



UPPSALA  
UNIVERSITET

*Digital Comprehensive Summaries of Uppsala Dissertations  
from the Faculty of Science and Technology 1204*

## How precise is cyclic life?

*Insights during a single molecule revolution of the  
bacterial cell cycle.*

MATS WALLDÉN



ACTA  
UNIVERSITATIS  
UPSALIENSIS  
UPPSALA  
2014

ISSN 1651-6214  
ISBN 978-91-554-9104-8  
urn:nbn:se:uu:diva-235546

Dissertation presented at Uppsala University to be publicly examined in BMC A1:111a, Husarg. 3, Uppsala, Friday, 5 December 2014 at 09:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Peter Graumann (Philipps Universität Marburg).

#### **Abstract**

Walldén, M. 2014. How precise is cyclic life? Insights during a single molecule revolution of the bacterial cell cycle.. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 1204. 77 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-554-9104-8.

Bacterial cells reproduce by doubling in size and dividing. The molecular control systems which regulate the cell cycle must do so in a manner which maintains a similar cell size over many generations. A cell can under conditions of fast growth conclude cell cycles in shorter time than the time required to replicate its chromosome. Under such conditions several rounds of replication are maintained in parallel and a cell will inherit replication processes which were initiated by an ancestor. To accomplish this the cell has to initiate and terminate one round of replication during each cell cycle.

To investigate the effects of the cell cycle on gene-regulation in the gut bacterium *Escherichia coli*, an experimental method combining microfluidics, single molecule fluorescence microscopy and automated analysis capable of acquiring an arbitrary number of complete cell cycles per experiment was developed. The method allowed for the rapid exchange of the chemical environment surrounding the cells. Using this method it was possible to measure the dissociation time of the transcription factor molecule, LacI-Venus, from the native lactose operator sequence, *lacO<sub>I</sub>*, and an artificially strong operator, *lacO<sub>syn</sub>*, *in vivo*. The results indicated that regulation of gene-expression from the lactose operon does not occur at equilibrium in living cells. Furthermore, by studying the intracellular location of non-specifically interacting transcription factor molecules it was possible to determine that these do not form long-lived gradients inside the cell as was previously proposed.

By studying the replication machinery and the origin of replication it was found that replication is initiated according to a cell volume per origin which did not vary over different growth conditions. Further, division timing was found to be determined by the initiation event to occur after a fixed time-delay. A consequence of this mode of regulation is an uncertainty relation between the size at birth and the cell cycle time, in which cells will vary more in the cycle time during conditions of slow growth as compared to fast growth and vary more in birth length during conditions of fast growth as compared to slow growth.

**Keywords:** E.coli, cell cycle, replication initiation, transcription factor, gene-regulation, single molecule microscopy, microfluidics

*Mats Walldén, Department of Cell and Molecular Biology, Box 596, Uppsala University, SE-75124 Uppsala, Sweden.*

© Mats Walldén 2014

ISSN 1651-6214

ISBN 978-91-554-9104-8

urn:nbn:se:uu:diva-235546 (<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-235546>)

*“Il n'est pas certain que tout soit incertain.”*

Blaise Pascal



# List of Papers

This thesis is based on the following papers, which are referred to by their roman numerals.

- I Studying transcriptional interactions in single cells at sufficient resolution.  
Wallden M, Elf J.  
*Curr Opin Biotechnol.* 2011 Feb; 22(1):81-6.
- II High-throughput gene expression analysis at the level of single proteins using a microfluidic turbidostat and automated cell tracking.  
Ullman G\*, Wallden M\*, Marklund EG, Mahmutovic A, Razinkov I, Elf J.  
*Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 2012 Dec 24; 368(1611):20120025.
- III Direct measurement of transcription factor dissociation excludes a simple operator occupancy model for gene regulation.  
Hammar P\*, Wallden M\*, Fange D, Persson F, Baltekin O, Ullman G, Leroy P, Elf J.  
*Nat Genet.* 2014 Apr; 46(4):405-8.
- IV Fluctuations in replication initiation determine the generation time and size distribution in *E.coli* cells.  
Wallden M, Fange D, Ullman G, Marklund EG, Elf J.  
Submitted manuscript

Reprints were made with the permission from the respective publishers.

\* Authors contributed equally to this work.



# Contents

Introduction.....	11
Life .....	11
Outline.....	12
A bacterial individual .....	12
The central paradigm of molecular biology.....	12
Gene expression.....	13
Regulating gene expression .....	13
Molecular species and metabolism .....	13
The lactose operon, a bacterial gene.....	14
<i>Escherichia coli</i> , a model organism.....	16
The <i>E. coli</i> chromosome and DNA replication.....	17
Overlapping rounds of replication .....	18
Replication initiation .....	19
Requirements for concluding a cell cycle.....	19
Ensemble measurements .....	21
Fluorescence microscopy in molecular biology .....	21
Fluorescence .....	21
Fluorophores .....	22
Fluorescence in biology .....	22
Localization and the diffraction limit .....	22
Studying molecular interactions in living cells.....	23
Cell imaging .....	24
Present Work.....	26
Microscopy.....	26
Microfluidics and live cell imaging.....	26
Integrating methods .....	26
Micro-fabrication and chip assembly .....	27
Maintaining micro-flows .....	28
A trap motif .....	29
Immobilization of cells and growth conditions .....	29
Multiplexing and maximum throughput .....	30
Chip handling .....	30
Medium exchange.....	31
Data Analysis .....	32
Automation in data analysis.....	32
Data uniformity.....	33
Anti-drift acquisition .....	33
Active image registration.....	33
Trap detection.....	34
Image intensity normalization .....	34

MicrobeTracker .....	34
Cell mobility and tracking .....	35
Active pole tracking.....	36
Division classifier .....	36
Management of tracking errors.....	37
Active cell tracking.....	37
De-novo segmentation .....	37
An integrated method .....	38
Life cycles in a microfluidic chip.....	39
Cell size .....	39
Cycle time.....	40
Exponential growth.....	40
Variations over space and time .....	41
Interdependencies of cycle observables.....	41
Cycle time and growth rate.....	42
Inheritance of cycle properties.....	43
Gene expression and regulation over the cell cycle .....	44
Cell cycle dependent bursts of gene expression .....	44
Constructs for monitoring burst of expression .....	45
Counting fluorescent molecules as they are synthesized.....	45
Gene expression rates over the cell cycle .....	45
The dissociation time of a TF .....	46
An <i>in vivo</i> chase assay .....	47
Non-equilibrium models of regulation .....	48
Cell cycle dependent lapses in TF regulation.....	49
Investigating a post-replication refractory period for gene expression .....	50
Measuring the expression response to induction .....	51
The maturation time of fluorescent proteins.....	51
Regulation of the cell cycle.....	52
Chromosome management .....	52
Replication initiates according to size .....	53
The origin during the cell cycle.....	53
A time-delay in replication initiation.....	54
Predictions of the chromosome management model .....	55
Regulation of replication initiation.....	55
Regulation of division timing .....	56
Predictions under the cycle regulator model.....	57
Variance in cell cycle regulation .....	57
Conclusions and future outlooks.....	60
Technology development for the study of bacterial cells.....	60
Insights afforded by the method.....	60
Future developments .....	61

Uncertainty in the bacterial cell cycle .....	62
Appendix: Measurement and determination of the maturation time of Venus in live bacterial cells growing in a microfluidic chip .....	64
Introduction .....	64
Materials and methods .....	64
Results .....	65
Discussion .....	67
Swedish Summary .....	69
Acknowledgements .....	71
References .....	73
Papers I-IV .....	79

## Abbreviations

DNA	Deoxyribonucleic acid
DSCI	Double sided confidence interval
FOV	Field of view
FP	Fluorescent protein
IPTG	Isopropyl $\beta$ -D-1-thiogalactopyranoside
NA	Numerical Aperture
ONPG	Ortho-nitrophenyl- $\beta$ -galactoside
PALM	Photo-Activation Localization Microscopy
PDMS	Polydimethylsiloxane
PLL	Poly-L-Lysine
PSF	Point Spread Function
RNA	Ribonucleic Acid
SD	Standard Deviation
SEM	Standard Error of the Mean
STED	Stimulated Emission Depletion
STORM	Stochastic Optical Reconstruction Microscopy
TF	Transcription Factor

# Introduction

## Life

You and I are outnumbered. The unicellular beings of this earth are its majority. The familiar world conveyed to us by our senses and the multicellular beings we encounter there are a special rather than the general case of life. The earth, the oceans and even our own bodies are teeming with unicellulars. Consequently life is measured on their scale, the miniscule scale of one micro-meter, at which objects are well beyond our natural perception. This microscopic world is a place which macroscopic analogy and intuition often fail to describe. The dominant sensory faculties found here are more akin to smell and taste than sight and hearing. Locomotion is performed by swimming rather than walking. Should a macroscopic swimmer be miniaturized to this scale she would find herself completely immobilized but for a very peculiar range of swimming styles [1]. Gravity is all but suspended and instead of falling to the earth a small object is continually bombarded by its environment, forcing it to perform a random walk, to diffuse, if not attached or actively propelling itself [2]. Inertia is negligible, so that movement stops immediately when active propulsion is discontinued [2]. Ballistic weaponry is exchanged for a chemical arsenal as collisions carry no momentum. Sexual reproduction is rare and caring for progeny is rarer still. The genetic blueprint of each individual has a propensity to change, to mutate, and unicellular beings are less restricted in the ways they may mutate as little or no requirement for uniformity or concerted efforts between individuals is present. A multicellular being would have difficulty surviving and reproducing if each of its cells was a cancer. The evolutionary strategies of uni-cellulars are presumably adapted to this and are oriented towards fast reproduction with a high rate of mutability to generate a sufficient number of variants which can cope and thrive under the conditions they are subjected to. If there is a beating heart of unicellular life it is the cell cycle, during which an individual will duplicate its parts and give rise to a new individual. Each cycle holds the possibility for an individual blue-print to establish itself as the norm for future generations of its kind. Therefore uni-cellulars appear to have embraced the randomness which governs their existence, incorporating it into their life cycles. Their ability to regulate these cycles is unquestionable as nearly  $10^{30}$  cycles have concluded since you began reading this thesis. However, the mechanisms involved are largely unknown and consequently so are their

ability to cope with randomness. The precision in controlling the cell cycle will set the range of fates which a unicellular individual can expect to befall them. In order to understand what unicellular life is like, I ask and attempt to answer the following question in this thesis: How precise is cyclic life?

## Outline

This thesis is written as an account of the work done during my graduate studies, of the trials and errors, confusions and insights I and my co-workers have had. It largely follows the chronology of those events. Where it does not, anachronisms are introduced to illustrate specific points which seemed important to me. The narrative therefore does not specifically discuss the papers this thesis is based on separately beyond the timing of the details they contain. Measurement technology and biology are blended to a degree in the narrative. The earlier sections under *Present work* contain more detail of the technologies used, while later sections contain more biological findings. Paper I is a review of the field of single molecule fluorescence microscopy in live cells. It may serve as a compliment to the introduction. This thesis contains a self-contained manuscript, which is included as an appendix. It allows an interested reader access to details and results on which discussions in the main text are based.

## A bacterial individual

### The central paradigm of molecular biology

A bacterial individual can be regarded as being composed of a blue-print, an envelope and additional requisite parts to live and reproduce (see *Figure 1*). The blue-print contains the necessary information to construct and coordinate all constituents. This information is encoded in a molecular alphabet with four letters, commonly referred to as A, T, C and G. These are the bases of the DNA molecule and they will selectively form pairs so that A will pair with T and G will pair with C [3]. The bases, which in their free form are referred to as nucleotides, can be combined sequentially to form long polymers consisting of two strands which conform in space to a double-helix structure [3]. The sequence of each strand is complimentary to the other strand according to the aforementioned base pairing principles. The sequence carries a syntax which is common to all life, which allows the information to be translated through to up to two molecular languages before becoming realized as effector molecules. The additional languages are, as in the case of DNA, written as the sequential structure of two other polymer molecule types, RNA and proteins. The act of transmitting information from

DNA to RNA is called *transcription* and from RNA to proteins is called *translation*. The central paradigm of molecular biology stipulates that the flow of information is directional and traverses from DNA to RNA to protein [4].

## Gene expression

The blue-print of a bacterium is often a single circular DNA molecule, which is referred to as its chromosome. The chromosome is insulated from the environment inside the envelope along with most of the molecular effectors it encodes. This envelope is closed and therefore referred to as a cell. This spatial separation is the distinction of an individual's identity as the characteristics of each chromosome is represented by the contents of its cell. Therefore the term cell is commonly used as a synonym of the term individual. Transcription and translation are performed by two important molecular entities, RNA-polymerases and ribosomes. RNA-polymerases will synthesize a *messenger RNA*, mRNA, in a directed sequential manner using DNA as a template. The sequences of DNA which can be transcribed are genes and are bounded by sequence motifs which signify their start and end. The content of the mRNAs are in turn translated by ribosomes into protein molecules. This composite process is often referred to as gene expression (see *Figure 1* upper).

## Regulating gene expression

Gene expression is often regulated, presumably to match a cell's current requirements. This can be accomplished in several ways, at different stages of the expression process. The most common approach is to regulate the expression rate by either increasing or decreasing the rate of transcription. This is often accomplished by a designated class of proteins, transcription factors, TFs, which interact with a region close to the start of a gene, its promoter, to either facilitate or obstruct initiation of the transcription process. Accordingly, TFs can be classified by their mode of interaction as being either activators or repressors of transcription [5]. Selective targeting of a gene requires that the TF can search and recognize its genomic target. Therefore the specific target sequence, often referred to as an operator, and the gene encoding its designated transcription factor(s) have co-evolved to maintain this interaction [6].

## Molecular species and metabolism

Molecular effectors can be structural parts, enzymes which catalyze molecular reactions, regulators which modulate reaction rates or any combination of these categories. Bio-molecular reactions can to some extent be classified as

anabolic or catabolic; pertaining to if they integrate or disintegrate materials. An example of the latter is metabolism, where nutrients are disintegrated to release energy which can be used for work and refined materials which can be integrated into the cell. Photosynthesizing organisms, such as cyano bacteria, may use carbon dioxide from the atmosphere as a carbon source. Bacteria which cannot use this source must metabolize compounds which were ultimately synthesized by organisms which could. There is a multitude of different carbon sources. To utilize each source, specific designated molecular effectors which either catabolize or transform nutrients to a substrate which can be catabolized are required.

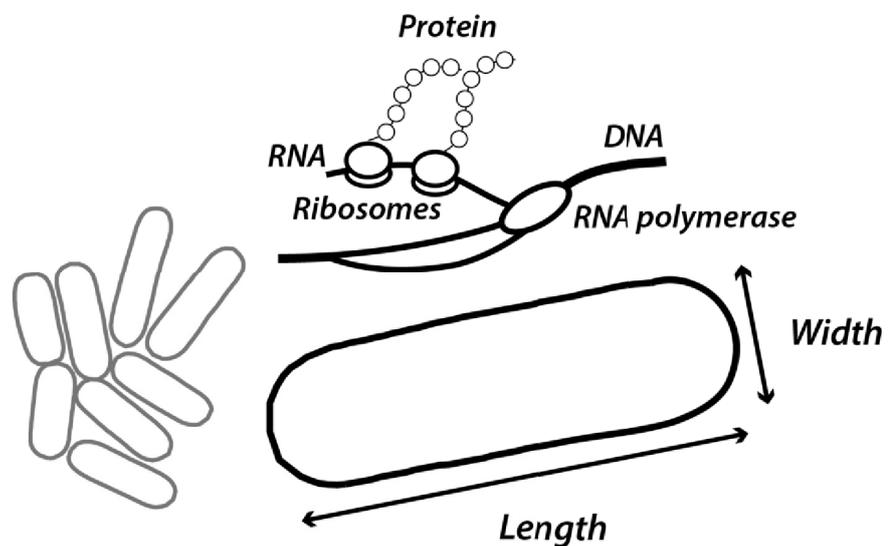


Figure 1. A bacterium. Left: outlines of a colony of cells. Right: Outline of one individual. Upper: intrinsic bio-molecular reactions.

### The lactose operon, a bacterial gene

Bacterial genes are sometimes grouped in succession on the chromosome according to the effects their products mediate, such that they can be co-expressed as a set (see Figure 2). These groups are called operons and their existence is presumed to be an evolutionary response to conditional needs for effectors which can act in concert. An example is the catabolic *lactose* operon of *Escherichia coli* which contains three genes, *lacZ*, *lacY* and *lacA*, which when expressed allow this bacterium to metabolize the sugar lactose. However, gene expression is costly [7] and it appears that natural selection has retained those bacteria which have the ability to selectively express genes only under conditions which necessitate the effectors they encode [5].

This is true for the lactose operon, or *lacZYA*, which is repressed in the absence of lactose and expressed in its presence [8]. This is accomplished by a conditional interaction between the repressing transcription factor, LacI, and the promoter of the gene, *p<sub>lac</sub>*. LacI has the ability to interact with DNA and with a molecule which is an intermediate in lactose catabolism, allolactose. When bound to allolactose, LacI will lose its selective affinity for its specific operator sequence leaving the promoter region unoccupied. LacI therefore promotes expression of the gene by means of its own absence in the presence of allolactose [9]. The native, or *wild type*, LacI molecule is a tetramer, *i.e.* four monomers of LacI form the molecule [9]. LacI has two DNA binding domains and can therefore bind to two operator sites simultaneously.

Three operator sites are found near the promoter, *lacO<sub>1</sub>*, *lacO<sub>2</sub>* and *lacO<sub>3</sub>*. The main operator site, *lacO<sub>1</sub>*, is located close to the lactose promoter, the auxiliary operators, *lacO<sub>2</sub>* and *lacO<sub>3</sub>* are located in the start of the *lacZ* gene and close to the end of the *lacI* gene. The region spanning all three operators is ~500 bp [9, 10]. The affinity of the interaction between LacI and the operator sites is determined by the DNA sequences of the operators [10]. When LacI is bound to the promoter, it is likely to be bound with one domain to *lacO<sub>1</sub>*. The interaction with *lacO<sub>3</sub>* will repress expression from the *lacI* gene and therefore this gene is auto-repressed. When bound to two operators simultaneously, LacI forms a loop which increases repression of the operon [9-11]. *E.coli* is reported to carry only a few copies of LacI per cell [12] and their search process for the operators combines free or three dimensional diffusion in the cytosol and one dimensional diffusion along DNA [13].

Several synthetic analogues to allolactose have been isolated. The most commonly used is *Isopropyl β-D-1-thiogalactopyranoside*, IPTG, a membrane permeable compound which induces the expression of the *lac* operon but cannot be metabolized itself [12].

The gene product of *lacZ* gene is an enzyme,  $\beta$ -galactosidase, which catalyzes the hydrolysis of lactose. A synthetic analog of its substrate has been isolated, *ortho-Nitrophenyl-β-galactoside*, ONPG, which will once hydrolyzed absorb light at a specific wavelength. Therefore the activity of the enzyme in a sample can be measured. The activity of the enzyme under repressed and active conditions, *i.e.* at low and high concentrations of IPTG respectively can be measured. The repression fold is the ratio of these values and for the native *lacZ* gene it has been established to be ~1000, *i.e.* the repressed cellular concentration of  $\beta$ -galactosidase is 1/1000 of the maximum activity [8]. In order to perform the assay, cells in the sample are lysed to release the  $\beta$ -galactosidase molecules.

The equilibrium occupancy model for gene regulation suggests that the average level of expression from a repressed gene is determined by the average fraction of the time that the promoter is unoccupied. One of its underlying

ing assumptions is that the rate of transcription initiation is constant from the time a TF dissociates from the promoter until a TF re-associates [14].

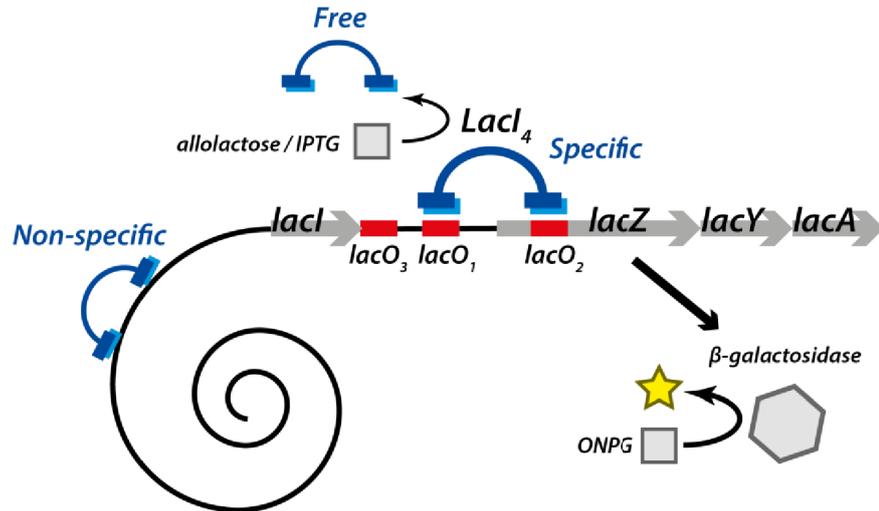


Figure 2. The lactose operon and regulation by LacI.

### *Escherichia coli*, a model organism

*E. coli* is regarded as a model organism in molecular biology. The bacterium is found in the lower intestine of warm blooded animals. In humans it constitutes ~0.1% of the gut flora [15]. The envelope of *E. coli* is rod-shaped and consists of a thin wall of peptidoglycan sandwiched between an inner and an outer phospholipid membrane layer. To reproduce, *E. coli* divides symmetrically by establishing and contracting a septum formed at the midsection. Constriction of the mother cell continues until the envelopes of the two nascent siblings are disconnected (see Figure 5). *E. coli* maintains its rod-shape during growth by expanding its length,  $L$  (see Figure 5). The volume of a cell,  $V$ , can be described as a cylinder with hemispherical end caps as

$$V(L) = \pi R^2 L - \frac{2\pi R^3}{3}$$

where,  $R$  is the cell radius. These endcaps are commonly referred to as the cell's poles and will constitute a constant sub volume of the cell described by a sphere with a radius  $R$ . A cell pole is established at the septum during division and as a result an *E. coli* cell will always have one pole which was made during the last division and one which was made any number of generations ago. The corresponding difference in the age of the poles of a cell is sometimes referred to as the age of the cell. Growth entails expanding the cell

volume and the *growth rate*,  $\mu$ , is the time derivative of the fractional volume expansion.

$$\mu = \frac{dV}{dt} \cdot \frac{1}{V}$$

The average growth rate of a population of *E. coli* cells is dependent on the living conditions. Factors which have proven to influence growth rate include the availability and quality of nutrients, the temperature and the absence of antibiotics. The cell cycle is bounded by two division events and its duration is referred to as the *cycle time*,  $\tau_c$ . The average of the cycle time is referred to as the *generation time*,  $\tau_g$ .

### The *E. coli* chromosome and DNA replication

In order to reproduce a bacterium has to copy its chromosome, which is accomplished by the process of DNA replication. The *E. coli* chromosome is circular and consists of  $4.6 \cdot 10^6$  bp [16]. With respect to replication the chromosome has an initiation and a termination site, which are referred to as the origin, *oriC*, and terminus, *ter*, of replication respectively. These are situated at opposite locations on the chromosome. The two distances between the origin and the terminus are therefore equal and are referred to as the arms or replichores of the chromosome [17]. During replication the strands of the DNA molecule are separated locally and each strand serves as a template for synthesizing a new complimentary strand (see *Figure 3*). This is achieved by integrating bases into the nascent strand successively after determining their complementarity to the corresponding template base. Each strand of the DNA molecule has a polarity, *i.e.* a given direction according to which replication will proceed. These are anti-oriented for two complimentary strands (see *Figure 3*). By opening or unwinding the strands of the chromosome locally at the origin, replication can proceed bi-directionally and continuously on the strands which polarity matches the direction of the corresponding replication process (see *Figure 3*). There is one such strand per chromosome arm, referred to as the *leading* strand. The remaining strand is referred to as the *lagging* strand.

The process of unwinding DNA is coupled to replication so that template material is made available as replication proceeds. Replication of the lagging strand is performed simultaneously as the leading strand, however here replication is discontinuous and piece-wise, resulting in a fragmented new strand. As sufficient stretches of the lagging strand are unwound, replication can proceed to the start of the last established fragment. These fragments, referred to as Okazaki fragments, are then joined to form a continuously connected strand [18]. At each site of replication two replicated DNA molecules will coalesce into one un-replicated molecule. This structure is often

referred to as a replication fork. The molecular effectors responsible for replication are associated to the replication fork, where they form a large complex. This complex is often referred to as the replisome [19]. A round of replication is completed at the terminus region and followed by segregation of the replicated chromosomes. The time required to complete a round of replication is referred to as the C-period, and is reported to be ~40 minutes in *E. coli* [20]. The time after terminating replication to the time of division is referred to as the D-period.

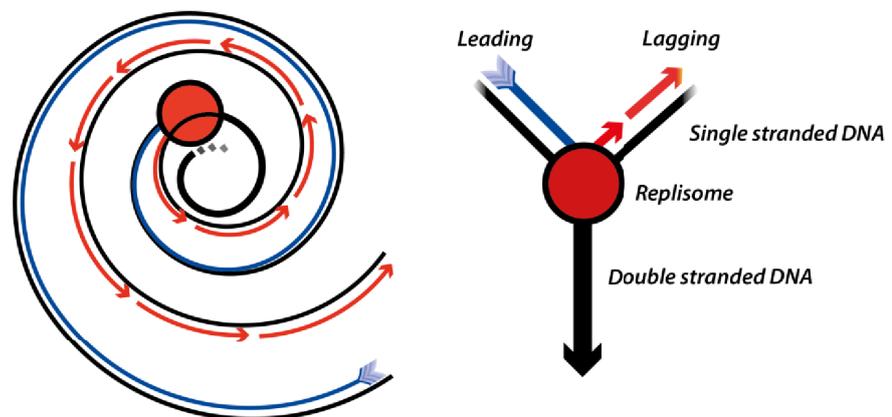


Figure 3. A replication fork. Black solid line: template DNA. Blue solid line: leading strand. Red solid line: Okazaki fragments on lagging strand.

### Overlapping rounds of replication

As *E. coli* has been reported to grow with generation times as brief as 18 minutes [21], the bacterium possesses the ability to complete its cell cycle in shorter time than required to complete one round of replication. This is accomplished by maintaining multiple rounds of replication initiated at different times. Therefore replisomes are separated from the terminus by different chromosomal distances (see Figure 4). In 1968 Cooper & Helmstetter proposed a model for predicting the average DNA content in continuously growing *E. coli* cells under different growth conditions [20]. This model predicts that overlapping rounds of replication occur if the generation time is less than the combined C + D period, *i.e.* when

$$\tau_g < C + D$$

Under such conditions, an individual inherits replication processes which were initiated prior to birth and its DNA is constantly undergoing replication. It also entails that cells have more than one origin [22].

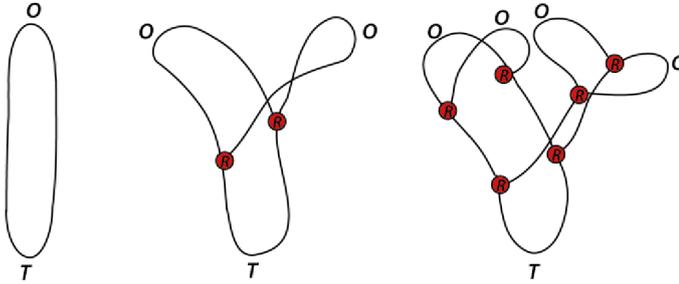


Figure 4. Chromosome replication. A circular chromosome with different number of replication rounds. Presented as origins (O), terminus (T) and replisomes (R).

### Replication initiation

Initiation of replication in *E. coli* is synchronous, meaning that all origins in a cell will initiate simultaneously [23] and the stability of an inheritance pattern of separated replisomes requires that initiation of replication occurs once and only once per cell cycle [24]. In 1968 Donachie proposed that *E. coli* will initiate replication at a stage of the cell cycle when it achieves a critical threshold of cell mass per number of origins and that this threshold is constant for different growth conditions [25]. In 2005 Bates & Kleckner reported that the time after division was more critical for the decision to initiate replication [26].

### Requirements for concluding a cell cycle

In order to complete a cell cycle under conditions of continuous growth, the average *E. coli* cell is required to

- Initiate one round of replication
- Terminate one round of replication
- Partition materials between the nascent daughters
- Establish a septum and divide

Further, the control systems which accomplish these tasks must do so in a manner which preserves cell size over many generations. The *relative expansion* of the cell,  $\Lambda$ , is the ratio between the current volume,  $V(\tau)$ , and the volume at birth,  $V_b$ , as

$$\Lambda(\tau) = \frac{V(\tau)}{V_b}$$

where  $\tau$  is the time after division. Stably maintaining similar sizes over many generations entails that the average cell divides after achieving twice the size it had at birth. *i.e.*

$$\langle \Lambda(\tau_c) \rangle_t = 2$$

If every segment of a cell will contribute equally to the total rate of growth, the cell will expand at a rate which is proportional to its size. This is referred to as exponential growth (see *Figure 10*) and under such conditions, the relative expansion is described by

$$\Lambda(\tau) = \exp(\mu \cdot \tau)$$

where  $\mu$  is constant with respect to  $\tau$ . The volume can then be described as

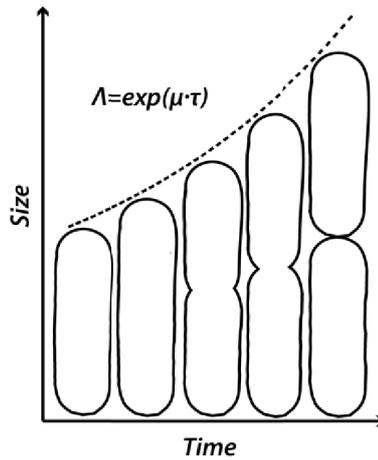
$$V(\tau) = V_b \Lambda(\tau) = V_b \exp(\mu \tau)$$

During exponential growth the requirement for maintaining a stable size is then

$$\langle \Lambda(\tau_c) \rangle_t = \langle \exp(\mu \cdot \tau_c) \rangle_t = 2$$

In the ideal case, *i.e.* when a cell cycle concludes as a cell is exactly twice the size it was at birth, the growth rate and the cycle time are related as

$$\mu = \frac{\ln 2}{\tau_c}$$



*Figure 5.* Cell growth. A cell elongates exponentially and divides.

## Ensemble measurements

Insights to the inner workings of bacteria have often been derived in accordance to measurements which are averages of large ensembles of cells. A population of cells is often asynchronous, *i.e.* cells are in different phases of the cell cycle. Averaging over an ensemble of asynchronous cells may mask cell cycle dependent behavior. A common approach to measuring cell cycle dependent events is therefore to synchronize cells, usually according to the time after division. This can be achieved in several ways. Using the baby machine approach, cells are immobilized in a porous membrane by filtering a sample of cell culture [27]. By reversing the filter, then slowly passing medium through it in the reverse direction, newly divided cells can be eluted and collected. Using a similar approach, the baby column technique, immobilization is achieved by using cells which are genetically modified to conditionally express a factor which will allow them to attach to the matrix of a column [28]. The expression is induced during the early stages of the experiment until a sufficient amount of cells are immobilized. By passing medium through the column, cells which have divided will be released from the column. In both these approaches, fractions containing synchronized cells can be collected.

The range of ensemble measurement techniques which can afford insights on the expression of genes, synthesis of DNA and cell growth have developed dramatically over the last 50 years. Further, increased proficiency in sequencing and modifying genomes has expanded the range of possible bacterial variants to apply these on. However, each ensemble can contain subpopulations. If these are very different, the average may not represent any of the subpopulations accurately. Techniques which could allow researchers to study bacterial individuals separately could also potentially allow them to identify and characterize subpopulations.

## Fluorescence microscopy in molecular biology

### Fluorescence

Fluorescent molecules can absorb light at one wavelength and once in this excited state, emit light at another wavelength. The emission wavelength of a molecule is longer than that of the excitation light and is described by a probability density function referred to as an emission spectrum. By introducing a filter which selectively only allows light of the emitted wavelength to pass, the location of and the quantity of fluorescing molecules can be recorded as a response to directing excitation light to a sample. To further increase the sensitivity, the excitation light can be filtered to only contain wavelengths which maximize the probability of exciting the material. This

probability is described by the excitation spectrum. The excitation and emission spectra are characteristic for each species of fluorescent molecule and by utilizing their characteristics; several fluorescent compounds can be combined to allow selective detection of each from a mixed sample.

