1 \), then we have \( \rho_{d} = \rho_B \). Conversely when the initiation process is limiting, we have \( \rho_{d} = \rho \). More generally, the level set curves for \( \rho_d \) are lines in the \((\rho_{id}, \rho)\) plane.

**Appendix E: I-value analysis**

Following the methodology proposed by Witz et al. (17), we first defined the covariance matrix (Eq. 13) of the analyzed variables. Then the \( I \)-value was computed as:
\[ I = \frac{\det(K)}{\prod h_i}, \]

where the \( k_{ii} \) are the diagonal elements of the covariance matrix, hence the variances. To compute the \( I \)-values of the four models listed in Figure 3 (top), we used the following combinations of physiological variables:

| Model | Variables |
|-------|-----------|
| RDA   | \( \lambda, \delta_{id}, \delta_{is}, s_i \) |
| IDA   | \( \lambda, \Delta_d, \delta_{is}, S_b \) |
| IA    | \( \lambda, \delta_{is}, \tau_{cyc}, s_i \) |
| sHC   | \( \lambda, \tau_{cyc}, S_s, S_i \) |

In Figure 3 (bottom), we defined 18 physiological variables in line with the definitions given by (17) (see Figure S1): \( \lambda, S_s, S_{d}, S_b, S_i, \Delta_d, \Delta_{is}, S_{b}^{(n-1)}, S_{b}^{(n+1)}, \delta_{is}, \delta_{id}, R_{is}, R_{id}, R_{is}, \tau, \tau_{cyc}, \) and \( \delta_{ii} \). We generated all possible combinations of 4 variables, that is \( 3060 \). We emphasize the difference between \( s_i \) and \( S_s \). The former is the cell size per origin of replication at the initiation event associated with the division in the current generation; therefore it could be in a previous generation (e.g. mother, grand-mother cell). The latter is the cell size when replication initiation occurs in the current generation. These 2 variables are only the same in the case of non-overlapping cell cycles (Figure 1). We also found that the results of this analysis were sensitive to processing of the experimental data. In their work, Witz and colleagues, instead of using the measured values for \( S_s, S_b, S_i, \) and \( s_i \), first fitted traces of cell sizes to an exponential function and took values interpolated by this fit. Yet it did not affect the relative ranking of the division-centric model with respect to the initiation-centric model.

**Figure S1:** Definition of the physiological variables in a scenario with 2 overlapping cell cycles. All variables with blue font are associated with the current generation \((n)\). We added a superscript when using physiological variables associated with another generation. We also defined variables derived from these quantities. Replication initiations are indicated with stars, and the red star is the initiation determining the division for the current generation. \( N_{ori} \) is the number of origins of replication just before initiation happens.
Appendix F: Simulations

In this study we have performed simulations of both initiation-centric and division-centric models. For this, we have re-used the code provided in (17). Few and minor modifications have been made, but these modifications did not affect the outcome or the essence of the original simulations. Simulations performed consist of:

- Repeats of the simulations performed by Witz and colleagues in their original study.
- Simulations using Witz and colleagues' original parameters, but with perfectly symmetrical division.
- Simulations of Witz et al. model, and our model, using experimental parameters taken from experimental datasets from Witz and et al. and datasets published in (11).

Witz et al.'s simulations of the RDA reproduced the adder behavior observed in their data, in apparent contradiction with our prediction in Eq. 5 that the RDA model is inconsistent with size homeostasis by the adder. We analyzed their simulations and found that they produced the adder-like behavior because of the additional fluctuations in the septum position (Figure S2a). Experimentally, septum position represents the most precise control among all measured single-cell parameters with CV < 5% (8, 27). Indeed, removing fluctuations in the septum position alone made the simulation deviate from experiment in a quantitative manner consistent with Eq. 5. We also conducted a similar analysis using our experimental data (11), and reached the same conclusion (Figure S2b). Based on this observation, we conclude that the RDA model does not self-consistently explain the adder phenotype.

![Figure S2](image_url)

**Figure S2:** Agreement of the RDA model with experimental data from a. (17), M9 + glucose condition and b. (11), MOPS + glucose condition (MG1655 *E. coli*). After removing fluctuations in the division ratio, simulations don’t agree with experimental data.
Supplementary Figures

**Figure S3** - **a.** Witz et al.'s validation of the RDA model is exclusively based on this data set. **b.** 3 out of the 4 datasets of Witz et al. present a slight negative correlation. **c-d.** Most of the experimental data we obtained with *E. coli* MG1655, NCM3722, BW25113 and *B. subtilis* NCIB 3610 strain also show varying degrees of negative correlation, consistent with the IDA model.

**Figure S4** - Varying degrees of positive correlation between the birth size $S_b$ and the initiation size per origin $s_i$. The experimental data from (17) shows the most positive correlation whereas experimental data acquired in our lab with *E. coli* MG1655 and with *B. subtilis* show moderate to zero correlation.