## Fluorophores

A wide range of fluorescent compounds have been isolated or synthesized and applied as reporters in various detection schemes. The quality of a fluorescent reporter molecule, or fluorophore, is described by several aspects of its performance. The brightness describes the number of photons emitted per unit time and is related to the strength of the signal which a reporter can generate. The Stokes shift is a measure of the separation between the excitation and emission spectra and is related to the feasibility of separating emitted light from interfering excitation light. Fluorophores bleach as they are exposed to excitation light and the relative ability to withstand exposure to light is referred to as their photo-stability. The true quality of a fluorescent reporter is however dependent on its application.

## Fluorescence in biology

The use of fluorescence in biology predates science and mankind as a wide variety of fluorescent organisms has been found in nature. The development of high-performance fluorophores has yielded increasingly better reporters. However, many of these are difficult to employ in the study of living cells. For this application, harnessing naturally occurring fluorophores has proven more successful. The green fluorescent protein, GFP, was isolated in the 1960s from the jelly fish *Aequorea victoria* [29]. With the development of protocols for genetic modification, the genes coding for fluorescent proteins could be identified and transferred to other organisms, evolved for better performance and alternative color and combined with native genes to produce fluorescent fusion proteins [30]. This allowed researchers to study the expression of genes in living organisms [31].

## Localization and the diffraction limit

A light microscope can be used to detect and localize fluorescent proteins within living cells. Due to the physical properties of light however, there is a limit to the spatial resolution which can be achieved. This limit is known as the diffraction limit and is the minimum distance at which two points in a micrograph image can be distinguished. The diffraction limit is roughly 250 nm [32] which constitutes 25% of the typical size scale of bacteria. Several

strategies have been reported for increasing the resolution beyond this limit, these are referred to as super-resolution techniques [33-35]. The most commonly used approach utilizes that light emitted from a stable point source will spread according to a point spread function, PSF. The PFS closely resembles a symmetric Gaussian function (see paper I, figure 1 d), the width of which corresponds to the diffraction limit. By comparing a diffraction-limited spot to this model, the accuracy in determining the center position converges as inverse of the square root of the number of recorded photons [36]. In the case of a few hundreds of emitted photons by an individual molecule, its position can be determined with an accuracy of 20-50 nm.

However, this can only be achieved if the target molecules are sparsely distributed within the cell [37]. One strategy for achieving this even for relatively abundant or *high-copy number* species is to genetically fuse these to a class of photo-activatable or photo-convertible FPs, which may alter their fluorescence properties in response to a pulse of light [38]. In this way a minute subpopulation of FPs can be activated, detected, localized and allowed to bleach before repeating the cycle. This technique has many names, two of the more commonly used are Photo-Activatable Localization Microscopy, PALM, and Stochastic Optical Reconstruction Microscopy, STORM [34, 35].

## Studying molecular interactions in living cells

Detection and localization of fluorescent particles is dependent on acquiring images on a time-scale during which the particle will appear to be immobilized. For example TFs bound to their genomic operators are immobilized for durations of more than 1 s [39], while freely diffusing [40] or non-specifically interacting TFs [39] are more mobile and may traverse the entire cell during the same time. Therefore, by acquiring images of fluorescently labeled TFs for long durations, the resulting fluorescence from the specifically bound fluorescent TFs are registered as localized spots. The fluorescence resulting from non-specifically bound or freely diffusing fluorescent TFs is distributed over all the locations at which these molecules resided during the acquisition [39]. In this way it is possible to selectively study molecular interactions rather than the total number of fluorescent fusion proteins. The mobility of molecular complexes or binding states varies from case to case. However, it is now possible to study freely diffusing individual FPs in live cells and consequently researchers are not limited by technology with respect to time resolution [40].

The time-scales of the binding states do however have to be sufficiently separated so that the resulting signal is generated by the molecules of the targeted state. Also as previously concluded, the interaction foci have to be sparsely distributed in the cell. Therefore the limits to which interactions can be studied in this way are set by biology itself. This is also true concerning

the most important aspect of this approach, namely that the modification of fusing a biomolecule to an FP a must not alter the native properties of that biomolecule. However, fluorescent molecular constructs can be engineered to meet these requirements. For example Yu *et al.* studied the expression from the *lacZ* gene of the lactose operon over the cell cycle by modifying the gene to encode for a fluorescent construct consisting of a fast maturing yellow fluorescent protein variant, Venus, fused to a membrane associating tag, Tsr. Tsr-Venus molecules were detected as diffraction limited dots as they were immobilized on the membrane for 100 ms [41].

## Cell imaging

Individual bacteria can be imaged using light microscopy. The light passing the sample is collected by the objective and the resulting image recorded by a camera. Bacterial cells are naturally transparent and in order to detect their outlines in micrograph data several contrast methods have been developed. The phase contrast method utilizes the shift in phase of light passing through a sample to create contrast, making structural information more accessible to the human eye. Using this method, bacterial cells can be detected as dark bodies on a lighter background.

In order to image live cells these have to be immobilized and kept under conditions which allow them to grow. Meeting the former condition entails the reduction or elimination of the inherent Brownian motion colloidal-particles such as bacteria are subject to. Further, cells have developed modes of active propulsion. Several techniques have been developed to immobilize cells in the focal plane. Tissue and cells can be fixed, *i.e.* chemically cross-linked for example using formaldehyde, allowing the organization of the cell to be maintained in stasis until the time of observation. Cells can also be chemically attached to a glass surface by coating it with poly-(L)-lysine, PLL.

Both these techniques fail to meet the criteria of maintaining cells in a state of growth. PLL is toxic to bacteria [42] and the attachment is temperature dependent and becomes unstable for temperatures above  $\sim 30$  °C. A common strategy to maintain cells is to sandwich them between a glass surface and a pad consisting of growth medium mixed with a gelatinous matrix, agarose. In this way cells are immobilized and have nutrients available for growth [43]. The number of successive generations which can be studied per experiment varies, but is limited by such things as availability of nutrients in the surroundings and crowding due to growth.

The field of microfluidics, which describes the behavior of fluids when confined in microscopic spaces, has found many recent applications in life sciences. In biochemical and biotechnological process the reagents are often expensive or time consuming to produce and sample size is often limited. By

fabricating miniaturized reaction vessels on a microfluidic chip, biochemical experiments can be performed using a fraction of the reagents. Further, these structural motifs can be multiplexed to allow many replicate reactions to occur simultaneously under near identical conditions.

The behavior of microscopically confined fluids also allows for added control and reliability as unpredictable aspects of macroscopic fluid mechanics are extinguished in miniaturized systems. The concept of a “lab on a chip” [44] entails that several chemical processes can be integrated on one miniaturized platform and realizing it has become increasingly more frequent.

Microfluidic designs have been used to study live bacterial cells. One example is the so-called mother machine approach [45]. Here, cells are immobilized by introducing them into a channel which is only slightly wider and deeper than the expected cell width. The channel is sealed at one end, open at the other and the length of the channel is considerably longer than the expected cell length. A progenitor or “mother” cell is captured at the start of the experiment and as this individual grows and divides the channel is filled with its progeny. The channel restricts the cells to grow in straight lines towards the open end from which cells are released and carried off by the flow through the main channel. The design removes or limits the freedom of cells to rotate or to move in the direction orthogonal to the direction of the channel.

## Present Work

### Microscopy

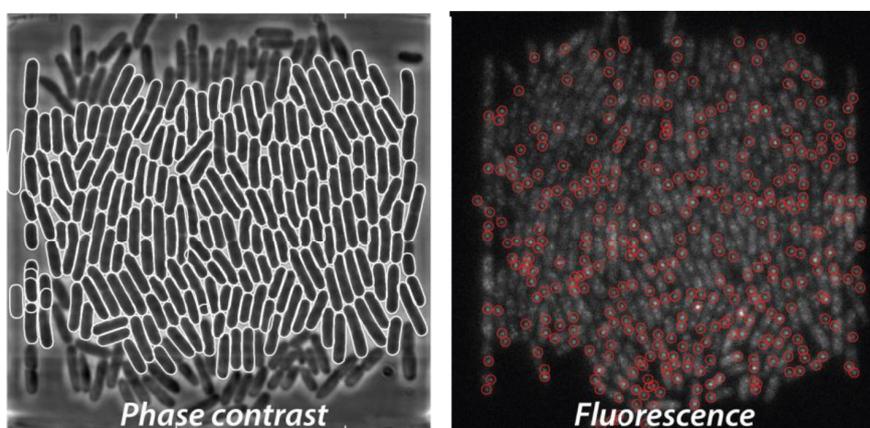
In the studies presented here, all data was collected using an inverted microscope (Nikon Ti Eclipse). To increase sensitivity, phase contrast was generated externally to the objective and as a consequence the phase contrast images and fluorescence images were recorded on separate cameras (see *Figure 6*). Phase contrast images of cells at 100 times magnification were recorded using a model CFW-1312M (Scion Corporation) camera at a contrast depth of 12 bit. The fluorophores in the sample were illuminated for excitation with collimated light, according to an epi-fluorescence configuration. The light source of excitation was provided by lasers (Ar-ion or solid state) emitting light at wavelengths appropriate for the excitation-emission characteristics of each fluorophore type. The resulting fluorescence was collected by the objective (APO TIRF 100x/N.A 1.49, Nikon), magnified by another lens (2X) and recorded by an EMCCD camera (Ixon EM plus (Andor Technologies)). The high numerical aperture, NA, of this objective increases the sensitivity by collecting a larger fraction of the light which was emitted by the sample. Acquisition of images was performed using  $\mu$ Manager [46] a free software package which allows for a high degree of flexibility in designing and customizing imaging protocols.

### Microfluidics and live cell imaging

#### Integrating methods

A method combining time-lapsed phase contrast and single molecule fluorescence microscopy could potentially provide insights to the relationship between the cell cycle and target molecular species. During experiments in which small colonies growing on an agarose pad were studied, several limitations to this approach were recognized: a targeted colony would often drift out of the field of view, the pad would dry out unexpectedly, cells would displace drastically, and as the colony grew cells would form local double layers.

In 2008 a study using a microfluidic device or chip to maintain *E. coli* cells which had been modified to express constituents of a synthetic genetic oscillator was published [47]. This or a similar system could potentially solve many of the problems we experienced using the agarose pad. As a result, we decided to design and develop a microfluidic sample management system tailored to our experimental needs. A common approach constructing microfluidic devices for microscopy is to bond a glass coverslip to a matrix containing microscopic structures so that the bonded interface, the glass and the matrix will serve as boundaries which contain fluids once introduced into the chip (see *Figure 7 c*). The matrix is often a cast of a mold containing a negative of the intended design of the microfluidic chip interior.

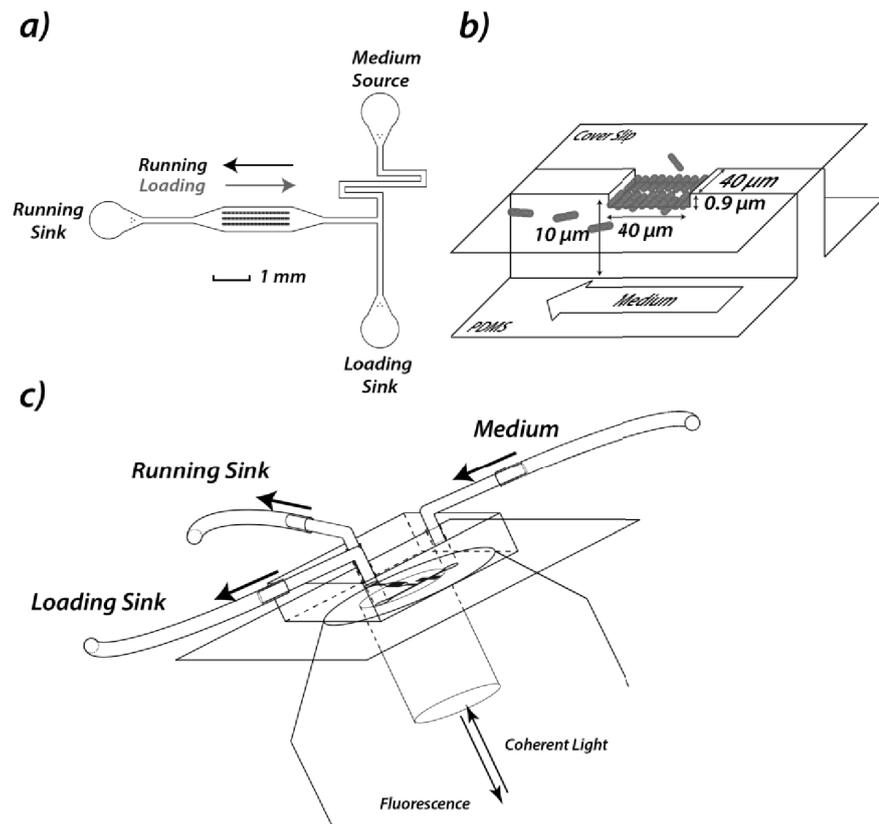


*Figure 6.* Cells imaged in phase contrast and fluorescence. White solid line: cell outline predictions. Red circles: detected fluorescent molecules

### Micro-fabrication and chip assembly

Micro-fabrication of structures can be achieved accurately down to sizes smaller than the average *E. coli* cell using a variety of different methods. To construct our molds we used UV-soft lithography to deposit microscopic structures on a polished surface. During this process a chemical which solidified when exposed to UV light, SU8, was spun to form a coat on a flat polished silicon substrate. The height of the coat could be controlled precisely by controlling the frequency and duration of the spin. The layer of SU8 was exposed to ultra-violet light through a binary mask. Once the microstructures were deposited, the substrates could be used as a mold for a high-fidelity casting agent, *polydimethylsiloxane*, PDMS. PDMS is once cured a hydrophobic, clear and gas permeable material. Therefore microfluidic chips constructed from PDMS would contain liquids, transmit light and allow for

the exchange of gases between the liquid and the surrounding air. For further detail on how micro-fabrication was performed, the reader is encouraged to consult the materials and methods section of paper II under *Design, fabrication and use of the microfluidic device*.



*Figure 7. Microfluidics. (a) Chip model A. The chamber holds 51 traps in three rows. (b) A trap with cells. (c) A microfluidic chip during microscopy. Medium is supplied and removed through tubing connected to the ports of the chip.*

### Maintaining micro-flows

To sustain growth we required a continuous flow of medium to supply the cells with nutrients. We utilized the principle of gravity flow, which is accomplished by connecting a chip to reservoirs filled with medium and suspending these at different heights to allow liquids to flow from *sources* to *sinks*. The relative heights of the reservoirs will determine the direction and the magnitude of the flow through the chip in a very predictable and reliable

manner. Once configured, the properties of the flow can be stably maintained for very long time periods. The reservoirs were connected to the port regions of the chip using tubing (see *Figure 7 a-c*). If air bubbles were retained in the tubing the flow could cease and end an experiment. To counter this risk, several measures were taken, including degassing the medium.

### A trap motif

Using the mother machine design a great number of cells can be monitored during one experiment [45]. If cells grow longer than the channel, the protruding end will be caught in the stream flowing through the main channel and the cell will be pulled out of the channel. If this happens to the mother cell, *e.g.* if it filaments, the channel will be rendered empty. Therefore, the number of populated channels in the mother machine decays over time.

An alternative approach was suggested by Mather *et al.* [48]. Here, cells grow in a square compartment in which the depth is the same as the expected cell width. As cells form colonies in the traps, they align to the direction of the walls and cells which reached the openings are released into the main chamber and removed by the flow of medium directed towards the sink. This trap motif relaxes the restrictions on cell shape, allowing filamenting cells to exit the trap without voiding it. However, the greater freedom to move adds complexity to the task of tracking cells over time. The maximum size of the colony is set by the dimensions of the trap and if growth conditions do not change the number of cells is constant over time.

We decided to use this design for our chips (see *Figure 7 b*). The traps were designed to maximize the number of cells which could be monitored simultaneously in one image acquired by our microscopy setup (see *Figure 6*). The upper limit is set by the projection of our samples on the chip of the fluorescence camera, which images regions of the sample which were  $41 \times 41 \mu\text{m}^2$ . Therefore we designed traps which were  $40 \times 40 \mu\text{m}^2$ .

### Immobilization of cells and growth conditions

In order to immobilize and maintain cells in a monolayer, the depth of the trap was fabricated to be the same as the expected cell width. As *E. coli* will alter its width in response to growth conditions, the molds used had to be specific for the intended growth condition of each experiment. For various growth conditions studied in our experiments, molds with trap depths of 1.1, 1.0, 0.9, 0.8, 0.7  $\mu\text{m}$  were fabricated. We defined two reference growth conditions which were intended to support fast and slow growth. We used a minimal medium, M9, as base for both conditions as it is chemically well-defined and not fluorescent. Glucose (0.4%) was selected as the carbon

source for fast growth, since it is present in the gut and *E. coli* cells will selectively metabolize it as long as it is present in their environment [49]. For slow growth succinate (0.4%) was used as carbon source. For both conditions we supplemented the medium with a blend of amino acids RPMI 1640 (Sigma-Aldrich). The temperature of the sample was controlled using a microscope incubator (OKO-lab). For conditions of fast growth we maintained the cells at 37°C as in the human gut. For slow growth we maintained cells at 30°C. Under the reference conditions we found that cells were sufficiently immobilized in the traps which were 0.8 and 0.9  $\mu\text{m}$  deep for slow and fast growth respectively.

### Multiplexing and maximum throughput

The trap motif was found to hold and maintain micro-colonies of  $\sim 350$  cells during slow growth and  $\sim 250$  cells during conditions of fast growth at any given time (see *Figure 6*). To further increase the number of cells which could be monitored during an experiment, we designed a chamber structure in which the trap motif was repeated 51 times distributed equidistantly over three rows (see *Figure 7 a*). The expected number of cells residing in the chamber was then in excess of  $10^4$ . The rate at which cell cycles are concluded in one chip was dependent on the generation time,  $\tau_g$ , as  $10^4/\tau_g \text{ min}^{-1}$ . As a reference for cells with a generation time of 30 min,  $10^6$  generations will have concluded in a chip after 5 hours.

### Chip handling

Two chip designs were used during the experiments, designated A and B (see *Figure 7 a* and *Figure 14 d*). Both chips were identical with respect to the design of the main chamber. Chip A was intended to allow continuous growth in a constant chemical environment. It had three media ports which were connected to the reservoirs by tubing and these were referred to as the *source*, *running sink* and the *loading sink* (see *Figure 7*). The role of the source was to provide a continuous supply of fresh medium to the cells. The role of the sinks was to receive medium which had passed through the chip from the source.

Operating a chip of design A or B has two phases, *loading* and *running*. During loading, cells are loaded through the running sink and allowed to flow to the loading sink (see *Figure 7 a* and *Figure 14 d*). The growth medium will flow continuously from the source port to the loading sink, thereby avoiding the risk of contaminating the medium source port. Cells flowing through the main chamber are captured by introducing pressure waves in the medium, by tapping or flicking the tubing until a sufficient number of cells have entered each trap. This concludes the loading phase and the running sink is re-adjusted to receive medium flowing from the source port. Chang-

ing the direction of flow is achieved by adjusting the relative levels of the reservoirs they are attached to.

Chip B was designed to allow the user to rapidly exchange the chemical environment of the cells while recording the effects of that exchange in real time (see *Figure 14 d*). Chip B has an additional source port as compared to chip A. The roles of the running and loading sinks are identical during the loading phase of operating chip A and chip B. During the running phase of chip B, the magnitude of flow from each source is balanced so that medium from both source ports reach the loading sink but only medium from one source reaches cells and ultimately the running sink.

### Medium exchange

Fluids behave very differently when confined in macroscopic as compared to microscopic systems. One characteristic trait of microscopic systems is the lack of turbulent mixing of fluids. As fluids meet they will form stable lamina at the interface. Here, mixing is slow as no turbulence is present. Therefore lamina containing different compounds can co-exist inside a microfluidic device with minimal exchange. In the manifold region of the chip (see *Figure 8* and *Figure 14 d*) the flows from the two sources meet and establish a stable laminar interface. Medium from one source, the active source, flows towards both the running and loading sink. As it flows towards the running sink, it passes the chamber and the cells. The medium flowing from the other source, the passive source, proceeds only towards the loading sink. Inert polystyrene particles are included in the medium in one of the source reservoirs as reporters of the bounds of the corresponding lamina inside the chip, allowing the user to configure the flows. During running of chip B, the active source reservoir will be elevated higher than the passive and to execute the medium exchange, their respective elevations are exchanged. This action can be performed by linear actuators and its execution program to synchronize with the acquisition program. By doing so the time-delay after executing the switch was determined to be 2 s (see *Figure 14 e* and supplemental under *IPTG induction*).

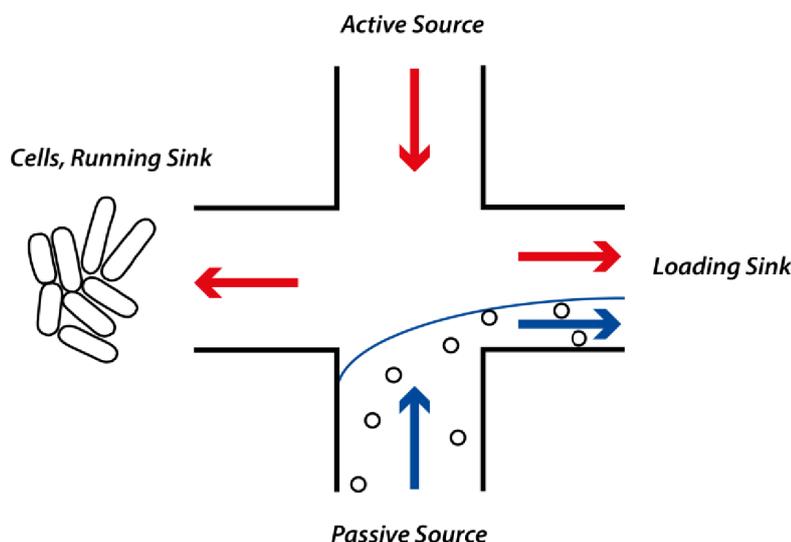


Figure 8. The manifold of chip B. Red and blue indicate different media. Blue solid: lamina interface. Black circles: polystyrene beads.

## Data Analysis

### Automation in data analysis

The data resulting from a time-lapsed microscopy experiment is gray-scale images (see Figure 6). Many studies based on microscopy data have relied on manual analysis. There are however potential problems to this approach; the classification criteria are inherently subjective and may vary between analysts or over the time an analyst works. Image analysis processes can be automated [50]. As an algorithm can work continuously, doing so can potentially increase the throughput in analysis. Further, automated analysis algorithms require objective and quantitative definitions for classification. Once defined, these enable the comparison between alternative definitions, allowing a progressive development towards greater accuracy, precision, speed etc.

We decided to develop automated analysis routines for our data. The objectives for these routines were to accurately and autonomously:

- detect and outline individual cells in phase contrast images
- detect and localize individual molecules in fluorescence images
- superimpose cell outlines and molecule positions
- track individuals between subsequent division events

## Data uniformity

Measurement noise can occur in microscopy data as variations in the focus, the background intensity, the contrast, the position of the sample relative the field of view, in the expected timing of acquisition etc. To reduce variations in our data, the microfluidic protocols were optimized (see paper II under *Materials and methods*). The microscope had an active focusing function (Perfect Focusing System, PFS, Nikon). However, this system was found to be sensitive to temperature variations and consequently minimizing fluctuations caused by the sample incubator improved focusing. The phase contrast camera would at times register images in which the total dynamic range was less than half of the average. This phenomenon was countered by introducing a feedback in the acquisition routines which would decide if a darker image was recorded in the phase contrast channel and if so re-acquire until an acceptable image was obtained or a maximum number of attempts had been made.

## Anti-drift acquisition

The sample would occasionally drift out of the field of view during the course of an experiment. We developed an active anti-drift algorithm which would correct for displacements between time points. This would require the user to record a reference image for each position at the start of the acquisition. The correction for each subsequent image was computed as the maximum cross-correlation between a test image and the reference image. Computing this in the frequency domain is fast and this allowed the correction to execute within 20 ms.

## Active image registration

Phase contrast images and fluorescence images were recorded using two separate cameras. Combining information derived from the different image types required a way to translate locations between images. This is often referred to as image registration. An active registration method was developed for our data. Since the fluorescence and phase contrast images hardly contain any common information, these two data types were deemed insufficient to base registration on. As an alternative source of registration information a bright-field image, in which the sample is exposed using a white light lamp, could be recorded using the fluorescence camera. The bright field image and the phase contrast image contained sufficient common information to construct an active super-resolved registration function. Consequently the acquisition of a phase contrast image and fluorescence image was accompanied by the acquisition of a bright field image.

## Trap detection

The trap motif could be used to span a two-dimensional coordinate system for each micro-colony. The edges of the trap walls could be detected in the phase-contrast image with super-resolution accuracy. However, the edges of open ends did not generate distinct signatures in the image. This was for two reasons: Firstly, there was no attachment between the PDMS and glass at these edges and the structures which were sought reside outside of the focal plane. Secondly, an interfering light phenomenon related to the generation of phase contrast appeared in the image which obscured the edge. Attempts to remove it by optimizing the phase contrast configuration proved fruitless. Detecting the open edge of the trap was therefore uncertain within several micrometers. By selecting a somewhat arbitrary position for one open edge and expecting that the opposing edge was located a known distance away, *i.e.* 40  $\mu\text{m}$ , a coordinate system could be erected in the first image of the series. The interference phenomenon was distinct for each trap and did not vary significantly over time. A signature profile could be extracted for each image. By comparing the signature profile at each time with the first recorded profile it was possible to maintain the coordinate system with super-resolution accuracy between images.

## Image intensity normalization

To counter variations in image intensity, an active intensity normalization method was applied to each phase contrast image. A further consequence of the interference phenomenon situated at the open edges of the trap was that cell signatures here were distorted. For this reason analysis efforts were abandoned in the afflicted regions, the bounds of which were determined by manual inspection. The active region in which cells and molecules could contribute to the analysis was roughly  $40 \times 30 \mu\text{m}^2$ . Each raw phase contrast image was cropped according to the predicted bounds of the trap, rescaled in intensity and finally reduced in size uniformly by 58%. The size reduction was performed to increase computational speed in subsequent protocols.

## MicrobeTracker

The process of outlining cells in an image is often referred to as cell segmentation. A software suite dedicated to automatic segmentation and tracking of cells in phase contrast images, MicrobeTracker [50] was used and further developed for our data. MicrobeTracker is implemented in MATLAB and uses a composite detection strategy where binary cell clusters are first predicted and subsequently refined by applying an active contour model to determine the cell outline. However, phase contrast was less efficient for our

cells which were closely grouped and the signature of an individual cell became highly dependent on its local environment.

Segmentation of an image can be done with or without prior knowledge of where cells are located. The former case is referred to as *de-novo segmentation*, which we found to be very challenging for our data. The latter case, *i.e.* when known previous locations are used to predict where the cell is currently located, was considerably easier. This will be referred to as *tracking based segmentation*. In the initial stages of development we would correct errors in *de-novo* segmentation performed on the first image manually then perform tracking based segmentation of the subsequent images.

### Cell mobility and tracking

When tracking cells over time we quickly realized that displacements would often overwhelm the algorithms. The first action was to improve the time resolution in imaging. We increased the frequency of acquisition from one image per 3 min to one per 30 s. This dramatically improved our ability to track cells. The motion of cell mass in the traps had at least two time scales. The longer time scale is related to growth rate and could be described by viewing the micro-colony as a continuum rather than a discrete collection of individuals. The flow of cells was directed toward the openings of the traps. The average rate of displacement was dependent on the location of cells relative the mid-line of the trap at which point the net flow was zero. The velocity increased exponentially with the distance from the midline. Every line in the trap parallel to the mid-line contained a roughly equal amount of cell mass which was expanding at the same rate. The rate of expansion at a particular line was then the sum (or integral) of all the lines which separate it and the mid-line. The velocity at the open edges of the trap would therefore be considerably greater as compared to the center of the trap.

The second and shorter time-scale required regarding the micro-colony as discrete individuals. Local displacements appeared to occur randomly within the micro-colony and could re-arrange sub-regions of it in less than one second. A cell which displaced would leave a gap in the colony and neighboring cells would proceed to fill it. Effects were local, however, the closest neighboring cells could displace drastically. As the time-scale of these displacements approached the exposure time of the acquisition, increasing the time resolution to account for these displacements was deemed practically impossible.

We therefore continued to manually correct segmentation between images. With manual curation of segmentations, 100 subsequent images (~50 minutes of experimental time) could be analyzed in 15-20 h of full time labor of one analyst.

## Active pole tracking

We guessed that if we could improve the initial guess for the location of cells poles, we could improve tracking. The strategy was to collect true examples of cell poles and extract a signature set of feature values and construct a probability model using these. During tracking, feature values sampled at randomly selected positions in the vicinity of the last known location of the pole could be evaluated using the model (see paper II, supplemental under *Pole searcher* and *Classification*). Two expectations of cell motion were incorporated into the initial guess. First, to capture the continuum motion of the cell colony (see under *Cell Mobility and tracking*), estimates of local displacement were made using the optical flow method [51]. Second, as cells were expected to grow, the increased separation between poles due to growth was estimated using a finite difference approximation based on the cell length at the two previous time points (see paper II, supplemental under *Pole searcher*). The introduction of the pole searcher brought with it a huge improvement to the quality of tracking.

## Division classifier

It was noticed that many cells were divided incorrectly. To develop a more accurate division classifier, we began by defined image features we believed were related to division. Next, examples were collected of true and false divisions and the corresponding feature values were extracted. Using linear discriminant analysis, an optimal linear threshold for separating the two classes was obtained (see paper II, supplemental under *Division function*). This threshold was then applied to the extracted feature values of each cell at each frame and the cells which exceeded the threshold were classified as divided. This addition greatly improved the confidence in the complete cell cycles we obtained from the analysis.

The division classifier was later analyzed for its accuracy and precision by monitoring cells which express fluorescently labelled division proteins, FtsQ-GFP [52]. FtsQ-GFP forms a band which contracts during cell division, ultimately collapsing to a point [52]. The cells were imaged at a frequency of  $1/1 \text{ min}^{-1}$  and it was found that as the contraction ended the point vanished in less than one minute. A second division classifier, a fluorescence classifier, was devised which could identify this abrupt loss of fluorescence signal (see paper IV, figure 2 and supplemental under *Detection of FtsQ-GFP signal drop in single cells*). The phase classifier was found to be slightly early, on average dividing cells  $\sim 1$  minute before the fluorescence classifier. The precision of the phase classifier was determined to be sufficient as the standard deviation of the difference between the predictions of the phase classifier and the fluorescence classifier was 2.7 min (see paper IV, figure 2).

## Management of tracking errors

The method was still too inaccurate for autonomous work. By developing a method capable of autonomously terminating erroneously segmented cells at each frame we hoped to improve the quality of the remaining data without the need for manual corrections. As in the case of the division classifier, an approach utilizing a binary class description of the task was adopted as in the cases of poles and division classification (see paper II, supplemental under *First error detector*). The effect of the introduction of the error classifier on the quality of the data was striking. As the pole searcher enabled the algorithm to cope with smooth cell displacements, the error classifier would terminate many of the errors committed as a result of fast, discontinuous displacements. The stability and reliability of the algorithm was markedly increased and as a result the length of the intervals under which we could allow it to analyze independently increased.

One of the features which were used for this error classifier was the intersection between a cell's outlines and those of its neighbors. This feature is a rational choice since cells should not overlap. However, the intersection between two cell outlines is the same for both cells and as a consequence correct cells could be classified as erroneous. The termination of cells would cause the number of remaining cells to decrease in a probabilistic fashion as the segmentation process advanced. Further, the error classifier would not be as effective as surviving cells would find themselves isolated.

## Active cell tracking

In order to slow the decay of our segmentation, we developed an active cell tracking algorithm, which would attempt to relocate the cells which were classified as erroneous. An additional error classifier was also developed (see paper II, supplemental under *Cell tracker* and *Second error detector*). With these additions the segmentation process was more accurate and sufficiently reliable to work autonomously. The processes could maintain an acceptable number of remaining cells per trap throughout 1000 subsequent time points (see paper II, figure 2 b). This corresponded to 8.3 hours of experiment time and an average of ~1000 complete cell cycles were acquired at each position. Unfortunately, the segmentation process was now extremely slow and to speed up the algorithm we developed a parallel computing version which would allow an image to be processed in 5 minutes.

## De-novo segmentation

To counter the decay of the segmentation process, we manually corrected images at intervals of 100 time points and allowed the re-found cells to replenish the colony. At this stage we had avoided the inevitable, the devel-

opment of an accurate de-novo segmentation algorithm, for as long as possible. As the strategy of binary classification had proven successful in its previous applications, we adopted the same approach for cell classification. The de-novo segmentation methods will when applied to an image, detect cells correctly, incorrectly or not at all for a given set of parameters. By performing the segmentations using a range of parameter sets, a large proportion of all cells could be found correctly at least once. To discard erroneous cells an error classifier was developed (see paper IV, under *Materials and methods, Cell segmentation and tracking*). Further, a likelihood of a remaining cell being a true cell could be derived and used as a score to select the most likely instance of a cell which was detected several times.

The de-novo segmentation method was applied to detect cells in all phase contrast images of a series. In the first frame, it was applied to the entire image. In subsequent frames it was applied, after all other operations had been performed, on the regions of the image which were vacant. The cells which were found were added to the segmentation process. This transformed the method to a non-decaying process which would maintain a quasi-steady number of segmented cells per trap (see *Figure 9* right). Under conditions of fast growth, the de-novo segmentation algorithm was found to detect roughly ~140 cells per image which represents ~60% of the total number present. The combination of continuous de-novo segmentation and tracking based segmentation increased the number of cells in the first ~30 images (15 min) of until reaching a steady state of ~200 cells per image (see *Figure 9* right). This represents 90 % of the cells that are expected to be present in the region of interest in each trap.

### An integrated method

The rate at which cell cycles are completed is expected to be  $5.1 \text{ min}^{-1}$  and position under conditions of fast growth and  $2.4 \text{ min}^{-1}$  and position under conditions of slow growth (see paper IV, figure S6). The segmentation and tracking method was estimated to capture more than 60% of all cell cycles (see paper IV, supplemental under *Efficiency in acquisition of cell cycles*). For example if six micro-colonies were monitored in parallel, ~20 complete cycles would be obtained per minute of experiment time during fast growth and ~15 cycles during slow growth. Fluorescent molecules were detected with super resolution accuracy (see *Figure 6* right) and the obtained coordinates were superimposed on the segmented outlines (see *Figure 6* left) using the registration information. The intracellular locations were determined by projecting these coordinates on the major (axial) and minor (radial) axis of the cells they were determined to belong to. The method was now able to meet all the stipulated criteria and was completely automated from the start of acquisition.

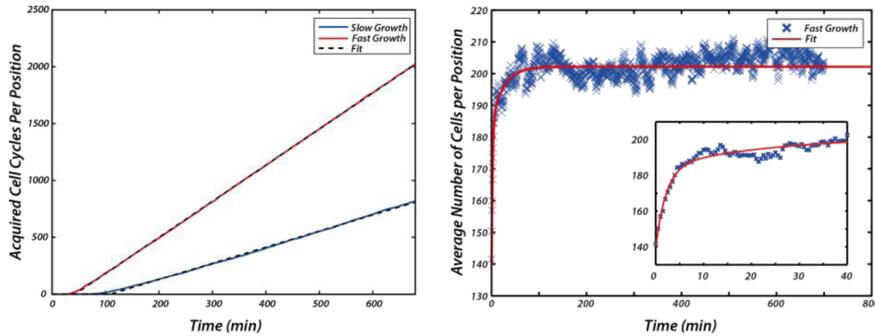


Figure 9. The acquisition of cell cycles and number of cells per position. Left: Acquisition rate for slow growth is 1.4 cycles per min and position and 3.2 per min and position for fast growth. Right: average number of detected cells per position over time. Right inset: initial increase.

## Life cycles in a microfluidic chip

### Cell size

The observed cell widths were highly uniform as the average cell under conditions of fast growth was  $1.15 \pm 0.02, 1.6 \cdot 10^{-4}$  (SD, SEM)  $\mu\text{m}$  wide and the average cell under conditions of slow growth was  $0.92 \pm 0.01, 2.0 \cdot 10^{-4}$  (SD, SEM)  $\mu\text{m}$  wide (see paper IV, figure S3 a). Here we realized that the measured widths were greater than the depth of the traps the cells were residing in under both conditions, *i.e.*  $0.9 \mu\text{m}$  for fast growth and  $0.8 \mu\text{m}$  for slow growth. It is therefore likely that cells were squeezed and not strictly circular at a cross-section perpendicular to their main axis. As variations were much greater in cell length than in width, the length of a cell could be used to represent its size (see paper IV, figure S3 b).

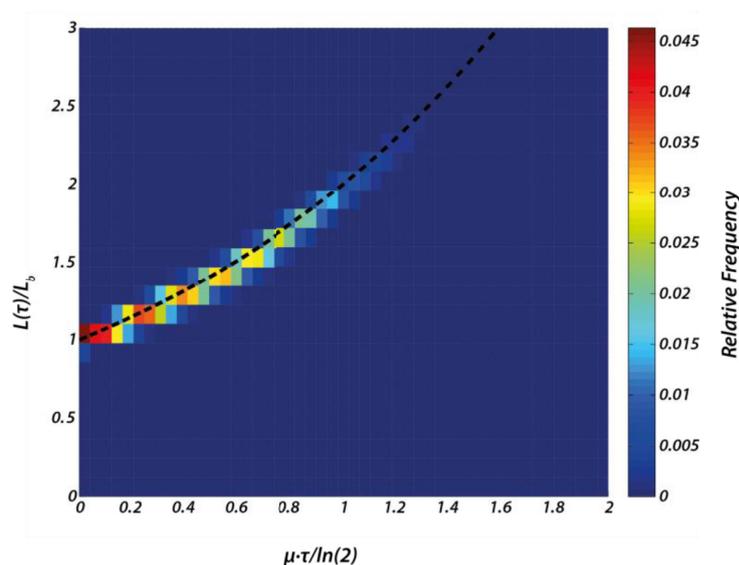
The average length at birth was  $3.00 \pm 0.42, 3.4 \cdot 10^{-3}$  (SD, SEM)  $\mu\text{m}$  and  $2.38 \pm 0.44, 6.8 \cdot 10^{-3}$  (SD, SEM)  $\mu\text{m}$  during fast and slow growth respectively (see paper IV, figure S3 c). The corresponding average division lengths were  $5.93 \pm 0.82, 6.7 \cdot 10^{-3}$  (SD, SEM)  $\mu\text{m}$  and  $5.93 \pm 0.82, 6.7 \cdot 10^{-3}$  (SD, SEM)  $\mu\text{m}$  (see paper IV, figure S3 d). We could therefore conclude that as expected, cells living under conditions of fast growth had a greater capacity of storing resources as biomass. The average division lengths were not quite twice the average birth length, which at first glance suggests that the cells were shrinking. This was however not reflected by birth- or division lengths over the course of the experiments (see paper IV, figure S9).

## Cycle time

The average cell cycle under conditions of fast growth lasted  $31.8 \pm 7.92$ ,  $0.064$  (SD, SEM) min and under conditions of slow growth the average cell cycle lasted  $83.2 \pm 35.1$ ,  $0.55$  (SD, SEM) min (see paper IV, figure S3 e).

## Exponential growth

To establish if cells grew exponentially under our experimental conditions, we compared the time course for the observed cell lengths for each cell cycle with an exponential growth model. We found that this description captured the data well (see *Figure 10* for conditions of fast growth and paper IV, figure S7 for both conditions). This allowed us to derive an additional dynamic observable for each cell cycle as the best fit, namely the growth rate,  $\mu$ . The average growth rate was  $0.0215 \pm 0.0043$ ,  $3.5 \cdot 10^{-5}$  (SD, SEM)  $\text{min}^{-1}$  for fast growth and  $0.0089 \pm 0.0029$ ,  $4.5 \cdot 10^{-5}$  (SD, SEM)  $\text{min}^{-1}$  for slow growth (see figure paper IV, figure S3 f). The growth rate expected from the cycle time (see introduction under *Requirements for concluding a cell cycle*) was computed for each cell cycle and the average was found to be  $0.023 \pm 0.0058$ ,  $4.73 \cdot 10^{-5}$  (SD, SEM)  $\text{min}^{-1}$  and  $0.0099 \pm 0.0055$ ,  $8.56 \cdot 10^{-5}$  (SD, SEM)  $\text{min}^{-1}$  for fast and slow growth. This predicted slightly faster growth for both conditions than we could observe.



*Figure 10.* Exponential growth. Observed values under conditions of fast growth of size and time have been scaled, as indicated on axes, to allow the comparison with an ideal curve (black dashed line). See paper IV for equivalent for slow growth.

## Variations over space and time

The growth rate, birth length, division length and cycle time, did not vary between different micro-colonies which were imaged in parallel (see paper IV, figure S8). This indicated that conditions were uniform for all micro-colonies in the microfluidic chips during acquisition. However, some variation did occur over the time course of the experiments. For example, during the beginning and the end of acquisition, the average cycle times were lower (see paper IV, figure S9). This was more noticeable for slow growth than for fast growth. Further, this trend also extended to the remaining cycle observables, and was always more pronounced for slow growth than for fast growth. This systematic variation over the course of our experiments appeared to be a consequence of the time-delayed nature of the cell cycle. At the beginning and end of the acquisition, only cells which divide early could contribute to the set of observed cell cycles.

A steady state for the average cycle time was reached after 50 minutes for fast growth and 200 minutes for slow growth (see paper IV, figure S9 lower left). This constituted roughly twice the generation time for each condition. This phenomenon should occur during any observations of cell cycles and as a rule of thumb, the acquisition time should always be much longer than four times the expected generation time.

Further, the time before losing a cell from the analysis, either by the cells physically vacating the trap or by failing to track them was well described as an exponential random variable. For this reason any cell which divided early retained a greater probability of contributing to the set of observed cell cycles. The increased representation in our data for early dividers could explain two inconsistencies; the faster growth expected from the cycle time and the average size at division being less than twice that of the average birth size. We noted that the magnitude of these inconsistencies decreased as our analysis accuracy increased. However, it was not possible to fully eliminate the phenomenon since no information could be gathered from cells which had vacated the traps.

## Interdependencies of cycle observables

Dependencies between variables are often expressed in terms of correlation. The Pearson correlation coefficient,  $r$ , will give a result between -1 and 1 and the value reflects the amount of the variation which can be accounted for by a linear model. However, if the sample is mixed and the different types are not related, the correlation coefficient will represent the mixture rather than one distribution. Our data contained correctly and incorrectly segmented and tracked cells. To allow cell cycles to contribute according to frequency rather than the observed values, we compared the joint relative frequen-

cies to the density function of the bivariate normal distribution and obtained the correlation coefficient,  $\rho$ , as a parameter of the best fit.

For example, the correlation between the observed growth rate,  $\mu$ , and the growth rate expected from the cycle time (see introduction under *Requirements to conclude a cell cycle*), was  $\rho = 0.50 \pm 5.2 \cdot 10^{-5}$  (95% DSCI) for fast growth and  $\rho = 0.70 \pm 4.2 \cdot 10^{-5}$  (95% DSCI) for slow growth. The corresponding Pearson correlation coefficients were  $r=0.50$  and  $r=0.49$  for the same samples. In the case of fast growth both estimates were consistent. However, in the case of slow growth they diverged. When inspecting the joint relative frequencies (see *Figure 11*) the eccentricity of the data appeared to be greater for slow growth, than for fast Growth, as indicated by  $\rho$  but not by  $r$ . It should however be noted that the alternative method requires considerably greater statistical power. In the cases where this power is strained, the Pearson correlation coefficient will be presented also for a comparison.

### Cycle time and growth rate

As was seen in the previous section in the case of slow growth, the dependency between the growth rate and the growth rate expected by the cycle time were such that they represented each other relatively accurately. For fast growth, the correlation was moderate. Therefore using for instance the cycle time of an individual as a representative of its growth rate would entail a weaker prediction for cells under conditions of fast growth than under conditions of slow growth. This could potentially explain the variation of the growth rate for slow growth during the early and late stages of our experiments (see paper IV, figure S9 lower right). By extension, similar co-dependencies between birth length and cycle time and division length and the cycle time could explain the variations we observed over the course of an experiment for slow growth (see paper IV, figure S9).

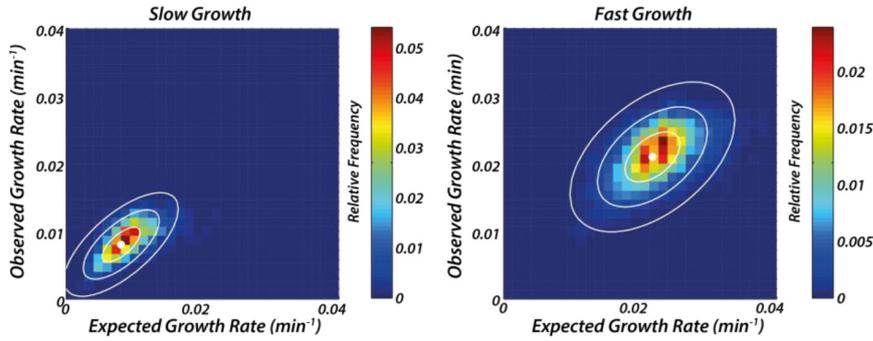


Figure 11. Joint distribution growth rate and growth rate expected from the cycle time. White ellipses: best fit of density function of the bivariate normal distribution, levels capture 0.6827, 0.9554 and 0.9973 of the observational mass. White dots: modes for the best fit.

### Inheritance of cycle properties

As expected for a symmetrically dividing bacterium, the birth lengths for siblings were highly correlated for both growth conditions as  $\rho=0.81 \pm 3.4 \cdot 10^{-3}$  (95% DSCI) and  $\rho=0.81 \pm 4.5 \cdot 10^{-3}$  (95% DSCI) for fast and slow growth respectively (see paper IV, figure S10, row 1, column 1-2). The growth rates were moderately correlated between mother and daughter for fast growth  $\rho=0.51 \pm 1.1 \cdot 10^{-4}$  (95% DSCI) and for slow growth  $\rho=0.54 \pm 4.21 \cdot 10^{-5}$  (95% DSCI) (see paper IV, figure S10 row 4 column 3-4). However, growth rates for sibling pairs were highly correlated both for fast growth  $\rho=0.81 \pm 6.0 \cdot 10^{-5}$  (95% DSCI) and for slow growth  $\rho=0.72 \pm 3.6 \cdot 10^{-5}$  (95% DSCI) (see paper IV, figure S10, row 2, column 3-4).

The cycle time was less correlated between mother and daughter for both conditions of fast growth  $\rho=0.097 \pm 0.17$  (95% DSCI) ( $r=0.14563$ ) and  $\rho=0.070 \pm 0.82$  (95% DSCI) ( $r=0.085362$ ) for slow growth (see Figure 12). Therefore the determination of the cycle time for a cell was essentially memory-less between generations. When we compared the cycle times for siblings we found that these were considerably more correlated as  $\rho=0.54907 \pm 0.094$  (95% DSCI) ( $r=0.60$ ) for fast growth and  $\rho=0.69 \pm 0.67$  (95% DSCI) ( $r=0.58$ ) for slow growth (see Figure 12).

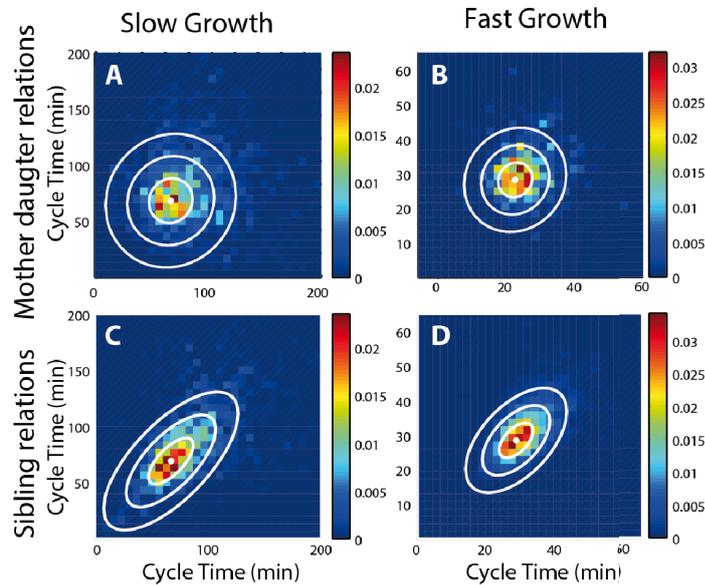


Figure 12. Inheritance of cycle times. Background: Density of relative frequencies. White ellipses: best fit of density function of the bivariate normal distribution, levels capture 0.6827, 0.9554 and 0.9973 of the observational mass. White dots: modes for the best fit.

## Gene expression and regulation over the cell cycle

### Cell cycle dependent bursts of gene expression

The promoter for the lactose operon is regarded as a strong promoter. The cell cycle contains two putatively critical stages for a gene which is regulated by a repressor TF.

- At division TFs are distributed between the nascent sibling cells and the relative fluctuations in the number of TFs received by each sibling is expected to increase for low abundant species such as LacI [53]. A cell may receive fewer TFs than gene copies, leaving one or more copies un-repressed until new TFs can be synthesized.
- As the gene is replicated, two new copies are expected to appear almost instantaneously ( $\sim 1$  s). At least one of these copies would be un-repressed until a repressor has time to occupy its operator.

An interpretation of the repression fold of the lactose operon under the occupancy model for gene regulation (see paper IV, supplemental under *The operator occupancy model at equilibrium*) is that it should be un-repressed for a maximum of 1/1000 of the cell cycle, which corresponds to  $\sim 1-2$  s under

conditions of fast growth. The search time for LacI was found to be ~30-60s, in living cells [39]. Taken together, this suggested that lapses in regulation should occur less frequently than once per cell cycle. The method we had developed could potentially allow us to determine the frequency and the distribution of expression events over the cell cycle, if we could monitor expression from the lactose operon over many cell cycles.

### Constructs for monitoring burst of expression

Choi *et al.* studied the expression from the *lacY* gene by genetically replacing the gene with a reporter construct, Tsr-Venus (see introduction) [54]. By studying the same strain we could profile the expression of the lactose operon. The *lacI* gene is auto-repressed and a cell cycle dependent lapse of repression could possibly lead to bursts of expression. The Tsr-Venus construct cannot be used to gather information about a TF, such as LacI, which relies on its mobility in the cytosol for its function. However, a previous study [39] suggested that LacI-Venus molecules could be immobilized for sufficiently long (~1 ms) while non-specifically bound to DNA to be detected. By studying expression of both *lacI* and *lacY* we could potentially [39] quantify the frequency and amplitude of expression events and determine any dependency on the cell cycle.

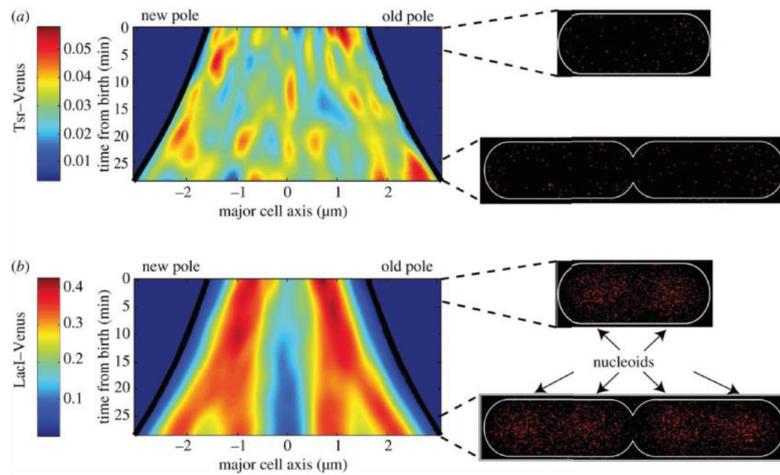
### Counting fluorescent molecules as they are synthesized

Accurate counting of fluorescent particles requires either completely photostable fluorophores which do not bleach or blink significantly or that all molecules can be irreversibly bleached by observation. As the bleaching regime was reported previously [41, 54] we attempted to bleach molecules between observations with various durations and intensities of laser exposure. When continuously exposing cells expressing Tsr-Venus we found that the number of molecules would decrease initially but that fluorescent spots would continuously appear. Further, it was realized that the cells were much more sensitive to laser exposure than anticipated. For the regimes which they seemed to be able to cope with, the number of fluorescent molecules was expected to decrease by 20% per observation. To account for the remaining fluorescent molecules we developed a maximum likelihood method to estimate how many new molecules had been produced since the last observation (see paper II, under *Materials and methods*). The accuracy of this estimation was however dependent on molecules bleaching quickly.

### Gene expression rates over the cell cycle

Cell cycles were synchronized at the time of birth and the observed average rates of expression for each construct was derived over the cell cycle (see

paper II, figure 5 c-d). For both the *lacY* and *lacI* genes, the rate of expression increased over the cell cycle. No distinct regions of increased sensitivity were apparent. We proposed that this could be due to too few observations. The intracellular distribution of LacI-Venus was found to adhere to two clusters at the start of the cell cycle and as the cycle progressed these would bifurcate into four clusters (see *Figure 13* lower).



Ullman G et al. *Phil. Trans. R. Soc. B* 2013;368:20120025

*Figure 13.* Intracellular locations of Tsr-Venus and LacI-Venus. Density of observed locations of Tsr-Venus encoded at *lacY* (upper) and LacI-Venus encoded at *lacI* (lower) over the cell cycle.

### The dissociation time of a TF

Gene expression can also be studied by studying regulation [39]. The equilibrium occupancy model of transcription factor regulation suggests that the association time for TFs to find their specific operator,  $\tau_{\text{on}}$ , and the dissociation time,  $\tau_{\text{off}}$ , determines the expression ratio of the regulated gene (see paper IV, supplemental under *The operator occupancy model at equilibrium*). The association time for LacI in living cells had previously been measured [13, 39]. If a bound TF was monitored directly by time-lapsed observations, distinguishing the cases when disappearance was due to dissociation rather than bleaching would be difficult. If observations were made less frequently to increase the longevity of the fluorophore, dissociation followed by association between observations could not be excluded.

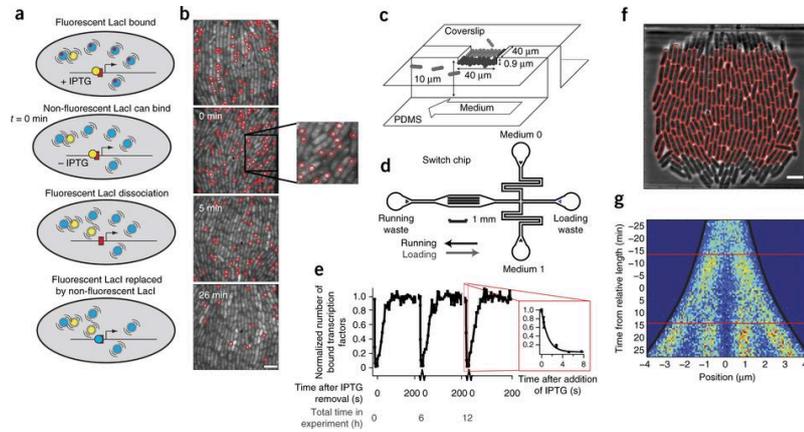
### An *in vivo* chase assay

An experiment which we hoped could allow us also to measure the dissociation time of LacI was designed (see *Figure 14 a*). The reporter construct was a fluorescently labeled version of LacI which had been mutated so as to be insensitive to chemical induction by IPTG, designated LacI<sub>s</sub>-Venus [55]. The gene coding for this construct was introduced into the chromosome to replace the native *lacI* gene. LacI and LacI<sub>s</sub> will when fused to a fluorescent protein, form dimers instead of tetramers [39].

The wild type *lacI* gene was then introduced on an extra-chromosomal genetic element, or plasmid, with a strong arabinose sensitive promoter, *p<sub>BAD</sub>*, which would express only when arabinose is present in the medium [56]. This plasmid was then introduced into the strain which carried the LacI<sub>s</sub>-Venus construct. Under conditions when arabinose was absent and IPTG was present in the medium, the majority or at least a significant proportion of all LacI molecules were expected to be LacI<sub>s</sub>-Venus and only LacI<sub>s</sub>-Venus molecules were expected to bind to the operator sequence. If arabinose was introduced in the medium and IPTG removed, the population of LacI molecules was expected to shift to mainly consist of non-fluorescent LacI tetramers, which could associate to the operator sequences. Each of the non-fluorescent particles could find the operator sequence at least as fast as a LacI<sub>s</sub>-Venus molecule. In this way re-binding of the operator site by LacI<sub>s</sub>-Venus could be avoided (see paper III, under *Materials and methods*).

Occupancy of the operator by LacI<sub>s</sub>-Venus could be detected by imaging the cells at long exposures, ~4 s. The lactose repressor was expected to remain bound for a substantial part of the cell cycle. Using microfluidic chip model B, the chemical environment of the cells could be rapidly exchanged to contain either IPTG or arabinose. Imaging a trap would bleach a fraction of LacI<sub>s</sub>-Venus molecules and therefore a position was only imaged once during an experiment. As there were 51 traps available in the chip, this became the upper limit to the number of time points which could be observed (see paper III, under *Materials and methods*).

After the experiment, the medium could be exchanged to deplete the cells of wild type LacI tetramers by growth. After sufficient recuperation, the experiment could be repeated as many times as required with high precision (see paper IV, figure 1 e).



**Figure 14.** (a) The single-molecule chase assay. Yellow: fluorescent LacI dimers red box: lac operator, Blue: non-fluorescent wild-type LacI tetramers. (b) Fluorescence images (4-s exposure) taken before and at different time points after the removal of IPTG. Scale bar, 4  $\mu\text{m}$ . Red circles : detected operator-bound LacI-YFP. (c,d) Microfluidic chip B (d) contains 51 traps as illustrated (c). (e) Medium switch-induced transcription factor dissociation and association. (f) Automatically segmented cells using a phase-contrast image. Scale bar, 4  $\mu\text{m}$ . (g) Intracellular positions of bound LacI-YFP molecules (x axis) mapped to the cell replication cycle (y axis). Individual cell replication cycles are synchronized so that the time of 0 min always implies a cell length of 4.25  $\mu\text{m}$ . Horizontal lines: average times for cell divisions.

## Non-equilibrium models of regulation

Using the chase method we determined that LacI<sub>s</sub>-Venus would dissociate from *lacO*<sub>l</sub>, after 5 minutes and from *lacO*<sub>sym</sub>, after 10 minutes (see paper IV, figure 2 a). The result for the latter operator was inconsistent with the equilibrium occupancy model, given the observed repression fold. Instead we investigated three non-equilibrium models which could possibly account for the observations we had made (see paper IV, figure 3).

One of these (see paper IV, figure 3 d) was based on the reports by Kuhlman & Cox of gradients of transcription factors in the cell [57]. One possible implication was that local concentrations surrounding the operator site could be higher and lead to faster association than we could observe using a method which relied on chemical induction to start LacI search from a uniformly distributed starting point in the cell.

This model could potentially be tested using the automated method we developed. The intracellular distribution of non-specifically bound LacI-Venus molecules could be determined using short (2 ms) exposures to locate fluorescent TF molecules as in paper II. Kuhlman & Cox reported that LacI molecules would remain proximate to the chromosomal location from which they were expressed [57]. To test this we studied several strains with LacI-

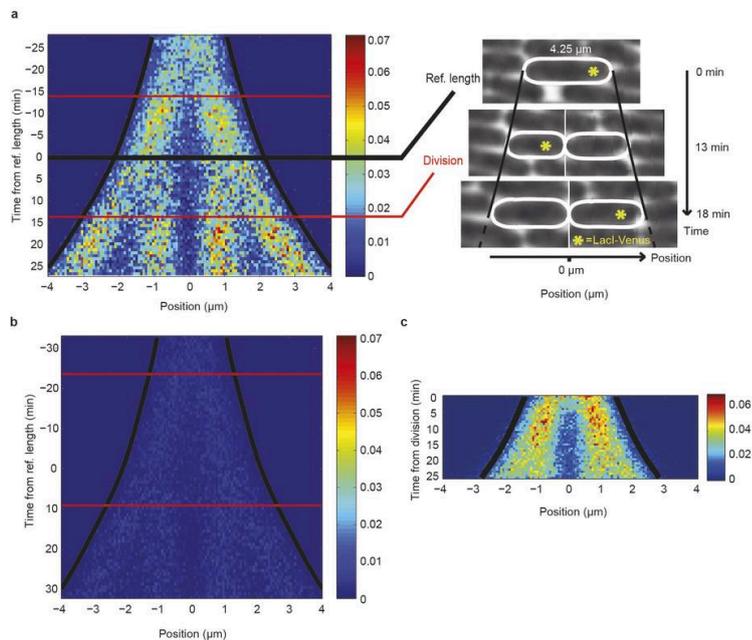
Venus encoded at different chromosomal positions. To ensure that only non-specific interactions were monitored, all strains had the specific operator sites removed. The results were compared to localization data of specific TF binding at the origin, mid replicore and the terminus.

From our observations we could not confirm that the chromosomal location of the *lacI-venus* gene influenced the localization of LacI-Venus (see paper III, figure S5 a-b). We could however see local gradients in the cell (see paper III, figure S5 a). To determine the importance of DNA binding on localization, we also studied a version of LacI-Venus, which has been mutated so that its DNA binding property was impaired [39]. We found that this version was uniformly distributed over the cell (see paper IV, figure S5).

We concluded that under our experimental conditions, gradients of LacI molecules were established by non-specific interaction to all DNA rather than the chromosomal location from which they were expressed. Further, using stroboscopic illumination it was possible to track the movements of individual LacI-Venus molecules within cells in the microfluidic chip. By mapping the trajectories to an intracellular coordinate system, it was determined that LacI-Venus traversed the cell within seconds (see paper IV, figure S6 and supplemental under *Are there spatial gradients of LacI in rapidly growing cells?*). We could therefore exclude this model as the cause for our findings.

## Cell cycle dependent lapses in TF regulation

The dependency of regulation of gene expression on the phases of the cell cycle was investigated by time-lapsed microscopy of cells expressing LacI-Venus molecules and having one *lacO<sub>sym</sub>* operator available for binding. Occupancy of the operator site would be detected as diffraction limited spots in fluorescence micrographs obtained by exposing cell colonies for long durations (4s) at a frequency of  $1/3 \text{ min}^{-1}$ . If occupancy was perturbed at cell division or at some distinct cycle coordinate corresponding to the replication of the operator site, this experiment could possibly allow us to observe it as lapses in occupancy. The resulting fluorescence information for individual cell cycles revealed no distinct patterns. We again tried to synchronize our observations at division and study the average number and intracellular location of LacI-Venus at time points after division. This did not help us arrive at any conclusions. We tried filtering the data on the basis of similar durations for cell cycles, with disappointing results. Finally we tried sorting observations according to the length of the cells. This provided a greater coherency in the localization patterns of LacI-Venus bound to the promoter. Further we found that if cell cycles were synchronized at the time the cells were a specific length, the coherency in localization was maintained (see *Figure 15*). However, no distinct lapses of occupancy corresponding to the search time of LacI were observed even when synchronizing our data in this way.



**Figure 15.** Occupancy of the lactose promoter by LacI-Venus over the cell cycle. (a) LacI-Venus binding to  $lacO_{sym}$ . (b) LacI-Venus localization without a specific operator sequence. (a-b) Intracellular location are organized so that at time zeros all cells are  $4.25 \mu\text{m}$  in length. (c) Data as in (a) organized according to time after division.

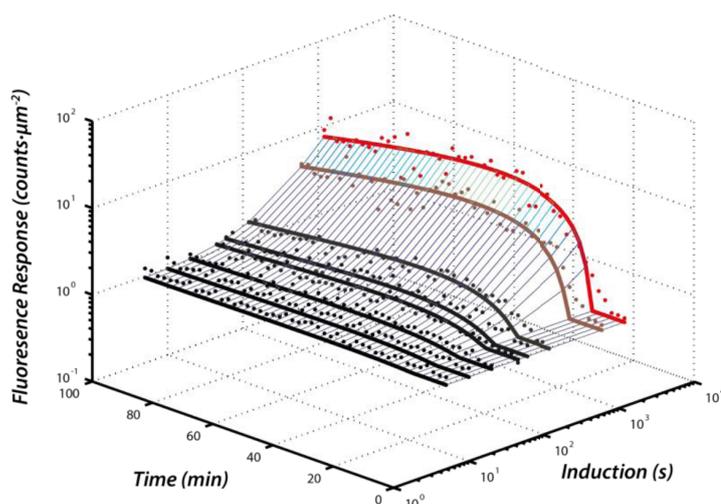
### Investigating a post-replication refractory period for gene expression

Although we could not confirm any cell cycle dependent perturbations to occupancy of the lactose promoter, we could not exclude that such events could be masked by averaging. We estimated that a lapse in occupancy of at least 10s would accompany replication of the lactose promoter. To achieve a repression fold of 1000 times, the operon should be expressed on average for 1-2 s per cell cycle. We therefore hypothesized there was some mechanism to dampen or silence expression during the time required for a transcription factor to associate to the promoter after replication. We designed an experiment to test this hypothesis which is described in detail under (*Appendix: Measurement and determination of the maturation time of Venus in live bacterial cells growing in a microfluidic chip*). The experiment, the results and their implications for gene expression are discussed in the following sections.

## Measuring the expression response to induction

Expression from the lactose operon can be chemically induced using IPTG and chip model B could be used to induce cells rapidly and reliably. If a refractory period existed, we expected that there should be a minimum induction time required for the gene to respond to induction. Cells modified to express a fluorescent reporter construct LacY-Venus from the *lacY* gene could generate a detectable fluorescence response to induction pulses of high concentrations of IPTG [54]. The durations of these pulses could be achieved with high precision and our tests indicated that induction pulses as brief as 1 s were possible.

The experiment was performed by acquiring images of colonies at designated and uniformly spaced time intervals. Each colony was imaged once to eliminate the effects of bleaching (see *Figure 20*). After the zeroth image, the acquisition program would execute the induction pulse and the remaining images acquired would allow us to determine the fluorescence generated as a response to induction over time. This procedure could be programmed to repeat several times using a predetermined set of induction durations. A time-delay between completing an acquisition and commencing a new session was included in each cycle during which the cells were allowed to deplete of LacY-Venus molecules.



*Figure 16.* Expression response of *lacY* to induction. See appendix for details.

## The maturation time of fluorescent proteins

We quickly realized that the fluorescence generated by an induction pulse would accumulate over long periods of time (see *Figure 16*). This time was in fact so great that the effect of dilution due to growth had to be accounted

for when estimating the fluorescence response. The reported maturation time for Venus is 7 min [41, 58]. By assuming that the maturation time was an exponentially distributed random variable and that molecules are diluted exponentially according to the average growth rate we determined the maturation time of Venus under our experimental conditions was 54 minutes (see Appendix: under *Results*). The maturation time dictates the certainty in predicting when a fluorescent molecule was synthesized. For example if a molecule is made at a known time the probability that the molecule would mature and fluoresce within twice the maturation time is 0.86. Under our conditions this would entail that molecules could mature and become fluorescent up to 3 times the generation time after they were synthesized. The implication of these findings was that we could not determine cell cycle dependent patterns for the synthesis of molecules either from individual cell cycles or by averaging over many cell cycles, regardless of how they were compared.

Information on the kinetics of localization of molecules as in the case of TF occupancy appeared to be less sensitive to the effects of a long maturation time. However, an implication was that the majority of all reporter molecules inside a cell were non-fluorescent. Therefore the variance in the fluorescence information was primarily related to maturation rather than variations in copy number. Under such conditions, our observations had to be combined and averaged in order to draw conclusions. Averaging can destroy information on the underlying distributions of a mixed sample. Consequently it was necessary to find ways of comparing data from individual samples which separates the underlying distributions. In the cases of measuring the association and dissociation times of LacI-Venus, it was possible to synchronize molecules with chemical cues and then compare molecular observations according to the time that elapsed since synchronization. In the case of occupancy of the lactose promoter LacI-Venus over the cell cycle, it was possible to achieve a greater separation of distributions by organizing our data according to size rather than according to time elapsed after division.

## Regulation of the cell cycle

### Chromosome management

Our results for the occupancy of the *lac* promoter suggested that the chromosome was managed according to size rather than time after division. DNA replication is an essential process involving the chromosome and during conditions of fast growth, the chromosome of *E.coli* is expected to be in a constant state of replication. To study the replisome, we utilized that the  $\epsilon$  subunit of the polymerase III holoenzyme, DnaQ, can be genetically fused with a yellow fluorescent protein, Ypet, without effects on growth [59]. The reporter construct, DnaQ-Ypet, is immobilized during active replication and

can be detected using fluorescence microscopy as diffraction limited spots [59]. We studied cells carrying DnaQ-Ypet using our method under conditions of fast growth. We found no distinct location patterns when organizing the resulting intracellular locations of replisomes according to the time after division (see paper IV, figure 3 d). Further, we did not find any distinct periods of sensitivity for the average number of replisomes per cell (see paper IV, figure 3 d lower).

When we organized the data according to cell size, here estimated by cell volume, we found that the coherency of the data increased substantially (see paper IV, figure 3 c). Replisomes were clustered in replication sites inside the cells and these sites would progress as intracellular trajectories over the cell cycle and bifurcate during a short size interval. These trajectories appeared to be overlapping. A size interval during which the average number of replisomes per cell would increase at a higher rate (see paper IV, figure 3 c lower) could be observed. We drew the conclusion that cells initiate replication according to size rather than the elapsed since division.

### Replication initiates according to size

By assuming that the initiation size was normally distributed we could derive a putative initiation size as the best fit of a cumulative normal distribution function to the observed average number of replisomes per cell. We found that under this model, the initiation volume was  $4.0 \mu\text{m}^3$  (see paper IV, figure 3 c lower red solid line). By organizing our data in time from the time cells achieved the initiation size, we found that the coherency was maintained (see paper IV, figure S4). Replication initiation was reported to be regulated according to size by Hill *et al.* for cells under condition of fast growth [60]. However, Kleckner & Bates reported that initiation was regulated according to time after division under conditions of slow growth [26].

We were therefore curious to see how replisomes behaved under conditions of slow growth. Again, organizing our data according to size gave a more coherent result as compared to time after division (see paper IV, figure 2 a-b). As in the case of fast growth we found that replisomes were clustered in replication sites which formed trajectories over the cell cycle. Here, these trajectories were not overlapping (see paper IV, figure 2 a). We found a distinct size interval during which the number of replisomes increased (see paper IV, figure 2 a lower red solid). The initiation volume obtained for slow growth was determined to be  $1.8 \mu\text{m}^3$ , which did not match that obtained during fast growth.

### The origin during the cell cycle

As the replisome trajectories were overlapping we studied the *oriC* region under conditions of fast growth. To do so, we introduced the gene coding for

a fluorescent reporter construct, Mall-Venus, near the *bglG* locus on the chromosome. Mall is a native TF of *E.coli* which associates to an operator site, *malO*, near the promoter of the maltose operon [61]. To allow Mall to bind at *bglG*, we introduced *malO* sites at this locus. Detection could then be achieved using the same approach as when imaging specifically bound LacI-Venus occupying the lactose promoter. To ensure that we only would monitor specific binding to the *bglG* locus by Mall-Venus, we removed the native *mall* gene and its operator sites completely. We found that the *oriC* locus was retained in specific sites inside the cells and that the trajectories over the cell cycle would bifurcate over the same size interval as the replisome trajectories (see paper IV, figure 2 e). The average number of *oriC* loci per cell would increase dramatically over a short size interval (see paper IV, figure 2 e lower red solid). However, the predicted initiation volume from this data suggested that initiation was most likely to occur at  $3.45 \mu\text{m}^3$ .

### A time-delay in replication initiation

The initiation sizes for fast growth indicated by replisomes and *oriCs* were inconsistent. Further, the production of *oriCs* was expected to follow initiation of replication rather than precede it. These apparent inconsistencies could however, be explained by a time-delay separating the replication of the arms of the chromosome [62]. A proposed model of chromosome management during fast growth is shown in *Figure 17*. According to this model, newly divided cells have four *oriCs* (white circles) and six replisomes (red circles) clustered into three intracellular replication sites (see *Figure 17* top). Further into the cell cycle the two replisomes residing in the inner most replication site terminate their corresponding round of replication. At this size, the 4 *oriCs* and remaining 4 replisomes reside in two sites which are symmetrically spaced around the mid-section. Replication initiates at the volume suggested by the *oriCs*, *i.e.* at  $3.45 \mu\text{m}^3$  which is indicated in *Figure 17* as the upper dashed white line. This is also the size at which the sites bifurcate. This results in the placing of 8 *oriCs* in four sites. Four new replisomes assemble to replicate the arms of the chromosome which are not delayed, and these are placed in four sites. The four replisomes which correspond to the previous rounds of replication are placed in the two sites closest to the mid-section. Therefore, the outer paths contain one replisome each and the inner sites contain three replisomes. As the cell grows to  $4.0 \mu\text{m}^3$  the time-delayed chromosome arms are replicated (*Figure 17* white dashed line) and four new replisomes are added, one to each site. Therefore the relative increase on the outer sites is 50%, while it is 33% on the inner sites. The time-delay has been reported to be 7 minutes under conditions of slow growth. The size interval  $3.45\text{-}4.0 \mu\text{m}^3$  represents 6.9 min according to the average growth rate during fast growth.

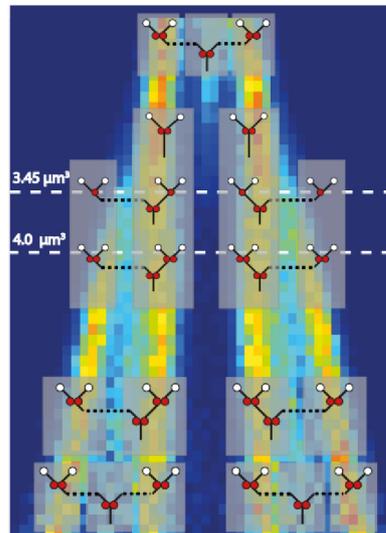


Figure 17. The chromosome management model. Background: color intensity represents replisome density along axial coordinates (horizontal) over the size cycle (vertical). Dashed lines: upper represents initiation of replication on the early starting chromosome arm, lower indicates initiation on the late starting chromosome arm. Chromosomes: white circles indicate *oriCs*, red circles indicate replisomes.

### Predictions of the chromosome management model

Under conditions of slow growth, when organizing data according to size a 50 % increase in the number of replisomes per site is expected ~7 minutes after initiation under the model. This time represents a size interval of  $0.12 \mu\text{m}^3$  at the average growth rate under conditions of slow growth. Therefore it is difficult to test this model using the replisome data. The *bglG* locus, is found on the arm which replicates first according to [62]. An identical Mall-Venus reporter construct placed at the same distance from the *oriC* on the time-delayed arm, should produce a 50% increase in *loci* per cell at  $4.0 \mu\text{m}^3$  if a time-delay exists.

### Regulation of replication initiation

In 1968 Donachie proposed that *E. coli* will initiate replication at a constant cell volume per *oriC* [25]. During slow growth, cells were expected to have 2 *oriCs* at the initiation volume,  $1.8 \mu\text{m}^3$ . Therefore the initiation volume per *oriC*, is  $0.9 \mu\text{m}^3$ . In the case of fast growth, cells had 4 *oriCs* at the initiation volume,  $3.45 \mu\text{m}^3$ . The initiation volume per *oriC* was therefore  $0.86 \mu\text{m}^3$

for fast growth. Our observations were therefore consistent with Donachie's prediction.

### Regulation of division timing

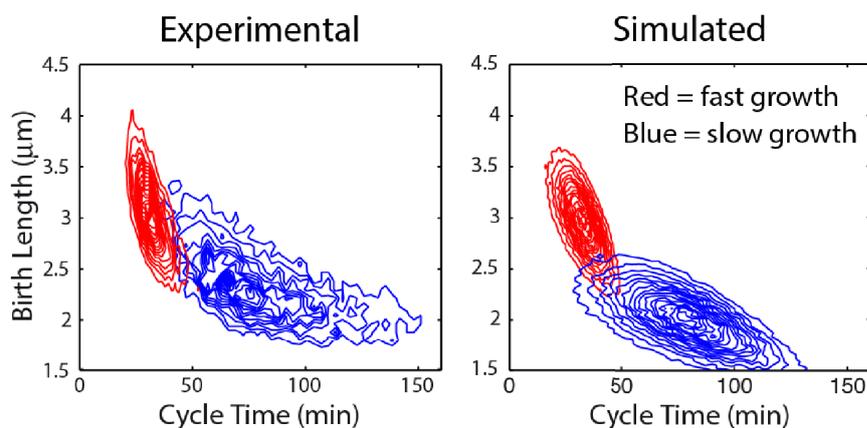
The average *E. coli* has to expand its envelope to twice the size at birth, to initiate and terminate a round of replication, to segregate chromosomes and to divide the cell (see introduction under *Requirements to conclude a cell cycle*). To accomplish the task of doubling the cell during exponential growth, it is necessary to regulate birth sizes and the cycle time. This can be achieved by having a regulatory mechanism which is either a timer, which divides the cell according to the time which has elapsed since division, or a sizer, which divides the cell according to size. It was recently proposed that *E. coli* growing under conditions of fast growth would divide according to a sizer and a timer [63]. Here the authors investigate the division rate of *E. coli* and devise a phenomenological stochastic model in which the time spent in the cell cycle until achieving a given size increases the propensity of the cell to divide.

Before a cell is ready to divide it must have completed a round of replication. This round is for both fast growth and for slow growth as studied in our experiments, not initiated within the same cell cycle, but is instead inherited from an ancestor cell. We knew when in the cell cycle the average round initiated. Further, by assuming that termination of replication occurred in the replication site at the center of the cell (see *Figure 17* top), we could determine the length at which it occurred. This allowed us to compute average replication time, or the C-period for both growth conditions which was found to be 56 min for fast growth and 63 min for slow growth (see paper IV, supplemental under *Estimates of the C and D periods*). Also, the time from termination to division, or the D-period, could be computed in this way and was found to be 25 min for fast growth and 50 min for slow growth.

Since these were averages, the joint distribution between the initiation length and the corresponding division time were not known to us. The least assuming model was that these were perfectly correlated, *i.e.* that division was executed after time-delay of C+D min after initiation, as in the Cooper-Helmstetter model. This control system was a sizer followed by a timer. We simulated cell cycles according to the cycle regulator using the experimentally determined values and the average growth rate for each condition (see paper IV, supplemental under *Cell cycle simulations*) and stored sampled values of birth length and cycle times for each cycle. The birth lengths and cycle times observed during our experiments were jointly distributed according to an uncertainty relation (see *Figure 18* left) in which variation is conserved between these properties. The weight of uncertainties was shifted according to the growth conditions. During conditions of fast growth, the variation in birth length was greater than under conditions of slow growth.

During conditions of slow growth the variation in cycle time was greater than under conditions of fast growth.

We found that this model captured the distributions of observed birth lengths and cycle times and their co-relation relatively well (see *Figure 18*). This would entail that variation in the initiation volume between individuals propagates to cause variations in birth size and cycle time according to the average growth rate.



*Figure 18.* Joint distribution of size at birth and cycle time. Density is indicated by level curves. Left: experimental values. Right: simulated values.

### Predictions under the cycle regulator model

A prediction under this model was that siblings should share a common cycle time. This is due to a common initiation event during a previous generation. As described earlier, the cycle times of mothers and daughters are not correlated (see *Figure 12* top). The cycle times of siblings are more correlated for both slow and fast growth (see *Figure 12* lower). An increased correlation between cycle times of siblings is expected for all growth conditions under which active replication is inherited. For growth conditions in which the same round of replication is initiated and terminated within the average cell cycle, *i.e.* at very slow growth, the correlation between siblings is expected to be extinguished.

### Variance in cell cycle regulation

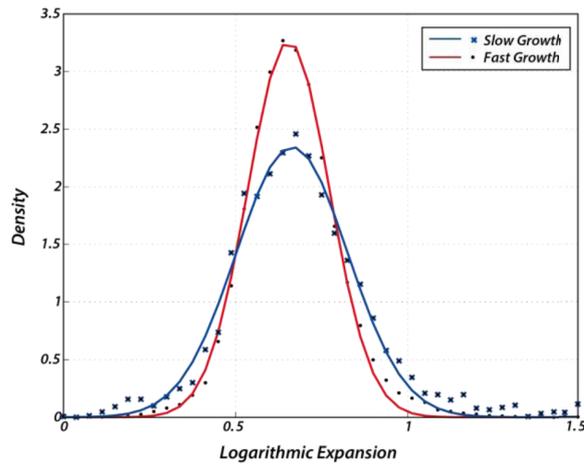
The time average of the relative expansion of cell size at the time of division is required to be 2 to maintain a constant cell size (see introduction under *Requirements to conclude a cell cycle*).

$$\langle \Lambda(\tau_c) \rangle_t = \langle \exp(\mu \cdot \tau_c) \rangle_t = 2$$

The expansion under conditions of exponential growth is a function of the logarithmic expansion,  $\lambda$ .

$$\lambda = \mu \cdot \tau_c$$

We found that the average values were  $\lambda=0.69 \pm 0.21, 0.0032$  (SD, SEM) for slow growth and  $\lambda=0.67 \pm 0.14, 0.0011$  (SED, SEM) for fast growth. These were close to the ideal value, *i.e.*  $\lambda=\ln 2 \approx 0.69$  (see introduction under *Requirements for concluding the cell cycle*). Under both conditions  $\lambda$  appeared to be normal distributed (see *Figure 19*) and the spread of the distributions could be obtained as the best fit of the normal density function. We found that  $\sigma_\lambda=0.16 \pm 0.0078$  (95% DSCI) for slow growth and  $\sigma_\lambda=0.12 \pm 0.0028$  (95% DSCI), which represented 23% and 17% of the ideal logarithmic expansion respectively.



*Figure 19.* The distribution of the logarithmic expansion. Observed values are shown as black dots (fast growth) and crosses (slow growth).

We showed that the uncertainty relation between size at birth and cycle time (discussed under *Regulation of division timing*) could be accounted for without considering population variation in growth rate (see *Figure 18* left). According to the uncertainty relation, the variance in cycle time was dependent on the growth conditions. The variance in the logarithmic expansion is dependent on the variance in the cycle time, the variance in growth rate and covariance between both. We had established that growth rate and growth rate expected from the cycle time were less correlated for fast growth as compared to slow growth (see under *Interdependencies of cycle variables*

and under *Cycle time and growth rate*). Further, the variance in growth rate under conditions of fast growth was 40% greater than the variance under conditions of slow growth. The variance in the logarithmic expansion was very similar between the two conditions. Under conditions of fast growth this was due to a greater population variance in growth rate as compared to slow growth.

It appeared that *E. coli* had an additional uncertainty relation, between the cycle time and the growth rate. The exponential growth model assumes that individuals maintain their growth rate throughout the duration of their cell cycle. The growth rate of a mother cell and daughter cell are moderately correlated for slow and for fast growth alike (see paper IV, figure S10, row 4, column 3-4). The growth rates of sibling pairs are highly correlated during slow growth and even more so during fast growth (see under *Interdependencies of cycle variables* and paper IV, figure S10, row 2, column 3-4 and).

If the growth rate of the mother alters over the cycle, we would expect that the adherence of the data to the exponential model would be worse in the cases where mother and daughter have greater differences in their growth rates. We do not observe such a dependency in our data for fast growth. It therefore appears that the determination of the growth rate of an individual is partly determined by the growth rate of the mother and partly determined by some other factor.

## Conclusions and future outlooks

### Technology development for the study of bacterial cells

As the microscopic world lies beyond our natural senses, insights of it have only been afforded at the rate with which technology has advanced. Here, the integration of microfluidics and single molecule fluorescence microscopy has afforded insights of the regulation of the bacterial cell cycle and how genes are regulated in living cells.

Integrating methods adds complexity and the development of a composite process requires a co-development of many factors. Optimization can easily lead to a combinatorial explosion of conditions to test and as a result, complex processes are rarely exhaustively optimized. After establishing accuracy of the process, the stability or reliability is often prioritized. The reliability of a composite process is often only as good as its weakest link or the least reliable factor. In order to have a reliable composite process all sub-processes must be very reliable.

The integrated method we developed allows for an arbitrary number of observations to be acquired from one experiment and is automated from the start of acquisition. The rate of acquiring complete cell cycles is currently set by the number of positions which can be monitored in parallel. If all positions in the chip were monitored, data containing  $\sim 10^4$  cell cycles would be acquired per hour during conditions of fast growth.

### Insights afforded by the method

Early dividing cells are likely over-represented in data from time-lapsed experiments. This bias will affect all factors which are dependent on the cycle time. To minimize this bias, experiments should be conducted for long durations and cell tracking should be as accurate as possible.

Under our experimental conditions, the timing of gene expression events is difficult to determine using fluorophores such as Venus due to a long maturation time. This can be a result of low oxygen levels in the microfluidic

chips during our experiments. The performance of Venus under other experimental conditions are difficult to predict, however it appears that it is crucial to be able to test the maturation of fluorophores under the conditions at which they are used.

Information on binding states, such as specifically bound TFs or active replisomes, appears to be more reliable, if a scheme of synchronization can be achieved which allows for averaging over many observations without masking underlying mixed distributions. In this way it was possible to test models of gene regulation in living cells. Our tests revealed that regulation does not occur at equilibrium. This illustrates that the study of molecules in living cells is important to understand their behavior.

Further, the chromosome was found to be maintained in a size dependent manner in *E. coli*. This is due to a size dependent mechanism for initiation of replication, where cells will initiate according to a constant volume per number of origins. Initiation is the main regulatory event of the cell cycle and variations in establishing its timing determine the distributions of cell sizes and cycle times. A consequence of this regulation is an uncertainty relation between size at birth and cycle time, in which variance for both entities is conserved and their relative magnitudes are set by the growth conditions.

## Future developments

One of the major limitations to studying individual cells using fluorescence microscopy is the performance of fluorophores. A great variety of fluorescent protein variants are available. However, fast maturing variants are rare. As we have seen the conditions under which the fluorophores are used may influence their properties. Using microfluidics it is possible to measure the maturation time *in situ*. This approach could potentially be used to develop new fast maturing fluorophores.

For applications such as single particle tracking, high photo-stability is essential to obtain long trajectories. Conversely, when counting individual fluorescent molecules to determine when they were synthesized, fluorophores which are less photo-stable are required. During our experiments cells have been more sensitive to laser exposure than expected and great care had to be exercised not to injure them. Bleaching the majority of Tsr-Venus molecules present in a population with a frequency of  $1/3 \text{ min}^{-1}$  under the conditions studied in paper II caused cells to grow slower, to filament, and rupture.

Fluorescent proteins are often the same size as the proteins they label and many attempts to label proteins impair their native function. Further, fluo-

rescent proteins have a propensity to form aggregates which can be misinterpreted as intracellular localization patterns of the native protein [64].

The tools of molecular biology are currently limited to studying the cases which conform to the requirements of those tools. The development of improved fluorophores appears to be the most important undertaking in order to widen the range of molecules which can be studied using fluorescence microscopy.

## Uncertainty in the bacterial cell cycle

*E. coli* regulates its cell cycle in a manner which will shift variance between the sizes at birth and the cycle times over different growth conditions. There appears to be at least one more uncertainty relation in this regulation, between the growth rate and the cycle time. For the two conditions we have studied here it appears that the variance in the product of these two entities, the logarithmic expansion is conserved to some degree. In the case of fast growth, this is due to an increase in the variance in growth rate.

The growth rate is often attributed to the concentration of constitutive cell parts which can generate more parts, *e.g.* ribosomes or RNA polymerases. The growth rate appears to be consistent over the duration of the cell cycle. At division we assume that a sibling pair will retain the constitutive parts which were present in the mother. However, growth rate is only moderately correlated between mothers and daughters (see under *Variance in cell cycle regulation*). Between siblings the growth rate is instead highly correlated. It is difficult to explain these findings solely according to inheritance of parts.

If an additional factor contributes to the determination of growth rate in a manner which increases variance in growth rate under conditions of fast growth, it could mean that the population is not growing at a maximum rate. If one bacterium in such a population should mutate to eliminate this factor and consequently grow faster than its neighbors, it is expected that this cell and its descendants would outcompete its neighbors and ultimately establish this mode of determining growth rate as the norm for future generations. For this not to occur, an evolutionary penalty for taking this path is expected.

The logarithmic expansion represents a concise description of the regulation of the bacterial cell cycle. We can therefore provide an answer to the question raised in the introduction. How precise is cyclic life? Well, it's 20% as measured by the coefficient of variation. The fascinating aspect of this finding is not the value itself, but that it is relatively conserved between growth conditions. The mechanisms involved are not known, but are related to the establishment of growth rate. Conservation in variance between size at birth and cycle time is a consequence of regulating the cell cycle to maintain constant cell sizes over-time by initiating replication according to size.

Whether variance in growth rate is a resulting consequence of regulation or a factor contributing to it remains to be discovered.

# Appendix: Measurement and determination of the maturation time of Venus in live bacterial cells growing in a microfluidic chip

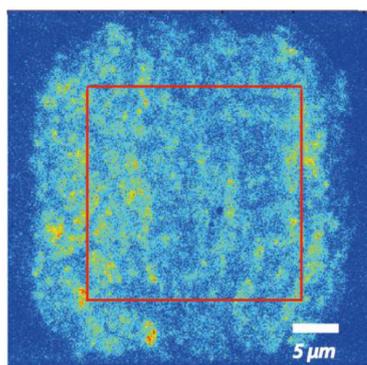
## Introduction

The reported maturation time of the yellow fluorescent protein variant, *Venus*, is 7 minutes [41, 58] and for that reason it is regarded as a fast maturing reporter suitable for studies of gene expression in living cells [41]. To measure the maturation time under the conditions present during a microfluidic time-lapsed fluorescence microscopy experiment, cells modified to express a reporter construct, LacY-Venus, from the native *lacY* gene [54] were subjected to chemical induction pulses of high IPTG concentrations for various durations. The microfluidic chip allows for continuous exponential growth and rapid induction of the lac promoter (~2 s) [65]. The response was measured as the fluorescence density in micro-colonies during and after induction. We found that under the conditions of fast growth in a microfluidic device, the average maturation time for Venus was 54 minutes, which constitutes roughly twice the duration of the cell cycle.

## Materials and methods

Preparations of the microfluidic device and microscopy were performed according to [65]. *E. coli* cells of strain SX700 [54] were cultivated in a microfluidic chip in M9 minimal medium, supplemented with 0.4% glucose and 1xRPMI 1640 amino acid (sigma-aldrich) and a surfactant Pluronic F108, 0.85 gL<sup>-1</sup>. The temperature of the chip was maintained at 37°C for the duration of the experiment. Both media reservoirs contained this as a base. One reservoir also contained 1 mM IPTG and trace amounts of inert polystyrene beads (2 µm diameter) for visualization of flows in the microfluidic device. Acquisition of images was synchronized with induction pulses executed by linear actuators according to scheduling in the acquisition software RITAcquire2 [65]. The duration of induction pulses in the experiment were

[0, 5, 10, 20, 45, 60, 120, 600, 1200] seconds. For each induction pulse micro-colonies were imaged only once to minimize bleaching. Each of the 48 micro-colonies was regarded as an individual time point after the start of induction and was imaged with 2 minutes separation. After the last micro-colony was imaged (~100 minutes after induction), the cells were allowed to recuperate for 4.7 hours before repeating the cycle for another induction duration. During this time the LacY-Venus molecules formed in response to induction were depleted by growth. The responses to nine different conditions were measured in one session lasting nearly 60 hours, during which the experimental setup executed all commands autonomously. Further, to ensure that no interfering time-dependent process would be misinterpreted as an induction response, the order of the induction durations was randomized. A centered sub-region of the fluorescence images for each micro-colony corresponding to  $23 \times 23 \mu\text{m}^2$  was found to yield similar cell density for all colonies (see *Figure 20*). To obtain the fluorescence response for each time point after induction, the fluorescence counts were averaged over and divided with the surface area over the sub-region.



*Figure 20.* A fluorescence micrograph of a colony of cells residing in a microfluidic chip. Red square indicates the region which is averaged in analysis.

## Results

The fluorescence response to induction by IPTG can be seen in *Figure 21*. The fluorescence response for all micro-colonies under un-induced conditions was used as the fluorescence response background,  $x_b$ , and yielded  $0.934 \pm 0.031, 0.0044$  (SD, SEM) counts  $\cdot \mu\text{m}^{-2}$ . The recuperation time appeared sufficiently long as the fluorescence response for the zeroth time point for each induction duration was similar to the un-induced  $1.05 \pm$

0.125, 0.10 (SD, SEM) counts  $\cdot \mu\text{m}^{-2}$ . The fluorescence response to induction was noisy but showed a dramatic increase for longer induction durations. The response increased to a maximum value, then decreased over time. This trend could be seen for inductions as brief as 20 seconds although the relative fluctuations between time points were large. The response to induction was modelled assuming that a fluorescence mass,  $x_{tot}$ , was synthesized during the induction pulse. The immature fluorescence mass,  $x_0$ , was assumed to be non-fluorescent. The maturation time was regarded as an exponentially distributed random variable with average maturation time,  $\tau_{mat}$ . The maturation rate,  $k_{mat}$ , was therefore

$$k_{mat} = \tau_{mat}^{-1}$$

As molecules mature, the mature fluorescence mass,  $x_{mat}$ , was formed. As the cells were growing exponentially [65, 66], both the immature and mature fraction of all molecules were diluted by growth at rate,  $\mu$ . This could be modelled as a system of linear ordinary differential equations as

$$\frac{d}{dt} \begin{bmatrix} x_{imat} \\ x_{mat} \end{bmatrix} = \begin{bmatrix} -(k_{mat} + \mu) & 0 \\ k_{mat} & -\mu \end{bmatrix} \begin{bmatrix} x_{imat} \\ x_{mat} \end{bmatrix}$$

Which could be solved analytically so that

$$x_{mat}(t) = x_{tot} \cdot (\exp(-\mu t) - \exp(-(k_{mat} + \mu)t))$$

Further we assumed that the total fluorescence mass was proportional to the induction duration,  $\Delta t_i$ , as

$$x_{tot} = k_{synth} \cdot \Delta t_i$$

where  $k_{synth}$  was the fluorescence synthesis rate for the lactose promoter while fully induced. The background,  $x_{back}$ , was assumed to be invariant over time and space and was added to the expression for the fluorescence response to induction as

$$x_{mat}(t) = k_{synth} \cdot \Delta t_i \cdot (\exp(-\mu t) - \exp(-(k_{mat} + \mu)t)) + x_{back}$$

The average growth rate  $\mu$  was obtained by segmentation and tracking according to [65, 66] as  $0.024 \text{ min}^{-1}$  and the fluorescence background could be measured for un-induced cells. Two unknown parameters,  $k_{synth}$  and  $k_{mat}$  remained. However, we found that this expression was inadequately complex to capture the observed induction response.

We observe that there was an initial period during which the fluorescence response was lagging before a more dramatic increase in the response en-

sues. The model assumed an immediate synthesis of immature protein molecules. This lag may be an indication that this was not so. However, the model does appear to capture the features of the fluorescence response with the addition of a time-delay,  $\tau_d$ . This transforms the model for the response to

$$x_{mat}(t) = H(t - \tau_{del}) \cdot k_{synth} \cdot \Delta t_i \cdot \left( \exp(-\mu \cdot (t - \tau_d)) - \exp(-(k_{mat} + \mu) \cdot (t - \tau_d)) \right) + x_b$$

where H is the Heaviside function. The parameters were determined as the best fit using the Levenberg-Marquardt algorithm as implemented in MATLAB 2012b. The fluorescence synthesis rate for LacY-Venus is  $k_{synth} = 0.058 \pm 0.011 \text{ min}^{-1}$ , the maturation rate of Venus was determined to be  $k_{mat} = 0.0184 \pm 0.005 \text{ min}^{-1}$  and the expression delay found to be  $\tau_d = 11.15 \pm 1.18 \text{ min}$ . The uncertainty measures stated are all 95% double sided confidence intervals for each parameter.

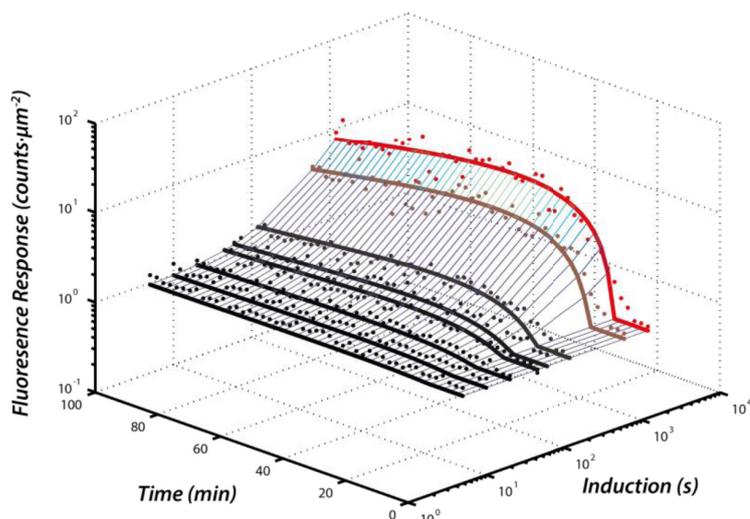


Figure 21. Fluorescence response to induction. The fluorescence response is averaged for a uniformly dense sub region of each micro-colony and divided with the surface area of the region. The cells residing in the chip are exposed to induction pulses of durations according to “induction time (s)” and measurements (dots) are recorded at one micro-colony per time point according to “time (min)”. The induction durations tested were  $\Delta t_i = 0, 5, 10, 20, 45, 60, 120, 600$  and 1200 seconds. The best fit of the model presented under result is shown as solid lines.

## Discussion

We find that the induction response of the *lac operon* is detectable for induction durations of 20 seconds or longer. Further, we find that measurements are repeatable over several days and that the qualitative features of the re-

sponse are not time-dependent. We find that the description of the maturation process as a first order reaction captures the observations well only if a time-delay between the start of induction and start of maturation is present. Under this model the corresponding maturation time of Venus is 54 minutes, which is roughly twice the generation time of the cells. The fluorescence response appears to have a substantial lag which appears to be invariant to the duration of induction, suggesting that even for brief inductions a response is generated ~10 minutes later.

From previous tests of the induction capacities of the microfluidic chip we have concluded that the LacI-Venus binding to its native operator site is induced to dissociate within 2 seconds from executing the command to exchange the medium [65]. Therefore we exclude the possibility that the lag is due to a delayed switch of medium in the chip. The process of protein synthesis contains several phases which could contribute to a lag before any LacY-Venus molecules are produced. The *lacY* gene is preceded in the *lac* operon by the *lacZ* gene. However, the *lacZ* gene is reported to synthesize functional  $\beta$ -galactosidase molecules within 75 s after saturating induction by IPTG [67] and the additional synthesis time of LacY-Venus is not expected from the length of the gene and the translation rate to account for the residual duration of the lag.

An alternative explanation for the kinetics of the response could be that the reporter mRNA is stable for long periods. The expected life-time of the mRNA would then have to rival the cell cycle, which is considerably longer than that of the average mRNA, 3-8 min [68], and that reported for *lacZYA*, 3 min [41]. For this reason we suggest that induction response is more likely a consequence of fluorophore maturation. The implication of the maturation time of Venus under the conditions studied here is that its use as a molecular reporter is likely to obscure information of the time when corresponding molecules were synthesized. The effect is expected to be greater for fast growing cells although it is not known if the maturation process is different under different growth conditions. The translatability of these results to other fluorescent reporters and other conditions under which they can be used is not known. However, they illustrate the importance of *in situ* characterization of the maturation properties of protein fluorophores to understand the results they generate.

## Swedish Summary

Bakterieceller fortplantar sig genom att växa och dela sig. För att åstadkomma detta måste den genomsnittliga cellen dubblera sin massa under cellcykeln. Bland annat behöver bakteriens arvsmassa, eller kromosom, dubblas. Detta sker genom DNA-replikation. Tarmbakterien *Escherichia coli* kan under snabb tillväxt ha en kortare cykeltid än tiden det tar att replikera kromosomen. Detta görs genom att flera rundor av replikation bibehålls samtidigt. Som en konsekvens ärver celler då replikationsprocesser som startades, eller initierades, under tidigare cellcykler. Kromosomen innehåller gener som kan uttryckas för att tillgodose cellens behov av enzymer, strukturella delar och reglermolekyler.

För att studera cellcykelns inverkan på gen-reglering i levande bakterieceller utvecklades en metod som kombinerade mikrofluidik, fluorescensmikroskopi och automatiserad bildanalys. Mikrofluidiken möjliggör att behålla celler i ett tillstånd av konstant tillväxt samtidigt som deras rörelser begränsas tillräckligt för att upprepade gånger avbilda dessa i ett mikroskop. Känsligheten i modern fluorescensmikroskopi är tillräckligt god för att möjliggöra detektion och lokalisering av enskilda fluorescerande proteinmolekyler inuti levande bakterieceller.

Geners uttryck regleras ofta av transkriptionsfaktorer. Transkriptionsfaktorer är proteiner som binder till DNA. Ofta binder dessa särskilt hårt till en eller några specifika DNA-sekvenser, s.k. operatorsekvenser. Gener är ofta grupperade på kromosomen för att samuttryckas, i s.k. operon. Laktosoperonet i *E. coli* ger bakterien möjligheten att metabolisera sockret laktos, och operonets uttryck regleras av transkriptionsfaktorn LacI. Då LacI binder till sin operatorsekvens, undertrycks uttrycket av operonet. När LacI släpper denna möjliggörs uttryck. Genom att genetiskt modifiera bakterieceller för att uttrycka transkriptionsfaktorer som är sammansvetsade med fluorescerande proteiner, kan enskilda bundna transkriptionsfaktorer detekteras.

Mikrofluidiken möjliggör ett mycket snabbt byte av den kemiska miljön som bakterierna lever i. Detta kunde utnyttjas så att söktiden, då en molekyl söker sin operatorsekvens, och bindingstiden hos en fluorescerande variant av LacI kunde mätas i levande celler. Resultaten kunde inte förlikas med en modell för genreglering som förutsätter att de molekylära reaktionerna är i jämvikt. Cellens tillstånd är dynamiskt och i detta fall var det nödvändigt att

studera molekylerna i levande celler för att dra slutsatser om deras beteende och effekterna på genens uttryck.

Fluorescenta proteiner har en mognadstid, under vilken molekylerna inte kan detekteras. Det snabba bytet av den kemiska miljön i mikrofluidiken möjliggjorde att denna mognadstid kunde mätas för ett fluorescent protein, Venus. Resultatet visade att Venus mognade avsevärt långsammare än förväntat, under ungefär dubbelt så lång tid som en genomsnittlig cellcykel. Mognadsprocessen är dessutom stokastisk och noggrannheten med vilken en fluorescent molekyls tillverkningstid kan bestämmas påverkas drastiskt av långa mognadstider.

Studier av bindingstillstånd, så som bundna LacI molekyler, påverkas mindre av lång mognadstid. Studier av bindingen av LacI-Venus till operatorsekvensen under cellcykeln visade att kromosomens delar enklast lokaliseras genom att jämföra observationer av celler som har samma storlek. Vidare studier av *E. coli*s replikationsmaskineri avslöjade att initieringen av replikationen sker i enlighet med cellers storlek, snarare än tiden från delning. Replikation startar vid origin-regionen på kromosomen. Studier av denna visade att replikation initieras vid en viss cellvolym per origin, oberoende av levnadsbetingelserna.

För att avsluta en cellcykel behöver den genomsnittliga cellen initiera och terminera en runda av replikation, segregera sin arvs massa och dela sig. En enkel modell av regleringen av detta skeende, där tidpunkten för cellens delning bestäms som en tidsförskjutning från det att cellen initierade replikationen, visade sig återskapade de observerade värdena av storlek vid födsel och cykeltid. En konsekvens av regleringen är att dessa observabler ingår i en osäkerhetsrelation, där variationen i födslolängd är stor under snabb tillväxt jämfört med långsam tillväxt, och variationen i cykeltiderna är stor under långsam tillväxt jämfört med snabb tillväxt.

## Acknowledgements

**Johan Elf:** Thank you for making me a scientist. I know it wasn't easy.

**David Fange:** You carried me the last bit of the way. Thank you.

**Petter Hammar:** I miss you making fun of me and everything I say. It's strange really, but there it is.

**Prune Leroy:** Dear DJ SlivoviSka, I can't imagine life without you. Thanks.

**Anel Mahmutovic:** You are the smartest person I know. Also, you are really kind to me. It's an honor to have you as a friend.

**Gustaf Ullman:** You were a mentor to me and I'll never forget the things you taught me.

**Erik Marklund:** You are my favorite hipster scientist. Everything you do is fabulous.

**Brian English:** Dude, I miss you. Thanks for the early days.

**Vasili Hauryliuk:** Thanks for forcing me to climb a boat once. It was in the spirit of adventure.

**Gemma Atkinson:** You are one of the nicest, kindest people I've ever met.

**Stoyan Tankov:** Thanks for some fun times. I hope to see you again.

**Arash Sanamrad:** You were always very kind to me and patient when I stole your supplies always.

**Irmeli Barkefors:** You are the wisest person I know. I appreciate it and everything you've done for me.

**Fredrik Persson:** You are the most skilled person I know and it's been fun working with you. I miss it.

**Özden Baltekin:** Thank you. I'm sorry I yelled at you that time.

**Ebba Gregorsson-Lundius:** Thank you for proof-reading the thesis.

**Michael J. Lawson:** The loudness. Awesome. Thanks.

**Martin Lindén:** Thanks for all the advice. Sorry if I made too much noise in the office.

**Labmates:** Thank you so much and sorry for being so stressed out lately.

**Diarmaid Hughes:** Thanks for the discussions early in my graduate studies. I miss them.

**Santanu Dasgupta:** You taught me so much about *E.coli* and replication. It was always thrilling.

**Måns Ehrenberg:** You continue to surprise me. In a good way, I think. Our discussions are really fun.

**Otto Berg:** It's always really fun to talk to you about scientific problems. Thank you.

**Ivan Razinkov:** You're helpful, enthusiastic, idealistic and inspiring. Thanks.

**Charles Moolman:** Thanks for the advice. You were right.

**Octavio Palomino-Mondragon:** I remember our discussion. Thanks, it was really helpful.

**Zhigang Wu:** You've always helped me when I needed it. I really appreciate it.

**Jeff Hasty:** Thanks for allowing me to join your group for a spell. It was fabulous.

**Klas Hjort:** Your expertise has helped me greatly. Thank you.

**Peter Graumann:** Thank you for being my opponent.

**Members of Sputnik Kollektiv:** brothers and sisters in arms and legs I salute you. Thank you.

**Marek Skupinski:** Thank you for the cover art. You are brilliant.

**Charles A. Tannen:** You are my nemesis. Thanks.

**Family:** I love you.

**Harriet Mellenius:** I'm the luckiest guy ever. Thank you.

**Branka Walldén-Mellenius:** Pappa älskar dig.

## References

1. Purcell, E.M., *Life at low Reynolds number*. American Journal of Physics, 1977. **45**.
2. Berg, H.C., *Random Walks in Biology*. 1993.
3. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid*. Nature, 1953. **171**(4356): p. 737-8.
4. Crick, F., *Central dogma of molecular biology*. Nature, 1970. **227**(5258): p. 561-3.
5. Jacob, F. and J. Monod, *Genetic regulatory mechanisms in the synthesis of proteins*. J Mol Biol, 1961. **3**: p. 318-56.
6. Savageau, M.A., *Design of molecular control mechanisms and the demand for gene expression*. Proc Natl Acad Sci U S A, 1977. **74**(12): p. 5647-51.
7. Dekel, E. and U. Alon, *Optimality and evolutionary tuning of the expression level of a protein*. Nature, 2005. **436**(7050): p. 588-92.
8. Kuhlman, T., et al., *Combinatorial transcriptional control of the lactose operon of Escherichia coli*. Proc Natl Acad Sci U S A, 2007. **104**(14): p. 6043-8.
9. Lewis, M., et al., *Crystal structure of the lactose operon repressor and its complexes with DNA and inducer*. Science, 1996. **271**(5253): p. 1247-54.
10. Oehler, S., et al., *Quality and position of the three lac operators of E. coli define efficiency of repression*. EMBO J, 1994. **13**(14): p. 3348-55.
11. Vilar, J.M. and S. Leibler, *DNA looping and physical constraints on transcription regulation*. J Mol Biol, 2003. **331**(5): p. 981-9.
12. Gilbert, W. and B. Muller-Hill, *The lac operator is DNA*. Proc Natl Acad Sci U S A, 1967. **58**(6): p. 2415-21.
13. Hammar, P., et al., *The lac repressor displays facilitated diffusion in living cells*. Science, 2012. **336**(6088): p. 1595-8.
14. Bintu, L., et al., *Transcriptional regulation by the numbers: applications*. Curr Opin Genet Dev, 2005. **15**(2): p. 125-35.
15. Eckburg, P.B., et al., *Diversity of the human intestinal microbial flora*. Science, 2005. **308**(5728): p. 1635-8.
16. Blattner, F.R., et al., *The complete genome sequence of Escherichia coli K-12*. Science, 1997. **277**(5331): p. 1453-62.

17. Lesterlin, C., et al., *Roles for replichores and macrodomains in segregation of the Escherichia coli chromosome*. EMBO Rep, 2005. **6**(6): p. 557-62.
18. Sakabe, K. and R. Okazaki, *A unique property of the replicating region of chromosomal DNA*. Biochim Biophys Acta, 1966. **129**(3): p. 651-4.
19. Yao, N.Y. and M. O'Donnell, *Replisome structure and conformational dynamics underlie fork progression past obstacles*. Curr Opin Cell Biol, 2009. **21**(3): p. 336-43.
20. Cooper, S. and C.E. Helmstetter, *Chromosome replication and the division cycle of Escherichia coli B/r*. J Mol Biol, 1968. **31**(3): p. 519-40.
21. Reshes, G., et al., *Timing the start of division in E. coli: a single-cell study*. Phys Biol, 2008. **5**(4): p. 046001.
22. Bremer, H. and G. Churchward, *An examination of the Cooper-Helmstetter theory of DNA replication in bacteria and its underlying assumptions*. J Theor Biol, 1977. **69**(4): p. 645-54.
23. Skarstad, K., E. Boye, and H.B. Steen, *Timing of initiation of chromosome replication in individual Escherichia coli cells*. EMBO J, 1986. **5**(7): p. 1711-7.
24. Skarstad, K. and A. Lobner-Olesen, *Stable co-existence of separate replicons in Escherichia coli is dependent on once-per-cell-cycle initiation*. EMBO J, 2003. **22**(1): p. 140-50.
25. Donachie, W.D., *Relationship between cell size and time of initiation of DNA replication*. Nature, 1968. **219**(5158): p. 1077-9.
26. Bates, D. and N. Kleckner, *Chromosome and replisome dynamics in E. coli: loss of sister cohesion triggers global chromosome movement and mediates chromosome segregation*. Cell, 2005. **121**(6): p. 899-911.
27. Helmstetter, C.E. and D.J. Cummings, *Bacterial Synchronization by Selection of Cells at Division*. Proc Natl Acad Sci U S A, 1963. **50**: p. 767-74.
28. Bates, D., et al., *The Escherichia coli baby cell column: a novel cell synchronization method provides new insight into the bacterial cell cycle*. Mol Microbiol, 2005. **57**(2): p. 380-91.
29. Shimomura, O., F.H. Johnson, and Y. Saiga, *Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusa, Aequorea*. J Cell Comp Physiol, 1962. **59**: p. 223-39.
30. Heim, R. and R.Y. Tsien, *Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer*. Curr Biol, 1996. **6**(2): p. 178-82.
31. Chalfie, M., et al., *Green fluorescent protein as a marker for gene expression*. Science, 1994. **263**(5148): p. 802-5.
32. James, J., *Light microscopic techniques in biology and medicine*. 1976, The Hague: M. Nijhoff Medical Division. xi, 336 p.

33. Willig, K.I., et al., *Nanoscale resolution in GFP-based microscopy*. Nat Methods, 2006. **3**(9): p. 721-3.
34. Betzig, E., et al., *Imaging intracellular fluorescent proteins at nanometer resolution*. Science, 2006. **313**(5793): p. 1642-5.
35. Rust, M.J., M. Bates, and X. Zhuang, *Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)*. Nat Methods, 2006. **3**(10): p. 793-5.
36. Thompson, R.E., D.R. Larson, and W.W. Webb, *Precise nanometer localization analysis for individual fluorescent probes*. Biophys J, 2002. **82**(5): p. 2775-83.
37. Betzig, E., *Proposed method for molecular optical imaging*. Opt Lett, 1995. **20**(3): p. 237-9.
38. Dickson, R.M., et al., *On/off blinking and switching behaviour of single molecules of green fluorescent protein*. Nature, 1997. **388**(6640): p. 355-8.
39. Elf, J., G.W. Li, and X.S. Xie, *Probing transcription factor dynamics at the single-molecule level in a living cell*. Science, 2007. **316**(5828): p. 1191-4.
40. English, B.P., et al., *Single-molecule investigations of the stringent response machinery in living bacterial cells*. Proc Natl Acad Sci U S A, 2011. **108**(31): p. E365-73.
41. Yu, J., et al., *Probing gene expression in live cells, one protein molecule at a time*. Science, 2006. **311**(5767): p. 1600-3.
42. Colville, K., et al., *Effects of poly(L-lysine) substrates on attached Escherichia coli bacteria*. Langmuir, 2010. **26**(4): p. 2639-44.
43. Young, J.W., et al., *Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy*. Nat Protoc, 2012. **7**(1): p. 80-8.
44. A.Manz, N.G., H.M.Widmer,, *"Miniaturized total chemical analysis systems: a novel concept for chemical sensing"*. Sens. Actuators B1, , 1990: p. 244-248
45. Wang, P., et al., *Robust growth of Escherichia coli*. Curr Biol, 2010. **20**(12): p. 1099-103.
46. Edelstein, A., et al., *Computer control of microscopes using microManager*. Curr Protoc Mol Biol, 2010. **Chapter 14**: p. Unit14 20.
47. Stricker, J., et al., *A fast, robust and tunable synthetic gene oscillator*. Nature, 2008. **456**(7221): p. 516-9.
48. Mather, W., et al., *Streaming instability in growing cell populations*. Phys Rev Lett, 2010. **104**(20): p. 208101.
49. Loomis, W.F., Jr. and B. Magasanik, *Glucose-lactose diauxie in Escherichia coli*. J Bacteriol, 1967. **93**(4): p. 1397-401.
50. Sliusarenko, O., et al., *High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics*. Mol Microbiol, 2011. **80**(3): p. 612-27.
51. Horn, B.K.P.S., B.G, *Determining optical flow*. Artificial intelligence, 1981. **17**(185-203).

52. Soderstrom, B., et al., *Disassembly of the divisome in Escherichia coli: evidence that FtsZ dissociates before compartmentalization*. Mol Microbiol, 2014. **92**(1): p. 1-9.
53. Huh, D. and J. Paulsson, *Random partitioning of molecules at cell division*. Proc Natl Acad Sci U S A, 2011. **108**(36): p. 15004-9.
54. Choi, P.J., et al., *A stochastic single-molecule event triggers phenotype switching of a bacterial cell*. Science, 2008. **322**(5900): p. 442-6.
55. Chang, W.I., P. Barrera, and K.S. Matthews, *Identification and characterization of aspartate residues that play key roles in the allosteric regulation of a transcription factor: aspartate 274 is essential for inducer binding in lac repressor*. Biochemistry, 1994. **33**(12): p. 3607-16.
56. Johnson, C.M. and R.F. Schleif, *In vivo induction kinetics of the arabinose promoters in Escherichia coli*. J Bacteriol, 1995. **177**(12): p. 3438-42.
57. Kuhlman, T.E. and E.C. Cox, *Gene location and DNA density determine transcription factor distributions in Escherichia coli*. Mol Syst Biol, 2012. **8**: p. 610.
58. Nagai, T., et al., *A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications*. Nat Biotechnol, 2002. **20**(1): p. 87-90.
59. Reyes-Lamothe, R., D.J. Sherratt, and M.C. Leake, *Stoichiometry and architecture of active DNA replication machinery in Escherichia coli*. Science, 2010. **328**(5977): p. 498-501.
60. Hill, N.S., et al., *Cell size and the initiation of DNA replication in bacteria*. PLoS Genet, 2012. **8**(3): p. e1002549.
61. Reidl, J., et al., *Mall, a novel protein involved in regulation of the maltose system of Escherichia coli, is highly homologous to the repressor proteins GalR, CytR, and LacI*. J Bacteriol, 1989. **171**(9): p. 4888-99.
62. Joshi, M.C., et al., *Escherichia coli sister chromosome separation includes an abrupt global transition with concomitant release of late-splitting intersister snaps*. Proc Natl Acad Sci U S A, 2011. **108**(7): p. 2765-70.
63. Osella, M., E. Nugent, and M. Cosentino Lagomarsino, *Concerted control of Escherichia coli cell division*. Proc Natl Acad Sci U S A, 2014. **111**(9): p. 3431-5.
64. Landgraf, D., et al., *Segregation of molecules at cell division reveals native protein localization*. Nat Methods, 2012. **9**(5): p. 480-2.
65. Hammar, P., et al., *Direct measurement of transcription factor dissociation excludes a simple operator occupancy model for gene regulation*. Nat Genet, 2014. **46**(4): p. 405-8.
66. Ullman, G., et al., *High-throughput gene expression analysis at the level of single proteins using a microfluidic turbidostat and automated cell tracking*. Philos Trans R Soc Lond B Biol Sci, 2013. **368**(1611): p. 20120025.

67. Dalbow, D.G. and R. Young, *Synthesis time of beta-galactosidase in Escherichia coli B/r as a function of growth rate*. *Biochem J*, 1975. **150**(1): p. 13-20.
68. Bernstein, J.A., et al., *Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays*. *Proc Natl Acad Sci U S A*, 2002. **99**(15): p. 9697-702.

## Acta Universitatis Upsaliensis

*Digital Comprehensive Summaries of Uppsala Dissertations  
from the Faculty of Science and Technology 1204*

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology".)

Distribution: [publications.uu.se](http://publications.uu.se)  
urn:nbn:se:uu:diva-235546



ACTA  
UNIVERSITATIS  
UPSALIENSIS  
UPPSALA  
2014

Paper I 





## Studying transcriptional interactions in single cells at sufficient resolution

Mats Wallden and Johan Elf

Our ability to dissect and understand the principles of gene regulatory circuits is partly limited by the resolution of our experimental assays. In this brief review, we discuss aspects of gene expression in microbial organisms apparent only when increasing the experimental resolution from populations to single cells and sub-cellular structures, from snap-shots to high-speed time-lapse movies and from molecular ensembles to single molecules.

### Address

Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, Husargatan 3, SE-751 24, Uppsala, Sweden

Corresponding author: Elf, Johan ([johan.elf@icm.uu.se](mailto:johan.elf@icm.uu.se))

**Current Opinion in Biotechnology** 2011, **22**:81–86

This review comes from a themed issue on  
Analytical biotechnology  
Edited by Matthias Heinemann and Uwe Sauer

Available online 10th November 2010

0958-1669/\$ – see front matter  
© 2010 Elsevier Ltd. All rights reserved.

DOI [10.1016/j.copbio.2010.10.004](https://doi.org/10.1016/j.copbio.2010.10.004)

### Introduction

Gene regulation depends on a broad spectrum of intracellular signals encoded and processed at a variety of spatial and temporal scales. In order to unveil the regulatory principles of a specific system, the experimental assay needs to resolve the regulatory cues of the process under investigation. Obviously, to investigate stochastic differences in gene expression between cells, means of studying cells individually have to be adopted, and to understand dynamics of gene regulation, the process needs to be monitored at sufficient time resolution. Biomolecules may also change conformation, location, states of binding and modification, etc. and resolving these states may be necessary to determine their role for regulation. In this review, we present a few examples of regulatory interactions revealed only at fine experimental resolution. Further, we present the use of microfluidics as a means of meeting the greater demands of conducting high-resolution studies.

### Single-cell resolution and gene expression

Isogenic cells living under seemingly identical conditions may display considerable heterogeneity in gene expression

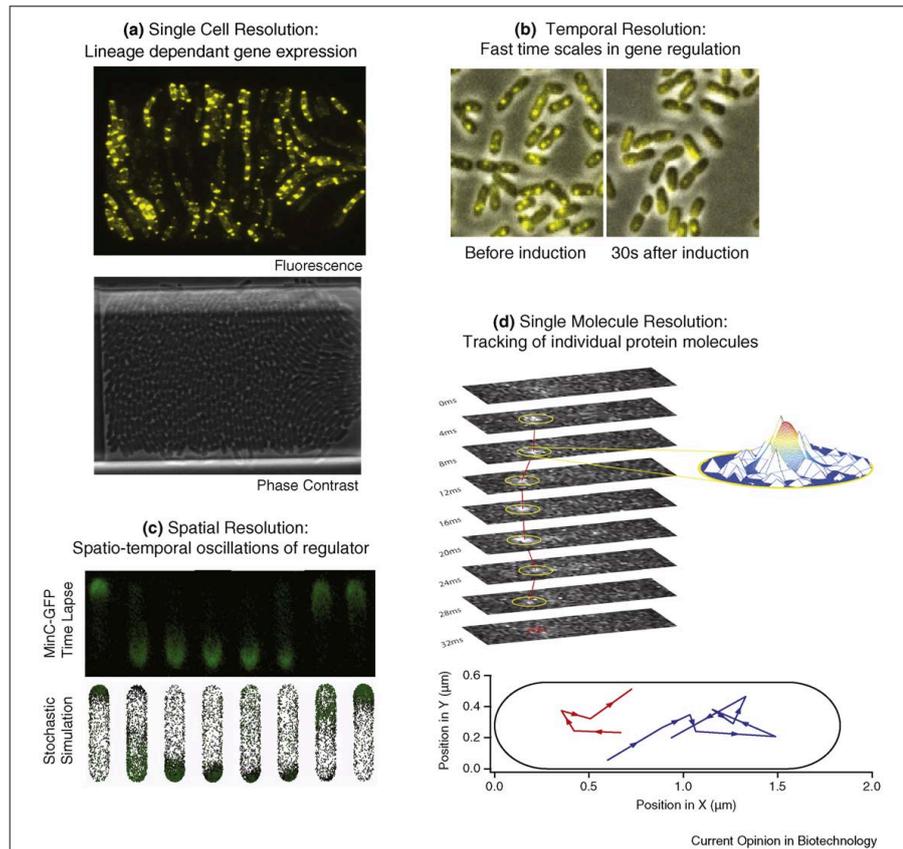
(Figure 1a). Population-averaged studies mask the cell-to-cell difference and assays with single-cell resolution are necessary to explore for example the reasons and consequences of heterogeneity. There are several different factors that can contribute to the variations between isogenic cells. Firstly, the chemical reactions involved in gene expression are inherently stochastic and will generate diversity [1–3]. Secondly, cells can be in different phases of the cell cycle or some other oscillating regulatory cycle [4]. Even if cells are initially synchronous, stochastic fluctuations will cause them to drift out of synchrony [5,6]. Thirdly, there are several epigenetic phenomena that will keep lineages with different histories different for generations even when growing under identical conditions [7,8]. Finally, unicellular organisms may also differ according to their age, as measured in the number of cell divisions they have experienced [9,10].

Gene expression can be studied in single cells using a variety of methods. For example by using flow cytometry the expression levels of fluorescent reporter proteins can be measured in many thousands of individual cells in a short time. Real-time PCR can also be used to quantify endogenous RNA molecules in single cells [11]. If the expression levels approach single copies per cell, methods with higher sensitivity are needed. For example, digital PCR makes absolute quantification of native mRNAs possible [12] and simultaneous quantification and localization of single intracellular mRNAs can be achieved by Fluorescence *in Situ* Hybridization, FISH [13]. Furthermore, expression levels of fluorescent reporter proteins can be quantified down to the level of single molecules in individual living bacterial cells by using sensitive microscopy [14]. This technique was recently used to quantify the genome-wide expression patterns of *Escherichia coli* at single-molecule sensitivity [15].

### Temporal resolution and dynamics

Static information on cell-to-cell variability can be used to infer dynamic properties of transcription based on the stationary distribution of a stochastic model [16,17]. However, important parameters are masked in time averaging and can only be obtained by monitoring cells over time [18,19]. Here, time-lapse fluorescence microscopy of fluorescent protein reporter constructs has proven an exceptionally powerful tool [20,21]. This approach has been used to study dynamic aspects of many regulatory circuits in single cells which may otherwise have been overlooked. For example, Dunlop *et al.* [22] recently reported on how time delayed correlations in gene expression can be used to

Figure 1



High-resolution fluorescence microscopy of gene expression. **(a)** Heterogeneous and lineage dependent gene expression from isogenic *E. coli* cells. *Top*: expression of fluorescent proteins from the lac operon in response to shift in carbon source (MW and JE, in preparation). *Bottom*: phase contrast image of the same cells. **(b)** Rapid induction of the *lac* operon as studied by dissociation of individual LacI-Venus repressors from the chromosomal binding site. The time-scale of induction would be masked in an assay based on detecting expressed reporter proteins. *Left*: cells without IPTG. *Right*: cells after addition of 100  $\mu$ M IPTG. (Petter Hammar, in preparation). **(c)** MinCD oscillations in *E. coli*. *Top*: the spatial oscillations of MinC regulates the position of cell division in *E. coli* [47]. *Bottom*: stochastic reaction-diffusion simulation in 3D based on the multivariate master equation [48,49]. **(d)** Single-molecule superresolution tracking of rapidly diffusing proteins in a bacterial cell (English et al. submitted).

infer active regulatory links and separate intrinsic from extrinsic noise. Another stunning example of time-dependent gene regulation is found in a study by Cagatay *et al.*, where they showed how design of competence circuits in *Bacillus subtilis* is under selection for fluctuations in the competence time [23\*].

However, the temporal resolution of assays based on expression of fluorescent proteins is limited, mainly by the reporters themselves. The instantaneous expression level of a gene is smeared over a period corresponding to

the maturation time of the fluorochrome, which can amount to several hours [24]. One way to improve the time resolution is to use fast-maturing fluorescent proteins, such as Venus, a yellow fluorescent protein variant [25] with a maturation time of around 7 min [18]. Another means of increasing the time resolution is to destabilize the reporter. By including a specific protease degradation tag, such as *ssrA*, the fraction of early matured proteins can be increased [26]. This is at the expense of the signal strength and also introduces noise from the proteolytic pathway into the signal.

Alternatively, it is possible to monitor pre-matured fluorescent protein reporters as they bind to nascent RNA molecules. This can be achieved by utilizing the affinity that fluorescently labeled MS2 phage coat protein exhibits to a 19 nt RNA hairpin structure which can be included repeatedly in the mRNA studied [19,27]. The time resolution is in this case limited by the association rate of the fluorescent fusion proteins. Higher expression levels of the fluorescent fusion protein will obviously decrease the time for binding, but also increase the fluorescent background such that more molecules will have to bind to detect the signal. The total expression level of the reporter has to be tuned so as to allow fast detection of nascent RNA molecules.

Several transcription factors exist at sufficiently low cellular copy number to enable detection over the background of a single fluorescently tagged protein molecule bound to its chromosomal binding site. This was used by Elf *et al.* [28\*] to determine the time it takes for a LacI transcription factor to find and repress its chromosomal operator in *E. coli* and its rate of dissociation after induction. In this case, the time resolution is only limited by the time it takes to distinguish specifically bound from freely diffusing proteins (<100 ms), which is much faster than the conventional indirect assays of gene regulation based on the expression of reporter proteins (Figure 1b).

### Spatial resolution and gene regulation

Important information encoded in the intracellular location of transcriptional regulators may also be lost when studying cells at insufficient spatial resolution. For example, in a recent study Cai *et al.* [29\*] demonstrated that the transcription factor Crz1 regulates expression of its target genes in response to the extracellular calcium concentration by oscillating between the soma and nucleus in *Saccharomyces cerevisiae* cells in a frequency modulated manner. The frequency modulation ensures that different regulated promoters can respond in proportion to the stimulus despite different binding strengths for the transcription factor. Striking examples of coordinated spatial and temporal oscillations of gene regulatory proteins can also be found in prokaryotic cells, such as in the cell cycle-dependent activities of master regulators [30] and mitotic apparatus [31] in *Caulobacter crescentus*; the regulated relocation of bacterial transcription factors between their chromosomal operators and the inner membrane [32]; as well as the irregular relocations from pole to pole or nucleoid to nucleoid of the Soj protein involved in sporulation and transcriptional regulation in *B. subtilis* [33].

All of these studies have relied on fluorescence microscopy to provide the spatial information. Conventional light microscopy is however diffraction-limited, in the sense that it is not possible to know from where in the sample

plane a photon originated with a better resolution than ~200 nm. The dynamics of cellular structures at finer resolutions have therefore remained hidden. However, several methods that allow far-field fluorescence microscopy at higher resolution have recently been devised [34]. For example Shroff *et al.* [35] used Photoactivated Localization Microscopy, PALM, to study dynamics within individual adhesion complexes in living mammalian cells at 60 nm spatial resolution and Huang *et al.* [36] used multicolor three-dimensional stochastic optical reconstruction microscopy, 3D-STORM, to visualize the topology of the mitochondrial network in fixed mammalian kidney cells at a 30 nm resolution. These methods will greatly improve our ability to study cellular structures and nanoscopic dynamics also in prokaryote cells. For example Biteen *et al.* [37] used superresolution time-lapse imaging to characterize the filamentous superstructure of the bacterial actin protein MreB in live *C. crescentus* cells at a resolution of 40 nm and Greenfield *et al.* [38] used PALM to characterize the organization and assembly of *E. coli* chemotaxis Tar receptors in fixed cells at a spatial resolution of 15 nm.

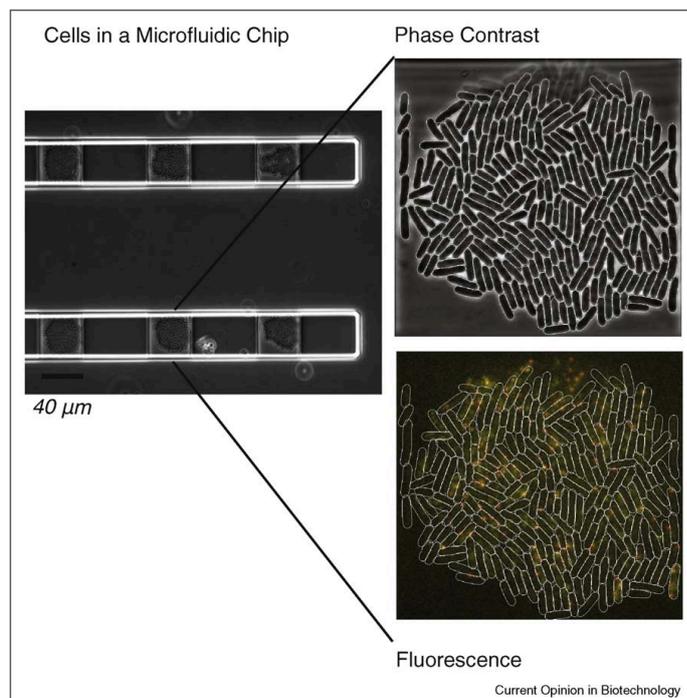
### Microfluidics in single-cell microscopy

Three pressing needs arise when performing single-cell gene expression studies: firstly observations of many cells and molecules have to be accumulated in order to draw significant conclusions, secondly the cells have to be in a representative physiological state and thirdly the chemical environment must be well-defined and preferably rapidly interchangeable. Recent advances in micro-fabrication have realized microfluidic devices capable of alleviating these needs for single-cell fluorescence microscopy [39,40] (Figure 2).

In order to keep microbial cells in constant good health during experiments, several micro-chemostats have been designed. For example Danino *et al.* [41\*] devised chips trapping a monolayer of bacterial cells in the focal plane. A constant occupancy of each trap was maintained as cells were released as they grew out of the trap.

The microscopic length scale characteristic of these devices shifts the properties of the fluids into a regime in which little or no turbulent mixing occurs. This allows for predictable and automatable on-chip composition of the cells chemical environment as well as its rapid exchange, increasing both the time and inductive resolution at which gene expression can be studied [6\*]. Most importantly, trapping regions are easily multiplexed on a chip, allowing a massive parallelization of a single experiment [15]. Combined with the indefinitely prolonged times that cells can be maintained and imaged [50] results in data sets, the magnitude of which can dramatically increase confidence in the conclusions drawn, even for rare events [42] and better merits the comparison with bulk experiments.

Figure 2



*E. coli* cells in a microfluidic device. *Right*: parallel trapping regions each containing 200–300 cells imaged with phase contrast. *Top left*: a single trap imaged with phase contrast allowing segmentation and tracking of individual cells. *Bottom left*: fluorescence image of membrane-bound Venus molecules resulting from leakage expression from the lac operon (MW and JE, in preparation).

### Resolving Individual molecular states

So far we have described phenomena that may be masked in experiments lacking resolution of single cells, in time or space. However, since macromolecules with the same primary sequences, that is isoform molecules, can be in different states of activity, conformation or binding, studying only the total molecule concentration at some intracellular location may obscure important biological information. Furthermore, the average properties of a molecular ensemble can only reveal kinetics of interconversion between such states if it is first perturbed out of steady state. The situation is different if the state transitions of single molecules can be traced directly. In this case the chemical rates can be determined also at steady state without the need to synchronize the ensemble of molecules, which is often not possible *in vivo*.

*In vitro*, very powerful single-molecule assays based on for example Förster Resonance Energy Transfer, FRET [43] or fluorogenic product reactions have been reported [44]. Unfortunately, many such schemes remain difficult to

realize *in vivo*, mainly because of high background auto-fluorescence, and in the case of FRET photo-instability of genetically encoded fluorescent reporters. One way to differentiate bound from free molecules in live cells would be to study their individual diffusion trajectories, as obtained by single-molecule tracking of fluorescent fusion proteins [51]. Inside living cells it has however proven a challenge to collect a sufficient number of photons from individual rapidly moving fluorescent proteins to pinpoint their positions. This has previously limited the range of possible targets to slowly diffusing molecules, such as membrane-associated proteins. For example, Niu and Yu [45] studied FtsZ labeled with the photo-convertible red fluorescent protein Dendra2b, and characterized the differences between two FtsZ subpopulations in *E. coli*.

By combining superresolution localization of individual fluorescently labeled proteins with stroboscopic laser excitation it is also possible to capture the movement of individual transcription factor molecules searching for their specific binding sites [28\*]. Recently this method

was improved such that it is possible to track individual rapidly moving fluorescent proteins in *E. coli* (Figure 1d) (English *et al.*, in preparation). This capability suggests that it is also possible to track fluorescent fusion proteins diffusing through the cytoplasm in free or bound states. By monitoring the transitions between such states of different diffusivities it will be possible to acquire binding and dissociation rates for proteins in living cells.

### Conclusions

Currently, gene expression and gene regulation can be studied in individual cells, and population-averaged models are now being understood from the perspective of individual contributions from heterogeneous ensembles of isogenic cells. Increased spatiotemporal resolution has enabled the correlation of expression to the diverse and dynamic states of individual regulator molecules. Furthermore, the chemical control and parallelization possible in microfluidic devices afford the exhaustive exploration of inductive responses as well as confidence in the results.

One of the challenges for the immediate future is to adapt these high-precision methods to the study of eukaryotic cells. Here the large volume and high auto fluorescent background make detection, accurate counting and fast tracking of single fluorescent proteins difficult, but probably not impossible. Also, more powerful schemes to interrogate the state of activity of individual molecules in living cells are needed. To date, single-molecule tracking can be used to measure binding kinetics and FRET to determine conformational changes *in vitro*, but the *in vivo* implementations of both methods remain limited. Brighter and more photo-stable fluorophores, as well as improved non-perturbative and specific labeling schemes may be a prerequisite and the search for both continues [52].

Another important challenge is to find ways to measure the dynamics of individual metabolites in single cells as they often act as primary signals for gene regulatory systems. Beyond this, metabolite concentrations are often the entity which the cell aims to regulate and a direct readout of the investment in metabolite pools may therefore be more revealing than that of the enzymes producing them [53]. Although some fast and bright FRET probes have been developed to this end [54], a wider range of probes and targets is needed.

Finally, the large amount of hi-resolution data available from single-cell experiments may be manageable only by automated analysis tools and intelligible only within the framework of correspondingly detailed quantitative models. New challenges therefore arise in how to automate the analysis of complex data as well as in how to model and simulate intracellular processes at the right level of detail to capture their central properties [55]. We should however remember that the key details may be hidden at spatial and temporal scales that we do not yet have the tools to explore,

in which case quantitative modeling should point the direction to what remains to be discovered.

### Conflicts of interest

The authors declare that they have no conflicts of interest.

### Acknowledgements

The authors would like to thank Gustaf Ullman, Gene-Wei Li and Petter Hammar for their contribution to this review and Prof. Jeff Hasty for help in constructing the  $\mu$ -fluidic devices. This work was funded by the ERC (no 203083), the Swedish Research Council and the Knut and Alice Wallenberg Foundation.

### References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest

1. Berg OG: **A model for the statistical fluctuations of protein numbers in a microbial population.** *J Theor Biol* 1978, **71**:587-603.
2. Elowitz MB, Levine AJ, Siggia ED, Swain PS: **Stochastic gene expression in a single cell.** *Science* 2002, **297**:1183-1186.
3. Mettetal JT, Muzzey D, Pedraza JM, Ozbudak EM, van Oudenaarden A: **Predicting stochastic gene expression dynamics in single cells.** *Proc Natl Acad Sci USA* 2006, **103**:7304-7309.
4. Laub MT, McAdams HH, Feldblyum T, Fraser CM, Shapiro L: **Global analysis of the genetic network controlling a bacterial cell cycle.** *Science* 2000, **290**:2144-2148.
5. Elowitz MB, Leibler S: **A synthetic oscillatory network of transcriptional regulators.** *Nature* 2000, **403**:335-338.
6. Stricker J, Cookson S, Bennett MR, Mather WH, Tsimring LS, Hasty J: **A fast, robust and tunable synthetic gene oscillator.** *Nature* 2008, **456**:516-519.
7. Robert L, Paul G, Chen Y, Taddei F, Baigl D, Lindner AB: **Pre-dispositions and epigenetic inheritance in the Escherichia coli lactose operon bistable switch.** *Mol Syst Biol* 2010, **6**:357.
8. Choi PJ, Cai L, Frieda K, Xie XS: **A stochastic single-molecule event triggers phenotype switching of a bacterial cell.** *Science* 2008, **322**:442-446.
9. Stewart EJ, Madden R, Paul G, Taddei F: **Aging and death in an organism that reproduces by morphologically symmetric division.** *PLoS Biol* 2005, **3**:e45.
10. Barker MG, Walmsley RM: **Replicative ageing in the fission yeast *Schizosaccharomyces pombe*.** *Yeast* 1999, **15**:1511-1518.
11. Bengtsson M, Stahlberg A, Rorsman P, Kubista M: **Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels.** *Genome Res* 2005, **15**:1388-1392.
12. Warren L, Bryder D, Weissman IL, Quake SR: **Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR.** *Proc Natl Acad Sci USA* 2006, **103**:17807-17812.
13. Femino AM, Fay FS, Fogarty K, Singer RH: **Visualization of single RNA transcripts in situ.** *Science* 1998, **280**:585-590.
14. Xiao J, Elf J, Li G-W, Yu J, Xie X: **Imaging gene expression in living cells at the single-molecule level.** In *Single Molecule Techniques A Laboratory Manual*. Edited by Selvin P, Ha T. Cold Spring Harbor Laboratory Press; 2008.

15. Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, Emilii A, Xie XS: **Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells.** *Science* 2010, **329**:533-538.  
Genome-wide gene expression levels at single-molecule sensitivity.
16. Friedman N, Cai L, Xie XS: **Linking stochastic dynamics to population distribution: an analytical framework of gene expression.** *Phys Rev Lett* 2006, **97**:168302.
17. Paulsson J: **Summing up the noise in gene networks.** *Nature* 2004, **427**:415-418.
18. Yu J, Xiao J, Ren X, Lao K, Xie XS: **Probing gene expression in live cells, one protein molecule at a time.** *Science* 2006, **311**:1600-1603.
19. Golding I, Paulsson J, Zawilski SM, Cox EC: **Real-time kinetics of gene activity in individual bacteria.** *Cell* 2005, **123**:1025-1036.
20. Pedraza JM, van Oudenaarden A: **Noise propagation in gene networks.** *Science* 2005, **307**:1965-1969.
21. Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB: **Gene regulation at the single-cell level.** *Science* 2005, **307**:1962-1965.
22. Dunlop MJ, Cox RS III, Levine JH, Murray RM, Elowitz MB: **Regulatory activity revealed by dynamic correlations in gene expression noise.** *Nat Genet* 2008, **40**:1493-1498.  
Shows how dynamic correlations can be used to infer activity in regulatory links.
23. Cagatay T, Turcotte M, Elowitz MB, Garcia-Ojalvo J, Suel GM: **Architecture-dependent noise discriminates functionally analogous differentiation circuits.** *Cell* 2009, **139**:512-522.  
Shows how selection of noise distributions may be needed to understand evolution of biological circuits.
24. Dong GQ, McMillen DR: **Effects of protein maturation on the noise in gene expression.** *Phys Rev E Stat Nonlin Soft Matter Phys* 2008, **77**:021908.
25. Nagai T, Ibata K, Park ES, Kubota M, Mikoshiba K, Miyawaki A: **A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications.** *Nat Biotechnol* 2002, **20**:87-90.
26. Andersen JB, Sternberg C, Poulsen LK, Bjorn SP, Givskov M, Molin S: **New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria.** *Appl Environ Microbiol* 1998, **64**:2240-2246.
27. Fusco D, Accornero N, Lavoie B, Shenoy SM, Blanchard JM, Singer RH, Bertrand E: **Single mRNA molecules demonstrate probabilistic movement in living mammalian cells.** *Curr Biol* 2003, **13**:161-167.
28. Elf J, Li GW, Xie XS: **Probing transcription factor dynamics at the single-molecule level in a living cell.** *Science* 2007, **316**:1191-1194.  
Shows how single-molecule imaging can be used to study gene regulation directly.
29. Cai L, Dalal CK, Elowitz MB: **Frequency-modulated nuclear localization bursts coordinate gene regulation.** *Nature* 2008, **455**:485-490.  
Illustrates the need for spatial resolution understand biological phenomena.
30. Collier J, Shapiro L: **Spatial complexity and control of a bacterial cell cycle.** *Curr Opin Biotechnol* 2007, **18**:333-340.
31. Ptacin JL, Lee SF, Garner EC, Toro E, Eckart M, Comolli LR, Moerner WE, Shapiro L: **A spindle-like apparatus guides bacterial chromosome segregation.** *Nat Cell Biol* 2010, **12**:791-798.
32. Bohm A, Boos W: **Gene regulation in prokaryotes by subcellular relocalization of transcription factors.** *Curr Opin Microbiol* 2004, **7**:151-156.
33. Doubrovinski K, Howard M: **Stochastic model for Soj relocation dynamics in Bacillus subtilis.** *Proc Natl Acad Sci USA* 2005, **102**:9808-9813.
34. Hell SW: **Far-field optical nanoscopy.** *Science* 2007, **316**:1153-1158.
35. Shroff H, Galbraith CG, Galbraith JA, Betzig E: **Live-cell photoactivated localization microscopy of nanoscale adhesion dynamics.** *Nat Methods* 2008, **5**:417-423.
36. Huang B, Jones SA, Brandenburg B, Zhuang X: **Whole-cell 3D STORM reveals interactions between cellular structures with nanometer-scale resolution.** *Nat Methods* 2008, **5**:1047-1052.
37. Biteen JS, Thompson MA, Tselentis NK, Bowman GR, Shapiro L, Moerner WE: **Super-resolution imaging in live Caulobacter crescentus cells using photoswitchable EYFP.** *Nat Methods* 2008, **5**:947-949.
38. Greenfield D, McEvoy AL, Shroff H, Crooks GE, Wingreen NS, Betzig E, Liphardt J: **Self-organization of the Escherichia coli chemotaxis network imaged with super-resolution light microscopy.** *PLoS Biol* 2009, **7**:e1000137.
39. Weibel DB, Diluzio WR, Whitesides GM: **Microfabrication meets microbiology.** *Nat Rev Microbiol* 2007, **5**:209-218.
40. Breslauer DN, Lee PJ, Lee LP: **Microfluidics-based systems biology.** *Mol Biosyst* 2006, **2**:97-112.
41. Danino T, Mondragon-Palomino O, Tsimring L, Hasty J: **A synchronized quorum of genetic clocks.** *Nature* 2010, **463**:326-330.  
Contains interesting design principles for genetic circuitry as well as microfluidic devices for maintaining cells for prolonged times.
42. Balaban NQ, Merrin J, Chait R, Kowalik L, Leibler S: **Bacterial persistence as a phenotypic switch.** *Science* 2004, **305**:1622-1625.
43. Roy R, Hohng S, Ha T: **A practical guide to single-molecule FRET.** *Nat Methods* 2008, **5**:507-516.
44. English BP, Min W, van Oijen AM, Lee KT, Luo G, Sun H, Cherayil BJ, Kou SC, Xie XS: **Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited.** *Nat Chem Biol* 2006, **2**:87-94.
45. Niu L, Yu J: **Investigating intracellular dynamics of FtsZ cytoskeleton with photoactivation single-molecule tracking.** *Biophys J* 2008, **95**:2009-2016.
47. Meinhardt H, de Boer PA: **Pattern formation in Escherichia coli: a model for the pole-to-pole oscillations of Min proteins and the localization of the division site.** *Proc Natl Acad Sci USA* 2001, **98**:14202-14207.
48. Hattne J, Fange D, Elf J: **Stochastic reaction-diffusion simulation with MesoRD.** *Bioinformatics* 2005, **21**:2923-2924.
49. Fange D, Elf J: **Noise-induced Min phenotypes in E. coli.** *PLoS Comput Biol* 2006, **2**:e80.
50. Balagaddé FK, You L, Hansen CL, Arnold FH, Quake SR: **Long-term monitoring of bacteria undergoing programmed population control in a microchemostat.** *Science* 2005, **309**:137-140.
51. Manley S, Gillette JM, Patterson GH, Shroff H, Hess HF, Betzig E, Lippincott-Schwartz J: **High-density mapping of single-molecule trajectories with photoactivated localization microscopy.** *Nat Methods* 2008, **5**:155-157.
52. Uttamapinant C, White KA, Baruah H, Thompson S, Fernández-Suárez M, Puthenveetil S, Ting AY: **A fluorophore ligase for site-specific protein labeling inside living cells.** *Proc Natl Acad Sci USA* 2010, **107**:10914-10919.
53. Elf J, Ehrenberg M: **Near-critical behavior of aminoacyl-tRNA pools in E. coli at rate-limiting supply of amino acids.** *Biophys J* 2005, **88**:132-146.
54. Horikawa K, Yamada Y, Matsuda T, Kobayashi K, Hashimoto M, Matsu-ura T, Miyawaki A, Michikawa T, Mikoshiba K, Nagai T: **Spontaneous network activity visualized by ultrasensitive Ca<sup>2+</sup> indicators, yellow Cameleon-Nano.** *Nat Methods* 2010, **7**:729-732.
55. David Fange, Otto Berg, Paul Sjöberg, Johan Elf: **Stochastic reaction-diffusion kinetics in the microscopic limit.** *Proc Natl Acad Sci* 2010, doi:10.1073/pnas.1006565107.

Paper II 



## High-throughput gene expression analysis at the level of single proteins using a microfluidic turbidostat and automated cell tracking

G. Ullman, M. Wallden, E. G. Marklund, A. Mahmutovic, Ivan Razinkov and J. Elf

*Phil. Trans. R. Soc. B* 2013 **368**, 20120025, published 24 December 2012

---

### Supplementary data

["Data Supplement"](#)

<http://rstb.royalsocietypublishing.org/content/suppl/2012/12/14/rstb.2012.0025.DC1.html>

### References

[This article cites 15 articles, 7 of which can be accessed free](#)

<http://rstb.royalsocietypublishing.org/content/368/1611/20120025.full.html#ref-list-1>

[Article cited in:](#)

<http://rstb.royalsocietypublishing.org/content/368/1611/20120025.full.html#related-urls>

### Subject collections

Articles on similar topics can be found in the following collections

[biophysics](#) (99 articles)  
[cellular biology](#) (209 articles)  
[microbiology](#) (53 articles)  
[molecular biology](#) (187 articles)  
[systems biology](#) (75 articles)

### Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)



CrossMark  
click for updates

## Research

**Cite this article:** Ullman G, Wallden M, Marklund EG, Mahmutovic A, Razinkov I, Elf J. 2013 High-throughput gene expression analysis at the level of single proteins using a microfluidic turbidostat and automated cell tracking. *Phil Trans R Soc B* 368: 20120025. <http://dx.doi.org/10.1098/rstb.2012.0025>

One contribution of 12 to a Theme Issue 'Single molecule cellular biophysics: combining physics, biochemistry and cell biology to study the individual molecules of life'.

### Subject Areas:

biophysics, microbiology, molecular biology, systems biology, cellular biology

### Keywords:

microfluidics, single molecule, image analysis, cell tracking, gene expression, growth

### Author for correspondence:

J. Elf  
e-mail: [johan.elf@icm.uu.se](mailto:johan.elf@icm.uu.se)

<sup>†</sup>These authors contributed equally to this study.

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rstb.2012.0025> or via <http://rstb.royalsocietypublishing.org>.

# High-throughput gene expression analysis at the level of single proteins using a microfluidic turbidostat and automated cell tracking

G. Ullman<sup>1,2,†</sup>, M. Wallden<sup>1,†</sup>, E. G. Marklund<sup>1</sup>, A. Mahmutovic<sup>1</sup>, Ivan Razinkov<sup>3</sup> and J. Elf<sup>1</sup>

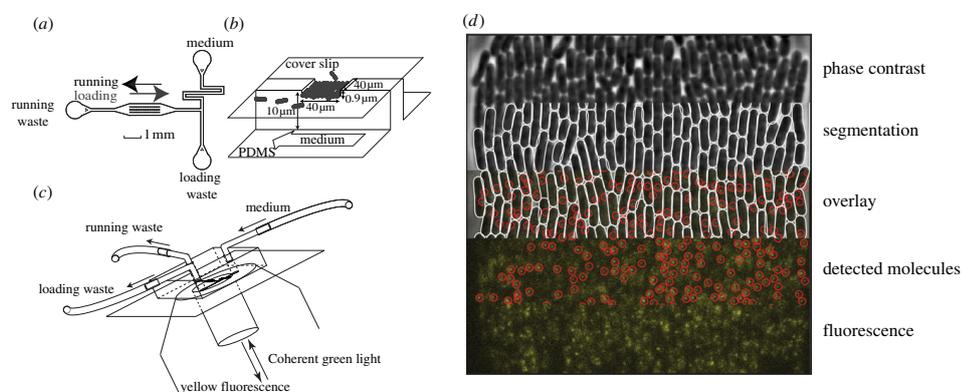
<sup>1</sup>Department of Cell and Molecular Biology, Science for Life Laboratory, and <sup>2</sup>Division of Scientific Computing, Department of Information Technology, Uppsala University, 75105 Uppsala, Sweden  
<sup>3</sup>Department of Bioengineering, University of California, La Jolla, San Diego, CA 92093, USA

We have developed a method combining microfluidics, time-lapsed single-molecule microscopy and automated image analysis allowing for the observation of an excess of 3000 complete cell cycles of exponentially growing *Escherichia coli* cells per experiment. The method makes it possible to analyse the rate of gene expression at the level of single proteins over the bacterial cell cycle. We also demonstrate that it is possible to count the number of non-specifically DNA binding LacI–Venus molecules using short excitation light pulses. The transcription factors are localized on the nucleoids in the cell and appear to be uniformly distributed on chromosomal DNA. An increase in the expression of LacI is observed at the beginning of the cell cycle, possibly because some gene copies are de-repressed as a result of partitioning inequalities at cell division. Finally, a size–growth rate uncertainty relation is observed where cells living in rich media vary more in the length at birth than in generation time, and the opposite is true for cells living in poorer media.

## 1. Introduction

Using time-lapsed phase-contrast and fluorescence microscopy, it is possible to monitor live bacterial cells and simultaneously quantify the expression of their highly expressed genes as the activity of introduced fluorescence reporters [1]. However, for many of its native protein species, a bacterial cell expresses only a few copies per generation [2]. In order to study processes involving these proteins, fluorescence microscopy methods sufficiently sensitive to resolve individual molecules have been developed. For instance, Yu *et al.* [3] reported on the use of a fast maturing yellow fluorescent protein (YFP) variant, Venus [4], fused to a membrane tag, Tsr, to profile the absolute expression of the *lacZ* gene, in live *Escherichia coli* cells, in its repressed state. The Tsr domain immobilizes the fluorophore at the membrane so that it appears stationary for periods of 50–100 ms and can be detected as a diffraction-limited spot. However, tethering to the membrane will disable molecules that rely on intracellular mobility for their function. For this reason, methods for counting expression events for cytoplasmic proteins are limited. A possible solution is suggested by the single-molecule tracking experiments where stroboscopic illumination pulses were used to image the transcription factor LacI–Venus non-specifically bound to DNA in live *E. coli* cells [5]. This suggests that short excitation pulses could be used also to profile the synthesis of cytoplasmic low copy number transcription factors or other proteins binding to relatively immobile intracellular targets.

Single-protein counting experiments *in vivo* reveal that isogenic cells under seemingly identical experimental conditions display considerable diversity in expression [6]. In order to confidently draw conclusions on the nature of this



**Figure 1.** The experimental set-up and data processing. (a) The microfluidic device has three ports designated for medium, running waste and loading waste. The chamber houses three rows, each containing 17 traps. The direction of the flow through the chamber is alternated between the loading and running phase of the experiment. The cells are introduced from the running waste and are caught in the traps. (b) Each trap is a  $40 \times 40 \times 0.9 \mu\text{m}$  compartment which is bounded by two rigid walls and two openings. Cells that reach the openings are released from the traps into the  $10 \mu\text{m}$  deep surrounding. (c) The device is connected to reservoirs at the ports and imaged using an inverted microscope. The various parts of the microfluidic chip are not drawn in scale. (d) Data processing: cells are detected and segmented from the phase-contrast image (top). Molecules are detected within the fluorescence images (bottom). The coordinates from the detected molecules and cells are used to map molecules to cells (middle).

diversity, it is necessary to sample a sufficient number of cells. Several microfluidic devices have been reported to substantially increase experimental throughput by harnessing the reproduction of bacterial cells to continuously regenerate the sample and also allowing imaging of many replicate colonies in parallel [7,8]. However, the sheer size of image datasets that can be generated in this fashion overwhelms manual analysis efforts and consequently several initiatives of automation have been undertaken [9,10]. In this study, we report on a method combining microfluidics, single-molecule fluorescence microscopy and automated image analysis, enabling the study of the expression and super-resolution localization of low copy number transcription factors throughout thousands of bacterial lifespans per experiment. To illustrate the performance of the method, we quantify the dynamics of synthesis and intracellular localization of the lactose repressor by monitoring LacI–Venus expressed from its native promoter in live *E. coli* cells. We compare these observations with those obtained under identical conditions for cells expressing the reporter construct Tsr–Venus from the lactose permease gene, *lacY*, of the lactose operon.

## 2. Material and methods

### (a) Design, fabrication and use of the microfluidic device

The chip design was inspired by Mather *et al.* [11]. The features of the microfluidic chip used in this study were designed in three layers using AutoCAD. The layers correspond to structures of different step heights of the mould and ultimately to the different depths of the structures of the finished microfluidic device (described under ‘mould fabrication’ and ‘chip fabrication’). The device contains four structural motifs: ports, channels, a chamber and traps (figure 1a). The chamber houses three

evenly spaced rows, each containing 17 traps (figure 1a). Each trap is  $40 \times 40 \times 0.9 \mu\text{m}$  (figure 1b), and is bounded by two opposite walls and two open sides connecting the trap to the  $10 \mu\text{m}$  deep surrounding. This geometry restricts the cells to form a monolayer colony in the focal plane while imaging. Cells close to the openings are released as the colony expands (figure 1b). The microfluidic device is connected to media reservoirs and imaged using an inverted microscope (figure 1c).

The master mould was fabricated using standard UV-soft lithography techniques. Three masks for microfabrication were printed in chrome. Custom formulations of SU8 Photoresist (MicroChem) were deposited on clean polished silicon wafers (University Wafer) using a spin coater. The wafers were then aligned to the mask and exposed using a mask aligner (Süss MA6). This process was repeated to deposit layers of step heights  $0.9$ ,  $2.7$  and  $10 \mu\text{m}$  per wafer. The first layer corresponds to the trap depth of the microfluidic device; the intermediate layer enables the alignment of the first and third layer, corresponding to the channels and ports. Each layer of the moulds was measured using a stylus profilometer and inspected under a microscope before applying the next.

A master cast of the mould was made from polydimethylsiloxane (Syngaard 184, Dow Corning), using the master mould. Bubbles were removed by vacuum desiccation. The cast was cured at  $80^\circ\text{C}$  for 30 min. One master cast contained 12 identical chip structures, which could be excised and used individually.

When fabricating each device, port holes ( $0.5 \text{ mm}$  diameter) were punched out of the device cast. Debris was removed from the cast by vortexing in ethanol. The chip cast was bonded to a coverslip ( $40 \text{ mm}$  diameter,  $200 \mu\text{m}$  thick, Thermo-Scientific) after oxygen/UV plasma treatment (UVO-cleaner 42–220, Jelight Co.) for 5 min at  $0.5 \text{ bar}$  oxygen pressure. The bond was stabilized by incubating at  $80^\circ\text{C}$  for 10 min. Just prior to loading and running the device, the ports were treated with a high-frequency generator (model BA 20 D, Electro-Technic Products Inc.), and the device was flooded with de-ionized water.

Gravity flow was used to control the direction and the magnitude of the flow inside the microfluidic device. The pressure gradients between the different ports of the device were established by differences in elevation relative to the sample of the

connected reservoirs. During loading, the seeding culture was introduced into the device through the running waste port. The cells were caught in the traps by introducing pressure waves into the tubing. Once all traps were sufficiently occupied (10–100 cells per trap), the direction of the flow in the chamber was reversed, exchanging the seeding culture with fresh medium (figure 1a). The cells were allowed to acclimatize and grow until the traps were fully occupied (approx. 4 h) before imaging. The temperature of the sample was maintained at 37°C using a custom-fitted incubator hood (OKO LAB).

### (b) Strains and medium

Two bacterial strains, SX701 and JE116, based on *E. coli* strain BW25993 [12], were used in this study. In strain SX701, the lactose permease gene, *lacY*, was replaced with the *tsr-venus* construct [13]. Strain JE116 is based on strain JE12 [5], in which the *lacI* gene was modified to encode a C-terminal fusion of LacI and Venus. The auxiliary lactose operator site, *O3*, was replaced with the main operator sequence, *O1*, to increase auto-repression by LacI threefold. Further, in strain JE116 the downstream sequence including the native *O1*, *O2* binding sites as well as parts of the *lacZ* gene was removed, leaving only one specific binding site sequence for LacI–Venus molecules per chromosome copy [14].

Cells were grown in M9 minimal medium, with 0.4 per cent glucose, either with or without supplemented amino acids (RPM11640 (R7131), Sigma–Aldrich). An overnight culture was diluted 200 times in 40 ml fresh medium and incubated for 3–5 h (6–8 h for cells grown without amino acids) at 37°C and shaking at 225 rpm. During this incubation, the microfluidic device was prepared. Cells were harvested into a seeding culture by centrifugation at 5000 × rcf for 2.5 min and the pellet resuspended in 50–100 µl fresh medium. In order to prevent the cells from sticking to the surfaces of the microfluidic device a surfactant, Pluronic F108 (prod. Number 542342, Sigma–Aldrich) was added to all media to a final concentration of 0.85 g l<sup>-1</sup>.

### (c) Microscopy and imaging

Imaging was performed using an inverted microscope (Ti Eclipse, Nikon) fitted with a high numerical aperture oil objective (APO TIRF 100×/N.A 1.49, Nikon) and external phase contrast to minimize loss of fluorescence signal. The phase-contrast channel and the fluorescence channels were imaged using separate cameras, a model CFW-1312M (Scion Corporation) and an Ixon EM plus (Andor Technologies), respectively. Focus was maintained by the *perfect-focusing-system* of the microscope. The light source for fluorescence excitation was an Argon ion laser (Innova 300, Coherent Inc.) dialled to 514 nm for excitation of YFP-reporters in the sample. For fluorescence imaging, a slower shutter (LS6Z2, Uniblitz) was used for strain SX701 (Tsr–Venus) and a fast shutter (LS2Z2, Uniblitz) was used for strain JE116 (LacI–Venus). The fast shutter was controlled using a signal generator (AFG3021B, Tektronix), which was triggered by the Ixon camera, exposing the sample for 1 ms. A 2× magnification lens was used in the fluorescence emission path to distribute the point spread function ideally on the 16 µm pixels of the EMCCD. Image acquisition was performed using RITACQUIRE, an in house GUI-based plugin for MICRO-MANAGER (v. 1.3.4.7, [www.micro-manager.org](http://www.micro-manager.org)). In each experiment, three positions (traps) were subjected to the following acquisition program in parallel: every 30 s (every frame), a phase-contrast image (125 ms exposure) was taken for all positions. Every 3 min (1/6 frames) for all positions, in addition to the phase-contrast image, two fluorescence images (50 ms exposure for SX701 and 1 ms exposure for JE116) were taken in rapid succession, followed by a bright field image (100 ms exposure) of the fluorescence channel, i.e. using the white-light lamp of the

microscope as illumination source. This programming cycle was repeated for 1001 frames (8.3 h). Fluorescence images were acquired in tandem to account for the effects of bleaching on molecular counting (see §2f). The bright field images were acquired to allow alignment of phase-contrast and fluorescence images for each frame. Our automatic method for cropping the phase images and aligning them to the fluorescent images is described in the electronic supplementary material, methods.

### (d) Cell segmentation and tracking

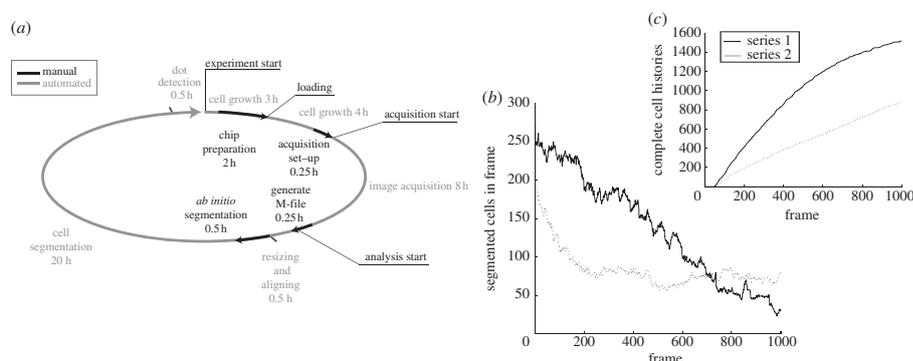
For segmenting and tracking individual cells in the microfluidic device, we have modified and further developed existing MATLAB software, MICROBETRACKER [10]. MICROBETRACKER uses the position of cells in the previous frame as an initial guess and applies an active contour model [15] to fit each cell with a sub-pixel resolution boundary. In order to accurately track mobile cells over several generations, three additional supervised algorithms [16] were implemented in MATLAB to complement MICROBETRACKER (see the electronic supplementary material, Methods): a *cell pole tracker* and two separate *error detectors*. The cell pole tracker is used to help the active contour model find the cell poles correctly for moving cells; otherwise this will lead to error propagation in the subsequent frames. The first error detector identifies errors made by the cell pole tracker. This is usually the result of an occasional large displacement of the cell between frames. This activates the cell tracker, which attempts to correct the segmentation of the erroneous cell. The accuracy of the cell tracker is in turn monitored by a second error detector. Any cell histories triggering this detector are terminated. In addition, a novel *division function* was added to MICROBETRACKER in order to more accurately detect cell divisions for densely growing *E. coli*. Each supervised algorithm was constructed by first identifying features that efficiently discriminate between two classes, for instance, true or false cell division. In the second step, training data were extracted manually from the image sets for creating training examples for the algorithm in order to achieve accurate classification. A linear classifier [16] was used in all supervised algorithms. The algorithms, cell tracker and the classification method are described in detail in the supplementary methods. To increase the computational speed, parts of MICROBETRACKER were rewritten to allow parallel computing, using MATLAB's parallel computing toolbox.

### (e) Single molecule detection, localization

Fluorescent particles in the sample were detected as diffraction limited spots in the fluorescence micrographs according to the method described in Ronneberger *et al.* [17], in which the normalized cross-correlation between the fluorescence image and an idealized optical point spread function (a symmetric bi-variate Gaussian function) is calculated. The standard deviation (s.d.) for this function is obtained experimentally by imaging and the signatures of immobilized highly fluorescent beads (data not shown). The image resulting from the correlation is transformed using the Fisher transform. A Fisher transformed Gaussian function with s.d. corresponding to the point spread function is fitted to the Fisher transformed correlation image using the Levenberg–Marquardt method [18] implemented in MATLAB, and the obtained parameters are used to localize each molecule with super-resolution accuracy and estimate the localization error.

### (f) Maximum-likelihood estimate of synthesis

For gene expression studies, we want to estimate how many molecules have been newly synthesized between two fluorescence images given that there is a chance that some of the fluorophores present in the previous frame have not been bleached. We formulate this as a maximum-likelihood problem



**Figure 2.** Throughput of the method. (a) Time distribution for an experiment. The minimum time expenditure for an experiment is around 36 h. Currently, the method generates around 3000 complete cell generations per experiment. The protocol contains mostly automated steps, i.e. they require no attention from the operator. The manual time expenditure accounts for less than 10% of the total time and less than 2% of the analysis time. (b) The number of segmented cells remaining at different frames during the analysis of two different time-series. Owing to large movements in the cell colony, the loss of correctly segmented cells varies between series. The nature of the decay is observed to depend on the pattern in which the cells grow in the trap, for which no sufficiently accurate prediction model has yet been found. (c) The integrated numbers of cells acquired from cell division to cell division, i.e. the number of complete cell histories, for the two series in (b).

where there are  $M$  molecules observed in frame  $i - 1$  and  $N$  molecules observed in frame  $i$ . The number of molecules surviving bleaching,  $m$ , can be calculated by maximizing the probability

$$p(m|M, N, p, \lambda) = \text{Bin}(m, M, 1 - p) \cdot \text{Po}(N - m, \lambda),$$

where Bin is the binomial distribution and Po is the Poisson distribution. The maximum-likelihood estimate of the number of new synthesized molecules is  $n_{\max} = N - m_{\max}$  where  $m_{\max}$  maximizes  $p(m|M, N, p, \lambda)$ . The parameter  $p$  is the bleaching probability per fluorophore per frame and is assumed to be constant.  $\lambda$  is the number of molecules synthesized between two frames.

In the special case of cell division between frame  $i - 1$  and  $i$ , where  $N_1$  molecules are found in one daughter cell and  $N_2$  in the other, the most likely number of newly synthesized molecules  $n_{\max}$  are calculated for both cells based on  $N = N_1 + N_2$ . Given  $n_{\max}$  the most likely number of newly synthesized molecules in daughter cell 1 is the  $n_1$  that maximizes  $\binom{N_1}{n_1} \binom{N_2}{n_{\max} - n_1}$  because this gives the number of possible combinations of picking  $n_1$  molecules from  $N_1$  and  $n - n_1$  from  $N_2$ .

### (g) Availability

All programs and scripts developed for this study will be made available on request.

## 3. Results

### (a) Throughput

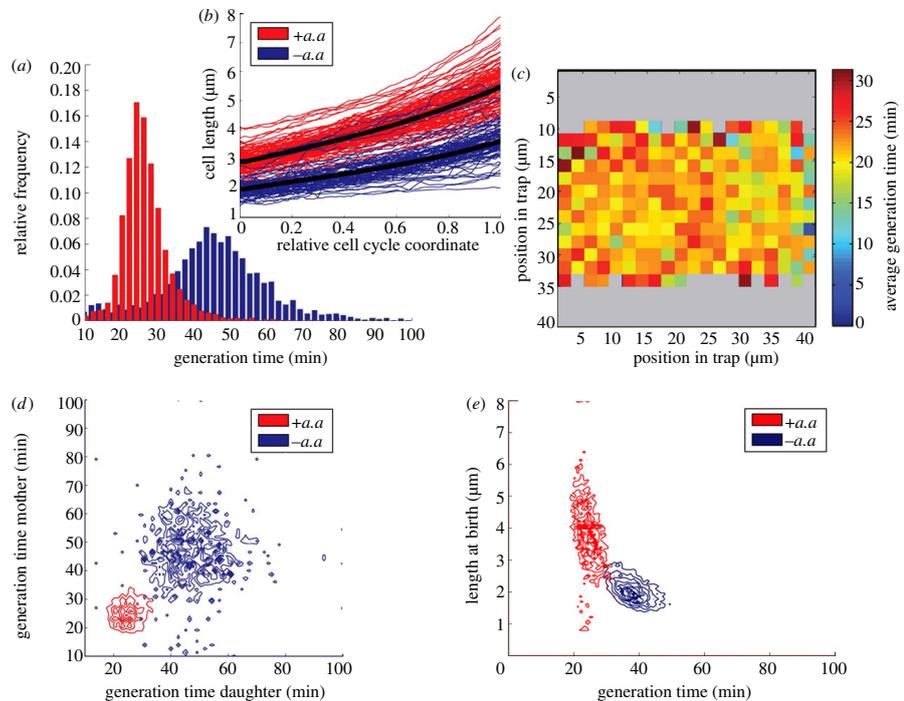
Currently, one experiment returns approximately 3000 complete cell histories from three traps imaged in parallel. The total time of expenditure is 36 h. The manual effort of a single operator amounts to 3 h, of which roughly 80 per cent is spent prior to image acquisition. The manual work effort to acquire and analyse the images constitutes less than 2 per cent of the total time required to complete these processes (figure 2a). Several overlapping experiments can

be performed to use the alternating availability of the microscope and the computational framework to further improve throughput. The number of cell histories acquired from an image series is determined in the segmentation process.

The cells sometimes make large displacements between two frames. When the cell tracker fails to track the cell, the cell history is terminated. Therefore, the number of cells that the program keeps track of decreases over time. The rate of decay varies considerably between image series, even when acquired under seemingly identical conditions (figure 2b). Only the set of cell histories that completely cover the time from division-to-division enter the analysis (figure 2c).

### (b) Morphology and growth in the microfluidic device

The generation time defines the growth rate of exponentially growing cells and is often used as an indicator of the health or fitness. We compare cells grown with and without amino acids in the medium (figure 3a, red and blue, respectively) and observe average generation times of  $26.4 \pm 7.2$  and  $46.8 \pm 17.0$  min, respectively. Further, we observe an exponential growth of the cell length over the cell cycle (figure 3b). In contrast to previous reports [11], we observe no obvious dependencies of the growth rate on the position the cell occupied in the trap (figure 3c). This uniformity also holds for morphology and bacterial age, i.e. the number of divisions during which the oldest pole of a cell has been observed. We find that the generation times of mother and daughter cells are weakly correlated ( $r = 0.27 \pm 0.02$  with amino acids,  $r = 0.07 \pm 0.05$  without amino acids; figure 3d). The relation between the length at birth and the generation time of a cell history displays a correlation (figure 3e), indicating that comparatively longer newborns complete their cell division faster. Although this holds qualitatively for cells grown both with and without supplemented amino acids (red and blue), it is less pronounced for cells grown without amino acids. Also, cells grown with amino acids vary more in length at birth than in generation time and the opposite is observed



**Figure 3.** Morphology and growth of cells in the microfluidic device. (a) The distribution of generation times of cells grown with (red,  $n = 6755$ ) and without (blue,  $n = 2298$ ) supplemented amino acids. (b) Cell length as a function of relative cell cycle coordinates, i.e. time from birth normalized by the generation time, for randomly selected cells with (red,  $n = 6755$ ) and without (blue,  $n = 2298$ ) supplemented amino acids. (c) Growth rate as indicated by average generation time over the geometry of the trap for cells growing with supplemented amino acids ( $n = 6755$ ). The figure is oriented so that the outlets of the trap are on top and bottom (figure 1b). (d) Joint distribution of generation times for daughter and mother cells with (red,  $n = 6755$ ) and without (blue,  $n = 2298$ ) supplemented amino acids. (e) Joint distribution of generation time and cell length at birth for cells with (red,  $n = 6755$ ) and without (blue,  $n = 2298$ ) supplemented amino acids.

for cells grown without supplemented amino acids. The correlation for cells with amino acids is  $r = -0.43 \pm 0.02$  and without amino acids  $r = -0.28 \pm 0.04$ . No significant differences in growth or morphology between strains SX701 and JE116 are observed.

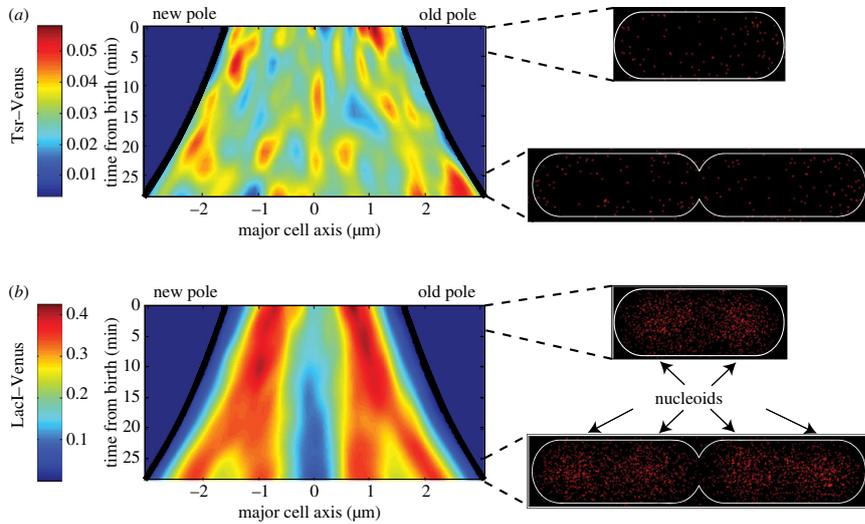
### (c) Localization of transcription factors during the cell cycle

In figure 4, we compare the intracellular localization of the reporter constructs, Tsr-Venus (figure 4a) and LacI-Venus (figure 4b), over the cell cycle. A localization distribution function (figure 4, left) is constructed by mapping the detected molecules to their position along the major axis of the cell ( $x$ -axis) at the time in the cell cycle they were detected ( $y$ -axis) and smoothed using a Gaussian filter. To increase synchronicity, only observations occurring in cells with generation times between 25 and 32 min and terminal lengths of 4–7  $\mu\text{m}$  are included (780 for SX701 and 1176 for JE116). In the right of figure 4, we visualize the detected molecules of each construct as bi-variate symmetric Gaussian functions to create a PALM style super-resolution plot of the intracellular distribution. We do not observe the typical polar

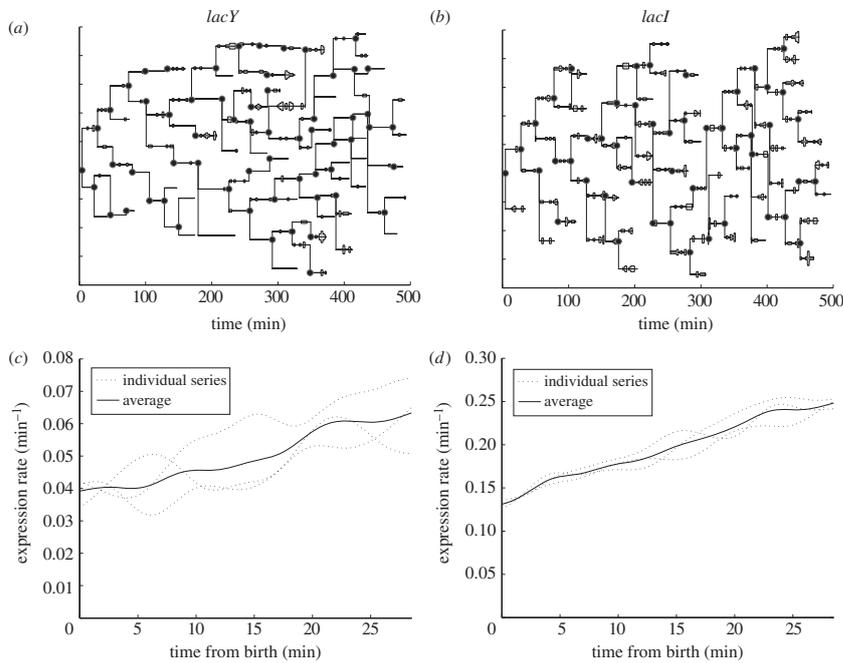
localization that may be expected for Tsr (figure 4a). This is most likely because the protein is inserted at random positions in the membrane and bleaches before reaching the Tsr clusters in the polar regions [4]. For LacI-Venus molecules (figure 4b), we observe a tendency to cluster at positions corresponding to the nucleoids of chromosomal DNA. The number of nucleoids doubles from two, early in the cell history, to four in the later stages, which is consistent with expectations for our growth conditions.

### (d) Synthesis dynamics of an auto-repressed transcription factor throughout the cell cycle

Figure 5 shows lineage trees of cell histories stemming from a single ancestral root of strain SX701 (figure 5a) and JE116 (figure 5b) with bars corresponding to the number of Tsr-Venus and LacI-Venus molecules at the times they were synthesized. The trees are pruned as cells are lost from the segmentation and/or from the trap. For Tsr-Venus expressed from the *lacY* gene, we observe  $1.5 \pm 0.1$  molecules per expression event and  $1.7 \pm 0.1$  events per cell cycle. For LacI-Venus,  $2.2 \pm 0.05$  molecules per expression event and  $2.5 \pm 0.04$  events per cell cycle are observed. The average expression



**Figure 4.** Intracellular localization of molecules over the cell cycle. On the left are the observed molecule densities projected onto the major axis of the cell as a function of time from birth to division for (a) Tsr-Venus and (b) LacI-Venus. The units are the average number of molecules per minute and  $\mu\text{m}$ . The black line at the edge indicates boundary of cell at the apex of the cell poles and expands as the cell grows. On the right is shown the localization of individual (a) Tsr-Venus and (b) LacI-Venus molecules with super-resolution accuracy for early and late stages of the cell cycle.



**Figure 5.** The rate of gene expression over the cell cycle. (a,b) Representative lineage trees for strains (a) SX701 and (b) JE116 stemming from one ancestral root. The absolute number of newly synthesized molecules expressed from the *lacY* and *lacI* gene are indicated as grey bars at the time they are detected. (c,d) The average expression rates from the (c) *lacY* and (d) *lacI* genes over the cell cycle. Solid lines show the average of all three series for each construct. The average of the individual series is shown as dotted lines as an indication of the uncertainty in determining the mean.

rates of Tsr–Venus and of LacI–Venus molecules over the cell cycle are shown in Figure 5c,d. Both show relatively large statistical errors, especially Tsr–Venus. The cell histories with generation time 25–32 min and terminal length 4–7  $\mu\text{m}$  are used. For strain JE116, 1418 complete cell histories and 7910 LacI–Venus molecules are observed. For strain SX701, 780 cell histories are retained from the experiment and 1176 Tsr–Venus molecules. The combination of fewer cell histories and lower expression levels leads to larger statistical uncertainty in determining the expression rate of Tsr–Venus from the *lacY* gene. However, our results indicate a greater expression rate of LacI–Venus at the beginning of the cell cycle.

#### 4. Discussion

In this study we report on a method combining microfluidics, time-lapsed single-molecule microscopy and automated image analysis capable of monitoring the growth and absolute number of gene expression events throughout approximately 3000 complete individual *E. coli* life-spans per experiment. Further, we demonstrate that it is possible to use a functional transcription factor, LacI–Venus, non-specifically interacting with DNA, to retrieve information on both expression dynamics and super-resolution localization dynamics throughout the cell cycle. We show that the microfluidic chip provides a beneficial and stable environment for exponentially growing *E. coli* cells and a high degree of control and reproducibility. We observe a significant variability in generation times of individual cells. However, we find that generation time is relatively memory-less from generation to generation. More interestingly, cells living in richer media vary more in length at birth than in generation time and that the opposite is true for cells living in poorer media. The underlying causes for this *size–generation time uncertainty relation* and for which range of conditions it holds are presently unclear. LacI–Venus molecules localize onto the nucleoids in the cell. It appears that non-specifically interacting transcription factors are uniformly distributed over the DNA. As expected, we find that LacI–Venus is more highly expressed than Tsr–Venus from the *lacY* gene position. Our result for the latter is consistent with the findings of Yu *et al.* [3] in the number of gene expression events from the *lacZ* gene position during the cell cycle. However, we observe fewer Tsr–Venus

molecules per expression event ( $1.7 \pm 0.1$  instead of  $4.2 \pm 0.5$ ). Given that *lacZ* and *lacY* are transcribed to a polycistronic mRNA, we conclude that the translation rate at the *lacY* position is two to three fold lower than that of the *lacZ* position. The average rate of LacI–Venus expression is slightly higher in the beginning of the cell cycle. We propose that this may be due to partition inequalities at cell division, in which disfavoured cells replenish their transcription factor pools. The experiments confirm the highly variable nature of *in vivo* single-molecule observations (figure 5). We estimate that to obtain a 5 per cent accuracy of the mean expression rate per minute for all points in the cell cycle would require to a total of 4000 and 16000 complete cell histories of JE116 (LacI–Venus) and SX701 ( $\Delta\text{lacY}::\text{Tsr–Venus}$ ), respectively. Sufficient observations could therefore be obtained with three additional experiments for JE116 and fifteen additional experiments for SX701. The Mather design can potentially sustain a population of bacterial cells in a state of exponential growth indefinitely. Many biological phenomena, such as the development of antibiotic resistance, occur in a small subpopulation of all cells and on longer time-scales than the current longevity of an experiment using our method. Further increasing the throughput and the longevity of the method to enable the study of such phenomena represents a formidable image analysis challenge. However, to our advantage is that the performance of supervised algorithms improves and can be made more advanced as more training data accumulate. We are confident that more advanced algorithms can be implemented to increase both accuracy and speed, which would make it possible to acquire an arbitrary number of cell histories from a single experiment.

The authors thank Jeff Hasty and the members of his laboratory for guidance in establishing the microfluidic protocols. We specifically thank William Mather and Octavio Mondragon-Palomino and colleagues for designing and characterizing the functions of the trap motif used in this study. We also want to thank Christine Jacobs-Wagner and Oleksii Sliusarenko for providing the original version of MICROBETRACKER. The SX701 strain was kindly donated from the laboratory of Sunny Xie. Further, we thank Fredrik Persson for helpful discussions on photonics and dot-detection. We are also grateful for the efforts of Intakhar Ahmad for validation work on the analysis protocols. This study was funded by grants from the European Research Council, the Swedish Research Council, Knut and Alice Wallenberg's stiftelse and Görans Gustafssons stiftelse. A.M. was funded by the Center for Interdisciplinary Mathematics (CIM), Uppsala University. The paper was accepted for publication July 16 2012.

#### References

1. Elowitz MB, Levine AJ, Siggia ED, Swain PS. 2002 Stochastic gene expression in a single cell. *Science* **297**, 1183–1186. (doi:10.1126/science.1070919)
2. Xie XS, Choi PJ, Li GW, Lee NK, Lia G. 2008 Single-molecule approach to molecular biology in living bacterial cells. *Annu. Rev. Biophys.* **37**, 417–444. (doi:10.1146/annurev.biophys.37.092607.174640)
3. Yu J, Xiao J, Ren X, Lao K, Xie XS. 2006 Probing gene expression in live cells, one protein molecule at a time. *Science* **311**, 1600–1603. (doi:10.1126/science.1119623)
4. Nagai T, Ibata K, Park ES, Kubota M, Mikoshiba K, Miyawaki A. 2002 A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* **20**, 87–90. (doi:10.1038/nbt0102-87)
5. Elf J, Li GW, Xie XS. 2007 Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191–1194. (doi:10.1126/science.1141967)
6. Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, Emili A, Xie XS. 2010 Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538. (doi:10.1126/science.1188308)
7. Wang P, Robert L, Pelletier J, Dang WL, Taddei F, Wright A, Jun S. 2010 Robust growth of *Escherichia coli*. *Curr. Biol.* **20**, 1099–1103. (doi:10.1016/j.cub.2010.04.045)
8. Danino T, Mondragon-Palomino O, Tsimring L, Hasty J. 2010 A synchronized quorum of genetic clocks. *Nature* **463**, 326–330. (doi:10.1038/nature08753)
9. Young JW, Locke JCW, Altinok A, Rosenfeld N, Bacarian T, Swain PS, Mjolsness E, Elowitz MB. 2012 Measuring single-cell gene expression

- dynamics in bacteria using fluorescence time-lapse microscopy. *Nat. Protoc.* **7**, 80–88. (doi:10.1038/nprot.2011.432)
10. Sliusarenko O, Heinritz J, Emonet T, Jacobs-Wagner C. 2011 High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. *Mol. Microbiol.* **80**, 612–627. (doi:10.1111/j.1365-2958.2011.07579.x)
  11. Mather W, Mondragon-Palomino O, Danino T, Hasty J, Tsimring LS. 2010 Streaming instability in growing cell populations. *Phys. Rev. Lett.* **104**, 208101. (doi:10.1103/PhysRevLett.104.208101)
  12. Datsenko KA, Wanner BL. 2000 One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645. (doi:10.1073/pnas.120163297)
  13. Choi PJ, Cai L, Frieda K, Xie XS. 2008 A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* **322**, 442–446. (doi:10.1126/science.1161427).
  14. Hammar P, Leroy P, Mahmutovic A, Marklund EG, Berg OG, Elf J. 2012 The lac repressor displays facilitated diffusion in living cells. *Science* **336**, 1595–1598. (doi:10.1126/science.1221648)
  15. Kass M, Witkin A, Terzopoulos D. 1988 Snakes: active contour models. *Int. J. Comput. Vis.* **1**, 321–331. (doi:10.1007/bf00133570)
  16. Bishop CM. 2006 *Pattern recognition and machine learning*. Berlin, Germany: Springer.
  17. Ronneberger O, Raffel M, Kompenhans J. 1998 Advanced evaluation algorithms for standard and dual plane particle image velocimetry. *Proc. 9th Int. Symp. Applications of Laser Techniques to Fluid Mechanics, Lisbon, Portugal, 13–16 July 1998*, paper 10.1.
  18. Moré J. 1978 The Levenberg–Marquardt algorithm: implementation and theory, chapter 10. In *Numerical Analysis* (ed. GA Watson), Lecture Notes in Mathematics, pp. 105–116. Berlin, Germany: Springer.



## Supplementary methods

### Cropping and Aligning Images

In raw phase contrast images, the trap only covers a fraction of the image. In the images used for the segmentation and tracking, the images were pre-processed in order to extract the trap. This was done by correlating a binary image containing a box of a size corresponding to the chip in the raw phase image. The size of the cut-out was reduced from 860x860 pixels to 500x500 pixels in order to increase speed of our algorithms. The cut-out from the phase image was correlated with the corresponding bright field image taken in the fluorescence channel in order to get an alignment with the precision of approximately one pixel. Vertical interference patterns of light that appears in the raw phase images also needed removal. This is done by calculating a background image averaged in time as well as in the horizontal direction. The background image is subtracted from each individual phase image.

### Division Function

The intensity landscape of a phase contrast image closely resembles a saddle node around the point where the division occurs, where a local maximum is found in the direction of the major axis and local minima in the direction of the minor axis. The original division function of MicrobeTracker classified division events solely by the magnitude of a local maxima relative its surroundings in the image. This method often mistakenly identifies unrelated intensity variations in densely growing *E. coli* as division events. Consequently the function was further developed to increase robustness and accuracy. The new division function applies a linear discriminant (described under “Classification”) to a space spanned by five features in order to classify a potential division event as either divided or non-divided. Test positions along the cell profile are identified using the original division function [1] and around each position a reference coordinate system of 11x11 points is established. The coordinate system uses an orthonormal base where the first direction corresponds to the direction between the cell poles and the second direction corresponds to the cell width. The distance between each point was one pixel, and a typical cell width in our images was 15 pixels. A test image was calculated from this coordinate system using bi-linear 2D interpolation. The features extracted from these images are the relative magnitude of the local maxima as in the original division function, the two Eigen-values to the Hessian matrix and the scalar products of the test image with two templates of a true division event and a false division event. The templates were constructed prior to analysis by averaging corresponding 11x 11 cut-outs from manually assembled training sets consisting of 82 true divisions and 259 falsely identified division events respectively.

### **Pole Searcher**

Cells continuously shift their position between the previous and the current frame. It was noticed that accurately identifying the cell poles was the most critical factor when tracking a cell and for this reason a method based on tracking the poles between frames was developed. The features of the pole searcher are the 20 principal components of the pixel intensities in a training set of 52370 manually segmented 15x15 pixel cell pole examples and an equal number of examples at random positions. A linear Bayesian probabilistic model was created from the training data (see Classification). When searching for the poles the method samples the surroundings of the previous poles stochastically according to a 4D Gaussian distribution for the spatial coordinates, cell length and an orientation angle.

This sampling and testing is done pairwise for both poles. The expectation value for the cell length is estimated using the Euler forward method and a finite difference approximation of the derivative. The expectation value of the movement of the entire cell is estimated by the optical flow method [2]. The expectation value of the orientation angle is set to the angle in the previous image. The standard deviations of these distributions are estimated from the training data. Test points are sampled from this 4D distribution and for each test point, a pair of 15x15 reference coordinate systems are established, corresponding to guesses of where the cell poles are localised in the current image. Sample images are calculated from the reference coordinate systems using 2D bi-linear interpolation. The probabilistic model described in the "Classification" section is used for calculating a probability for each pair of sample images to be correctly aligned with the cell poles. The weighted mean of all samples is used as an estimate of the new pole positions, length and angle.

### **First Error Detector**

Segmentation errors will propagate throughout the series and also cause neighbouring cells to be erroneously segmented. Therefore, it was necessary to develop methods of detecting errors as they were made by the segmentation algorithm. Different indicators of errors were used to create features for a classifier to detect incorrectly segmented cells. For training, 135 correctly segmented examples and 41 incorrectly segmented examples were used. The features used were 1) the relative difference in pixel intensity inside the cell contour, 2) the overlap between cells relative to the cell area, 3) cell movement between the current and previous frame and 4) angular movement between the current and previous frame.

### **Cell Tracker**

In order to increase the number of cell generations a cell tracker was developed in order to save cells lost due to the first error detector. A coordinate system of 44x17 points was created in order to describe a cell in a standardised framework, with equal number of points independent of cell length. The first and last 8x17 points of the coordinate system were used for describing the cell poles. The

intermediate 28x17 points cover the remaining cell with 28 equidistant ribs along the cell profile. The cell was tracked in a given frame searching for the minimal Mahalanobis distance,  $d$ , defined by

$$d^2 = (x_1 - x_2)^T \Sigma^{-1} (x_1 - x_2), \quad (1)$$

where  $x_1$  is the cell in the previous frame and  $x_2$  in the current frame.  $\Sigma$  is the covariance matrix calculated from 13880 training examples extracted in the neighbourhood of correctly segmented cells. The dimensionality of the covariance matrix was reduced by projecting the training data to the 40 first principal components. The dimension reduction was done for the purpose of removing noise in the covariance matrix due to the limited number of training examples. The search was performed by sampling from a 3D Gaussian distribution of positions and angles. The cell length was kept fixed as the same as in previous frame. The search space was reduced by not allowing a relative overlap with neighbouring cells of more than 25% of the cell area.

### Second Error Detector

Also the cell tracker occasionally makes errors. A second error detector was therefore developed in order to prevent these errors from propagating. It uses a set of five features that are different from those used in the first error detector. These features are: 1) the relative difference in the number of pixels above the threshold given by Otsu's method [3] inside the cell contour. 2) The sum of the cell profile's absolute deviation from the straight line that goes through both cell poles. This feature was used as a measure of overall curvature. 3) The difference in the number of edge pixels inside the cell profile. The edge pixels were defined as pixels having the sum of second derivatives in the image, in x and y directions, above a pre-defined threshold. 4) Differences in cell lengths between the two frames. 5) The cross correlation of the cell images in the current and the previous frame. The cell images were interpolated from the previously mentioned 44x17 points cell coordinate system.

### Classification

For all algorithms described above except the division function a Bayesian probabilistic model was used for classification between two classes denoted  $C_1$  and  $C_2$ . For example, the classes may denote whether a sample image is on the pole or off the pole. The probabilistic model can be derived from Bayes theorem using the assumption that the features for both classes have a multivariate normal distribution [4]. The probability of the input data  $\mathbf{x}$  belonging to class  $C_1$  is given by

$$p(C_1|\mathbf{x}) = \sigma(\mathbf{w}^T \mathbf{x} + w_0) \quad (2)$$

The vector  $\mathbf{w}$ , sometimes referred to as the linear discriminant, is calculated as

$$\mathbf{w} = \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \quad (3)$$

Where  $\Sigma$  is the shared within class covariance matrix of class 1 and 2,  $\boldsymbol{\mu}_1$  and  $\boldsymbol{\mu}_2$  are the mean feature vectors of class 1 and 2 respectively. Since class 1 and 2 may have different number of training samples, the shared covariance matrix is calculated as the maximum likelihood solution

$$\Sigma = \frac{N_1}{N} \Sigma_1 + \frac{N_2}{N} \Sigma_2 \quad (4)$$

Where  $\Sigma_1$  and  $\Sigma_2$  are covariance matrixes of class 1 and 2 respectively.  $N_1$  and  $N_2$  are the number of training samples from class 1 and 2 respectively.  $N$  is the total number of samples. The number of features must be significantly smaller than the number of training samples in order to get a linear discriminant with high statistical precision. If the number of features is larger than the number of training samples, the covariance matrix will not be invertible. The function  $\sigma$ , referred to as the sigmoid function, is defined as

$$\sigma(a) = \frac{1}{1 + \exp(-a)} \quad (5)$$

The quantity  $w_0$  is calculated as

$$w_0 = -\frac{1}{2} \boldsymbol{\mu}_1^T \Sigma^{-1} \boldsymbol{\mu}_1 + \frac{1}{2} \boldsymbol{\mu}_2^T \Sigma^{-1} \boldsymbol{\mu}_2 + \ln \frac{p(C_1)}{p(C_2)} \quad (6)$$

The prior probabilities are calculated using the maximum likelihood solution  $p(C_1) = \frac{N_1}{N}$ . For the division function, the linear discriminant was used without a probabilistic framework. In this case the scalar product  $\mathbf{w}^T \mathbf{x}$  is calculated and compared to a threshold to check whether the cell is divided or not.

### Parameter Estimation for Maximum likelihood Estimate of Synthesis

Two parameters,  $p$  and  $\lambda$ , were used in the Maximum likelihood method used for estimating the number of new spots. The parameter  $p$  was estimated by the fraction of dots that are lost by taking one fluorescence image rapidly after another. If there are  $k$  extra bleaching images in between the images where dots are counted  $p = 1 - (1 - p_1)^{k+1}$ , where  $p_1$  is the fractional loss of dots per bleaching frame.  $\lambda$  was estimated by the number of newly synthesised molecules per generation divided by the number of frames per generation where the dots are counted. The number of newly synthesised molecules per generation was estimated by counting molecules per cell in a sample where fluorescence images are

taken much more rarely than the generation time. It is equal to the average number of molecules in a newly divided cell.

### Spot Quality Estimation

The positions of the spots were estimated with the method by Ronneberger *et al* [5]. This method is briefly described in the main article. However, the spots vary in shape and quality and we only want to take into account the most significant spots. Therefore, an objective quality criterion was needed in order to reject spots with a quality below a certain threshold. This was done with linear regression according to the model

$$Y = \mathbf{1}\beta_1 + \mathbf{X}\beta_2, \quad (7)$$

where  $\mathbf{Y}$  is the 11x11 pixel cut out from the fluorescence image at the position of the spot,  $\mathbf{1}$  is a matrix of the same size with each matrix element equal to 1 and  $\mathbf{X}$  is a discretized 2D Gaussian function. The 2D Gaussian function has a mean estimated with the method by Ronneberger *et al* and a standard deviation corresponding to the point spread function. In general, the mean of the 2D Gaussian does not coincide with the central pixel in the 11x11 coordinate system. In the linear regression model, the coefficient  $\beta_1$  corresponds to background fluorescence and  $\beta_2$  to the intensity of the Gaussian spot. The  $Z$ -score was calculated for the  $\beta_2$  coefficient and we test that  $\beta_2 > 0$ . In this work, spots with a  $Z$ -score below 6 were rejected.

### Reduction of Memory Footprint

As the size of datasets that can be processed with MicrobeTracker increases, so does the memory footprint. This eventually becomes a critical issue when the managing of cell data starts to hamper performance. Originally, the cell list that contains all data of individual cells, including their contours, was stored as an array (a “cell array” in MATLAB terminology) where the number of elements for a particular frame grew exponentially with the number of frames. A new API was developed to store the cell list in a more compact way that grows linearly with the number of frames. To further reduce the memory usage by a factor of approximately 0.7 all floating-point data were stored in single precision, with no significant impact on the accuracy of the algorithms. Conversion from the old to the new format is automatically done when data is loaded, making our version of MicrobeTracker backwards compatible. The API for the new format was spliced off from the main code so that it may be used by external programs that uses the cell lists.

### **Stubs and Iterative Segmentation**

To progress the transition towards fully automated segmentation the concept of *stubs* were introduced and implemented in MicrobeTracker. Whenever a cell is manipulated such that its data is invalidated on subsequent frames, e.g., by manually forcing it to split or by refining its contour, it is turned into a stub. Stubs are easily distinguished in the GUI and are exempt from further automatic segmentation unless MicrobeTracker is explicitly told to process stubs. This enables automated bookkeeping of cells with invalidated downstream data stemming from manual editing, but the stubs API was also integrated with the error detectors to exempt cells that the latter identify as erroneously segmented from further processing; i.e. cells caught in the error detector turn into stubs. The present design of the first error detector renders false positives that wrongly exempt cells from processing at subsequent frames. To compensate for that shortcoming all newly formed stubs are optionally and automatically segmented one or several times to see if a sensible cell model can be constructed. A single iteration has proven to take care of most false positives, extending the expected number of frames that a cell can be automatically segmented before MicrobeTracker loses track of it. It should be emphasized that this effect will be reduced or eliminated if the error detection is further improved such that it gives less false positives. For this work, the manual effort was only done in the initial frame.

## References

- 1 Sliusarenko, O., Heinritz, J., Emonet, T. & Jacobs-Wagner, C. 2011 High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. *Molecular Microbiology*, **80**, 612-627, (doi:10.1111/j.1365-2958.2011.07579.x).
- 2 Horn, B. K. P. & Schunck, B. G. 1981 Determining optical flow. *Artificial Intelligence*, **17**, 185-203, (doi:10.1016/0004-3702(81)90024-2).
- 3 Otsu, N. 1979 A threshold selection method from grey level histograms. *IEEE Trans. Syst. Man Cybernet*, **SMC-9**, 62-66.
- 4 Bishop, C. M., 2006 *Pattern Recognition and Machine Learning*.(Springer).
- 5 Ronneberger, O., Raffel, M. & Kompenhans, J. (1998).



Paper III 



# Direct measurement of transcription factor dissociation excludes a simple operator occupancy model for gene regulation

Petter Hammar<sup>1,2</sup>, Mats Walldén<sup>1,2</sup>, David Fange<sup>1</sup>, Fredrik Persson<sup>1</sup>, Özden Baltekin<sup>1</sup>, Gustaf Ullman<sup>1</sup>, Prune Leroy<sup>1</sup> & Johan Elf<sup>1</sup>

Transcription factors mediate gene regulation by site-specific binding to chromosomal operators. It is commonly assumed that the level of repression is determined solely by the equilibrium binding of a repressor to its operator. However, this assumption has not been possible to test in living cells. Here we have developed a single-molecule chase assay to measure how long an individual transcription factor molecule remains bound at a specific chromosomal operator site. We find that the *lac* repressor dimer stays bound on average 5 min at the native *lac* operator in *Escherichia coli* and that a stronger operator results in a slower dissociation rate but a similar association rate. Our findings do not support the simple equilibrium model. The discrepancy with this model can, for example, be accounted for by considering that transcription initiation drives the system out of equilibrium. Such effects need to be considered when predicting gene activity from transcription factor binding strengths.

Transcription factors are the major regulators of gene expression. Transcription factor-based regulation of transcription initiation is often described by a simple operator occupancy model, where in the case of repressors it is assumed that transcription is 'off' when the repressor is bound and 'on' when the promoter is free<sup>1,2</sup>. In this scenario, the resulting ratio of expression levels with and without repressor, i.e., the repression ratio (RR), becomes

$$RR = \frac{\tau_{\text{on}} + \tau_{\text{off}}}{\tau_{\text{on}}} \quad (1)$$

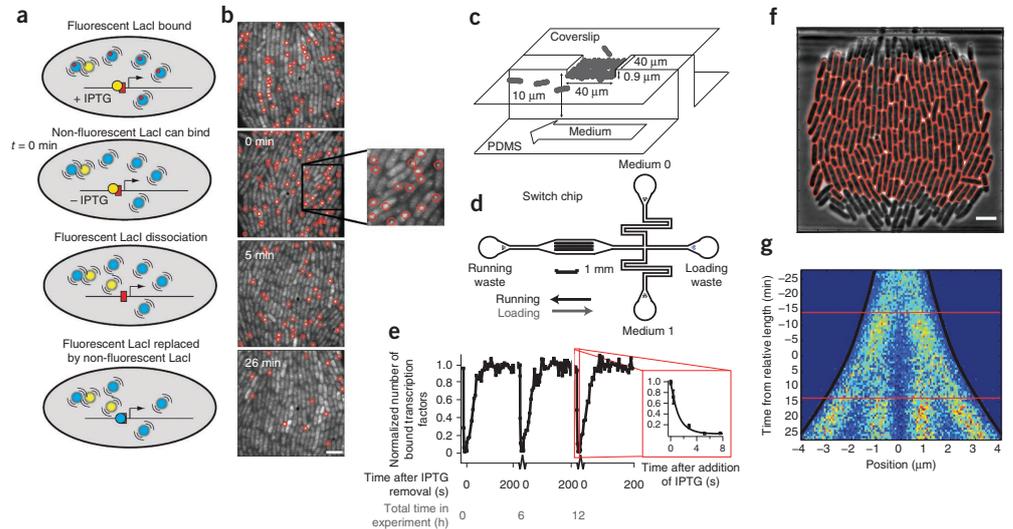
where  $\tau_{\text{off}}$  is the average time the repressor is bound and  $\tau_{\text{on}}$  is the average time the promoter is free (Supplementary Note). The repression ratio is high when the repressor is bound for a long time (large  $\tau_{\text{off}}$ ) or when the repressor concentration is high, which leads to fast binding (small  $\tau_{\text{on}}$ ). This simple equation has a central position in quantitative biology as it relates the state of the cell, i.e., transcription factor concentrations, to change in state, i.e., gene expression.

The equation is therefore used in most synthetic and systems biology studies although the underlying assumptions have not been tested in living cells, where cooperative binding, active transcription, DNA replication and chromosome dynamics could influence gene regulation.

The challenge of testing the operator occupancy model in living cells is to measure the rates of operator association,  $\tau_{\text{on}}^{-1}$  and dissociation,  $\tau_{\text{off}}^{-1}$ , directly in live cells rather than inferring them from reporter expression assays<sup>3,4</sup>. Recently, we developed a direct single-molecule microscopy assay to measure the rate of binding to a single *lac* operator site in the bacterial chromosome<sup>5</sup>. Here we present an *in vivo* version of a biochemical chase assay<sup>6</sup>, which enables direct measurements of spontaneous dissociation of the *lac* repressor protein, LacI, from individual chromosomal operator sites (Fig. 1a,b). In our assay, operator-bound fluorescent LacI-YFP dimers that spontaneously dissociate are replaced (chased) by non-fluorescent LacI tetramers. Non-fluorescent LacI molecules are present in excess (Supplementary Fig. 1a) and prevent rebinding of fluorescent LacI. The spontaneous dissociation process can thus be followed by counting the average number of bound fluorescent molecules per cell over time. To start the experiment with the fluorescent LacI bound, a point mutation has been introduced into the fluorescent LacI such that it cannot bind the inducer isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG)<sup>7</sup>. The presence of IPTG prevents binding of the non-fluorescent LacI until IPTG is removed at the start of the experiment (Supplementary Fig. 1b,c). To ensure that dissociation kinetics were independent of IPTG outflux, we showed that the intracellular concentration of IPTG within 1 min of its removal dropped to a level where non-fluorescent LacI bound effectively (Supplementary Fig. 2 and Supplementary Note), and we subsequently analyzed dissociation kinetics beginning at 1.5 min after the removal of IPTG. An extended analysis of how the finite concentrations of non-fluorescent LacI influenced the results is provided in the Online Methods. The model for replication-induced LacI dissociation is extended in the Supplementary Note. The kinetic assays were performed on *E. coli* cells residing in a microfluidic growth chamber (Fig. 1c,d), which allowed the cells to

<sup>1</sup>Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden. <sup>2</sup>These authors contributed equally to this work. Correspondence should be addressed to J.E. (johan.elf@icm.uu.se).

Received 21 July 2013; accepted 31 January 2014; published online 23 February 2014; doi:10.1038/ng.2905



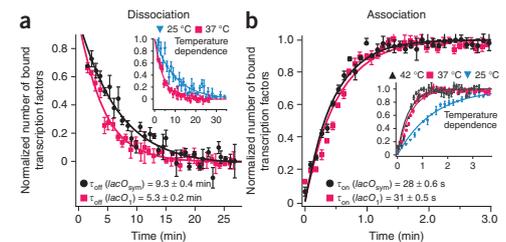
**Figure 1** The single-molecule chase assay. **(a)** Outline of the single-molecule chase assay. When fluorescent LacI dimers (yellow) dissociate from the *lac* operator (red box), they are replaced by non-fluorescent wild-type LacI tetramers (blue) present in excess. **(b)** Examples of fluorescence images (4-s exposure) taken before and at different time points after the removal of IPTG. Scale bar, 4  $\mu$ m. The inset image is magnified by 2 $\times$  relative to the original image. Red circles indicate detected operator-bound LacI-YFP. **(c,d)** The microfluidic switching chip **(d)** contains 51 traps as illustrated **(c)**. Each trap harbors  $\sim$ 250 *E. coli* cells and allows for sustained exponential growth and fast change of medium. **(e)** Medium switch-induced transcription factor dissociation and association. When medium is switched from high 2-nitrophenyl  $\beta$ -D-fucopyranoside (ONPF; anti-inducer) to high IPTG (inducer), transcription factors dissociate in a few seconds (inset). When medium is switched back, transcription factors associate in  $\sim$ 30 s. The graph shows three switching cycles separated by 6-h recovery periods. **(f)** Automatically segmented cells using a phase-contrast image. Scale bar, 4  $\mu$ m. **(g)** Intracellular positions of bound LacI-YFP molecules (x axis) mapped to the cell replication cycle (y axis). Individual cell replication cycles are synchronized so that the time of 0 min always implies a cell length of 4.25  $\mu$ m. Horizontal lines mark the average times for cell divisions.

be maintained in a constant state of exponential growth (generation time of 26 min)<sup>8</sup> as well as allowing rapid medium exchange (in 2 s). Image acquisition and medium exchange were automated and synchronized so that the experiment was repeatable with high precision (Fig. 1e). Cell segmentation and detection of fluorescent spots were also automated and enabled the mapping of individual molecules onto an intracellular coordinate system for an arbitrary number of cells (Fig. 1f). For example, Figure 1g (as well as Supplementary Fig. 3) shows the probability distribution of the intracellular location of specifically bound LacI-YFP molecules as a function of position in the cell cycle.

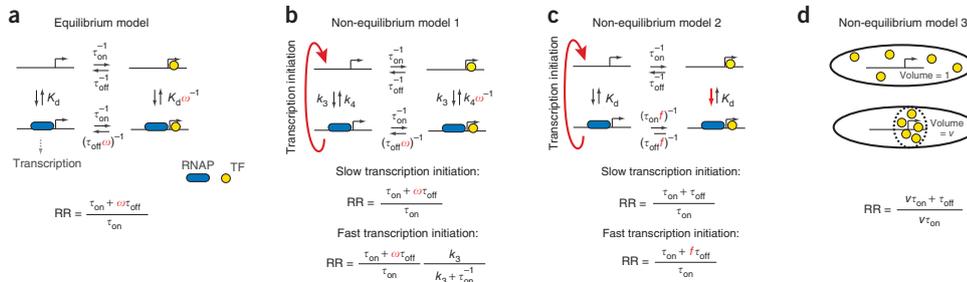
We used the *in vivo* chase assay to measure the kinetics for two operators of different strength, the natural *lacO*<sub>1</sub> operator and the stronger, symmetric artificial *lacO*<sub>sym</sub> operator. The dissociation curves for the LacI-YFP dimer from the *lacO*<sub>1</sub> and *lacO*<sub>sym</sub> operators at 37 °C are shown in Figure 2a. The average time LacI stayed bound to its operator ( $\tau_{off}$ ) was 5.3  $\pm$  0.2 (s.e.m.) min for *lacO*<sub>1</sub> and 9.3  $\pm$  0.4 (s.e.m.) min for *lacO*<sub>sym</sub>. The average time before the operator was bound by a repressor ( $\tau_{on}$ ) was measured under identical experimental conditions (Fig. 2b) and was 30.9  $\pm$  0.5 (s.e.m.) s for *lacO*<sub>1</sub> and 27.6  $\pm$  0.6 (s.e.m.) s for *lacO*<sub>sym</sub>. Thus, a stronger operator has a slower dissociation rate but a similar association rate.

We were then ready to ask whether the measured association and dissociation times could be used to predict the repression ratio using the simple operator occupancy model, i.e., equation (1), as given by the model in Figure 3a without any cooperative interaction between LacI

and RNA polymerase (RNAP) ( $\omega = 1$ , as defined in Fig. 3a and equations (3) and (4) in the Online Methods). Combining the association and dissociation measurements, we calculated that the repression ratio was expected to be 11.2  $\pm$  0.5 (s.e.m.) for *lacO*<sub>1</sub> and 21.2  $\pm$  0.9



**Figure 2** Kinetic measurements for individual *lac* operators. **(a)** Dissociation curves for *lacO*<sub>sym</sub> and *lacO*<sub>1</sub>.  $n = (i, j, k)$  implies  $i$  repetitions 6 h apart for chip 1,  $j$  repetitions for chip 2 and  $k$  repetitions for chip 3. Error bars,  $\pm$  s.e.m.;  $n = (2, 3, 2)$  (*lacO*<sub>1</sub>) and  $n = (3, 2, 3)$  (*lacO*<sub>sym</sub>). Inset, temperature dependence for dissociation from *lacO*<sub>1</sub>. Error bars,  $\pm$  s.e.m.;  $n = (2, 3, 2)$  (37 °C) and  $n = (2, 2)$  (25 °C). **(b)** Association curves for *lacO*<sub>sym</sub> and *lacO*<sub>1</sub>. Error bars,  $\pm$  s.e.m.;  $n = (2, 3, 2)$  (*lacO*<sub>1</sub>) and  $n = (3, 2)$  (*lacO*<sub>sym</sub>). Inset, temperature dependence for association with *lacO*<sub>1</sub>. Error bars,  $\pm$  s.e.m.;  $n = (2, 3, 2)$  (37 °C) and  $n = (3)$  (25 and 42 °C).



**Figure 3** Models of gene regulation. **(a)** At equilibrium, the repression ratio only depends on the fraction of time the operator is bound independent of kinetic schemes. Owing to cooperative binding ( $\omega > 1$ ), the ratio can be modulated by other factors. TF, transcription factor;  $K_d$ , equilibrium binding constant. **(b)** Transcription initiation can drive the system out of equilibrium such that the repression ratio depends on the rate of transcription initiation. **(c)** The transcription factor binds and dissociates slower when RNAP is bound. Transcription drives the system out of equilibrium such that the transcription factor associates at naked DNA and dissociates at RNAP-bound DNA. **(d)** When the transcription factors are maintained in a reduced volume,  $v$ , transcription factor association rates are in the simplest case increased by the corresponding factor.

(s.e.m.) for  $lacO_{sym}$  (Table 1). The corresponding measurements of the repression ratios for the LacI-Venus dimer based on an enzymatic reporter assay were  $10.0 \pm 1.3$  (s.e.m.) for  $lacO_1$  and  $29.7 \pm 3.4$  (s.e.m.) for  $lacO_{sym}$  (Table 1). We conclude that the operator occupancy model accounts for the repression ratio for  $lacO_1$  but not for the ratio for  $lacO_{sym}$ , where the observed repression ratio was higher than expected when considering association and dissociation rates alone.

This discrepancy for  $lacO_{sym}$  motivated the construction of more complex interaction models. One possibility was an equilibrium model where LacI interacts cooperatively with RNAP or another protein binding near the operator and where the degree of cooperativity depends on the operator sequence. This model is represented (Fig. 3a) using  $\omega = 1.5$  and  $\omega = 1$  for  $lacO_{sym}$  and  $lacO_1$ , respectively, and resulted in excellent agreement with the measured repression ratios. Such a difference in cooperativity between  $lacO_1$  and  $lacO_{sym}$  could be due to the markedly different bending of DNA when LacI is bound to the different operators<sup>8,9</sup>. Operator sequence-specific interactions between LacI and RNAP have previously been suggested when the operator is positioned upstream of the *placUV5* promoter<sup>10</sup>. Although this equilibrium mechanism is also possible with the operator located downstream of the promoter, a model with operator-specific cooperativity was not needed to describe our data. Cellular reaction dynamics are commonly out of equilibrium, and we therefore also considered more simple non-equilibrium schemes. In Figure 3b–d, we outline three such schemes that can increase the repression ratio beyond the

ratio predicted by the simple operator occupancy model. We discuss them individually below.

The first non-equilibrium scheme (Fig. 3b) is similar to the scheme with cooperative interaction with RNAP (Fig. 3a) except that active transcription initiation clears the promoter in the absence of LacI. Slow transcription initiation leads to a repression ratio as in the cooperative equilibrium model, whereas fast transcription initiation leads to a reduced repression ratio, as it is possible to synthesize transcripts before the repressor has equilibrated with DNA. Interestingly, we found that the transcription rate for the *lac* operon with full induction was  $5.4 \pm 0.5$  (s.e.m.) times higher in the strain with the  $lacO_1$  sequence than in the strain with the  $lacO_{sym}$  sequence next to the promoter (Supplementary Note). This difference in transcription rate, in combination with the measured association and dissociation rates, is sufficient to fully account for the measured repression ratios when  $\omega = 1.5$  for both  $lacO_1$  and  $lacO_{sym}$ . The reason for this is that  $lacO_{sym}$  is closer to the equilibrium case (with slow transcription) described above, whereas  $lacO_1$  is out of equilibrium (with fast transcription) and thus has a lower repression ratio than what is expected from the equilibrium model alone (Fig. 3b). As a consequence, no operator sequence-dependent interaction between LacI and RNAP is needed in this case, as the sequences are transcribed at different rates.

Also in the second non-equilibrium scheme (Fig. 3c), transcription initiation drives the system out of equilibrium but this time without any cooperative binding between RNAP and LacI. In this scheme, RNAP binds to one of the alternative *lac* promoters next to the operator-bound LacI but does not continue into open complex formation<sup>11</sup>. In contrast, when RNAP binds in the absence of LacI, it proceeds rapidly and irreversibly into transcription, clearing the promoter. Consequently, LacI will most often bind in an RNAP-free promoter region and dissociate from an RNAP-bound operator region. Thus,

**Table 1** Comparison of repression ratios from reporter expression assays and direct single-molecule *in vivo* measurements

Operator region	Repression ratio		Single-molecule kinetics	
	Reporter expression assay <sup>a</sup>	$\frac{\tau_{on} + \tau_{off}}{\tau_{on}}$	$\tau_{on}$ (s)	$\tau_{off}$ (min)
$lacO_1$	$10.0 \pm 1.3$	$11.2 \pm 0.5$	$30.9 \pm 0.5$	$5.3 \pm 0.2$
$lacO_{sym}$	$29.7 \pm 3.4$	$21.2 \pm 0.9$	$27.6 \pm 0.6$	$9.3 \pm 0.4$

Data are shown as mean values  $\pm$  s.e.m.;  $n$  indicates replicates from individual experiments (reporter expression):  $n = 9$  ( $lacO_1$ ) and  $n = 8$  ( $lacO_{sym}$ ). See Figure 2 for details of the single-molecule experiments.

<sup>a</sup>The repression ratio is induced (+ IPTG) divided by repressed (– IPTG) *lacZ* expression in terms of Miller units (normalized  $\beta$ -galactosidase activity) and is normalized to the lower repressor concentrations in the kinetic experiments (Supplementary Fig. 8 and Supplementary Note).

**Table 2** Binding kinetics dependence on roadblocks

	Binding kinetics		Repression ratio
	$\tau_{on}$ (s)	$\tau_{off}$ (min)	
Without roadblock	$27.6 \pm 0.6$	$9.3 \pm 0.4$	$21.2 \pm 0.9$
With roadblock	$37.1 \pm 0.6$	$11.6 \pm 1.4$	$19.7 \pm 1.1$

Association and dissociation rates measured for LacI-YFP with or without TetR binding next to one side of the operator  $lacO_{sym}$ . Data are shown as mean values  $\pm$  s.e.m.;  $n = \{i, j, k\}$  implies  $i$  repetitions 6 h apart for chip 1,  $j$  repetitions for chip 2 and  $k$  repetitions for chip 3;  $n = \{4, 2\}$  ( $\tau_{on}$ ,  $lacO_{sym}$  with roadblock) and  $n = \{4, 3, 3\}$  ( $\tau_{off}$ ,  $lacO_{sym}$  with roadblock). Data without roadblock are the same as in Table 1.

if a bound RNAP molecule slows down LacI dissociation, this would result in repression beyond that predicted in the equilibrium model, even if the binding strength for LacI is unaltered by the bound RNAP. The average times for LacI association and dissociation are expected to increase by up to a factor of two when a protein is bound next to the *lac* operator, as sliding along DNA in and out of the operator is blocked from one side<sup>5</sup>. To test this hypothesis, we positioned the *tet* repressor protein, TetR, next to the *lac* operator site and measured the times for LacI dissociation and association. We found that the time for association increased by a factor  $f = 1.35 \pm 0.04$  (s.e.m.) when TetR was bound next to *lacO*<sub>sym</sub> and that the effect on dissociation was similar (Table 2 and Supplementary Fig. 4), as was expected from detailed balance when steady-state binding is not altered. The effect was smaller ( $f = 1.16 \pm 0.03$  (s.e.m.)) for *lacO*<sub>1</sub>, for which the lower binding probability reduced the impact of the diffusion blockade, as the transcription factor will need multiple attempts to bind anyway<sup>5</sup>. If RNAP binds in a closed complex near LacI and blocks sliding in the same way as TetR, repression ratios would be expected to increase up to  $12.8 \pm 0.6$  (s.e.m.) and  $28.2 \pm 1.4$  (s.e.m.) for *lacO*<sub>1</sub> and *lacO*<sub>sym</sub>, respectively, from this effect alone.

In the third scheme (Fig. 3d), active transport or a combination of slow diffusion and degradation maintains a higher concentration of LacI close to the operator sites. This higher concentration of LacI can lead to faster association rates than we report above, as our association process started from any position in the cells when IPTG dissociated from LacI. A local gradient effect is expected to be greater for *lacO*<sub>sym</sub> than for *lacO*<sub>1</sub> as LacI is more likely to bind *lacO*<sub>sym</sub> before escaping to a random position<sup>5</sup>. Furthermore, previous studies have reported that the spatial distribution of LacI in the cell under poor growth conditions depends on where in the chromosome the protein is encoded<sup>12,13</sup>. However, under our experimental conditions, we could not observe any difference in the spatial distributions of non-operator-bound LacI expressed from different chromosomal loci with different intracellular locations (Supplementary Fig. 5 and Supplementary Note). Using single-particle tracking, we also did not observe that LacI could be trapped locally in the nucleoid for more than a few seconds. This timeframe is far shorter than what would be required to maintain a locally higher concentration of LacI close to the point of synthesis (Supplementary Fig. 6). In addition, we did not observe a change in the repression of the LacI-regulated *lacZ*YA operon when the *lacI* gene was moved to its mirror position on the other chromosome arm (Supplementary Note). Together, these results make it unlikely that LacI association is faster under steady-state growth than in our measurements owing to local concentration gradients of the repressor.

Our single-molecule chase method has allowed us to identify inconsistencies in the simple operator occupancy model of gene regulation in living *E. coli* cells, a model system where it is possible to conduct the experiment with sufficient accuracy. The inconsistencies are most easily explained by simple non-equilibrium mechanisms driven by transcription initiation itself. The same mechanisms are expected to operate in eukaryotic cells, where the added complexities of ATP-dependent chromatin remodeling<sup>14</sup> and clearing of the transcription factor binding region by divergent transcription<sup>15</sup> will contribute

to keeping operator occupancy out of equilibrium. Overall, non-equilibrium transcription factor kinetics add a new layer of complexity to the genomics puzzle beyond the steady-state mapping of transcription factor concentrations to gene activity.

## METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

## ACKNOWLEDGMENTS

We thank K.S. Matthews for advice on LacI mutants, G.-W. Li for helpful comments, X.S. Xie (Harvard University) for the Mall-Venus strain and I. Barkefors for critical reading of the manuscript. This work was supported by the European Research Council (ERC), the Knut and Alice Wallenberg Foundation (KAW), Vetenskapsrådet (VR), the Foundation for Strategic Research (SSF) and the Göran Gustafsson Foundation.

## AUTHOR CONTRIBUTIONS

J.E. conceived the project, J.E. and P.H. conceived the chase method, M.W. designed the microfluidics chip, P.H. and P.L. generated strains, P.H., M.W., Ö.B., F.P. and P.L. performed experiments, F.P., G.U., D.F. and M.W. developed and implemented analysis routines, P.H., F.P., M.W., G.U., D.F. and Ö.B. analyzed the data, J.E. and D.F. developed theoretical models, and J.E., P.H., D.F., M.W. and F.P. wrote the manuscript.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Bintu, L. *et al.* Transcriptional regulation by the numbers: applications. *Curr. Opin. Genet. Dev.* **15**, 125–135 (2005).
- Oehler, S., Amouyal, M., Kolkhof, P., von Wilcken-Bergmann, B. & Müller-Hill, B. Quality and position of the three *lac* operators of *E. coli* define efficiency of repression. *EMBO J.* **13**, 3348–3355 (1994).
- Oehler, S., Eismann, E.R., Kramer, H. & Müller-Hill, B. The three operators of the *lac* operon cooperate in repression. *EMBO J.* **9**, 973–979 (1990).
- Choi, P.J., Cai, L., Frieda, K. & Xie, X.S. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* **322**, 442–446 (2008).
- Hammar, P. *et al.* The Lac repressor displays facilitated diffusion in living cells. *Science* **336**, 1595–1598 (2012).
- Riggs, A.D., Bourgeois, S. & Cohn, M. The *lac* repressor-operator interaction. 3. Kinetic studies. *J. Mol. Biol.* **53**, 401–417 (1970).
- Chang, W.I., Barrera, P. & Matthews, K.S. Identification and characterization of aspartate residues that play key roles in the allosteric regulation of a transcription factor: aspartate 274 is essential for inducer binding in *lac* repressor. *Biochemistry* **33**, 3607–3616 (1994).
- Bell, C.E. & Lewis, M. A closer view of the conformation of the Lac repressor bound to operator. *Nat. Struct. Biol.* **7**, 209–214 (2000).
- Bell, C.E. & Lewis, M. Crystallographic analysis of Lac repressor bound to natural operator *O1*. *J. Mol. Biol.* **312**, 921–926 (2001).
- García, H.G. *et al.* Operator sequence alters gene expression independently of transcription factor occupancy in bacteria. *Cell Reports* **2**, 150–161 (2012).
- Sanchez, A., Osborne, M.L., Friedman, L.J., Kondev, J. & Gelles, J. Mechanism of transcriptional repression at a bacterial promoter by analysis of single molecules. *EMBO J.* **30**, 3940–3946 (2011).
- Kuhlman, T.E. & Cox, E.C. Gene location and DNA density determine transcription factor distributions in *Escherichia coli*. *Mol. Syst. Biol.* **8**, 610 (2012).
- Hermesen, R., ten Wolde, P.R. & Teichmann, S. Chance and necessity in chromosomal gene distributions. *Trends Genet.* **24**, 216–219 (2008).
- Vignali, M., Hassan, A.H., Neely, K.E. & Workman, J.L. ATP-dependent chromatin-remodeling complexes. *Mol. Cell. Biol.* **20**, 1899–1910 (2000).
- Seila, A.C., Core, L.J., Lis, J.T. & Sharp, P.A. Divergent transcription: a new feature of active promoters. *Cell Cycle* **8**, 2557–2564 (2009).

## ONLINE METHODS

**Strain construction.** Strains were constructed in a BW25993 background<sup>16</sup> using the  $\lambda$  Red<sup>16</sup> or pKO3 (ref. 17) protocols. Detailed strain descriptions can be found in the **Supplementary Note**, **Supplementary Figure 7** and **Supplementary Table 1**.

**Growth conditions.** Cells were grown in M9 minimal liquid medium supplemented with 0.4% glucose and RPMI amino acids (Sigma). For growth of strains harboring pBAD24 plasmids encoding *lacI*, *lacI-Venus* or *xylR*, the medium was supplemented with carbenicillin (Sigma).

For microfluidics experiments, saturated (overnight) cultures were diluted 1:200 in 40 ml of medium and grown at 37 °C for 4 h unless otherwise specified. Cells were collected by centrifugation and immediately loaded onto microfluidic chips as previously described<sup>18</sup>.

Information about growth conditions in other microscopy experiments and expression assays can be found in the **Supplementary Note**.

**Fluorescence microscopy and microfluidics.** *Microfluidic switching chips—design and preparation.* Microfabrication of the templates and construction of the individual devices were performed in accordance with the protocols described previously<sup>18</sup> with the exception that an extra medium port was added to allow for rapid exchange of medium. Inert polystyrene beads of 2  $\mu$ m in diameter (Sigma-Aldrich) were added to one medium reservoir. Beads allowed for the detection of flow rates and flow directions necessary for determining the induction states of the device during operation.

Relative height differences between medium reservoirs were used to control the pressure gradients and, thereby, flow rates and directions in the device during running and medium exchange. Medium exchange, i.e., anti-correlated elevation/lowering of reservoirs, was automated by using programmable linear actuators (Robocylinder, Intelligent Actuators), the control of which was synchronized with image acquisition using a custom-written Java program.

**Optical setup.** We used a Nikon Eclipse Ti-E microscope (with Nikon's Apo TIRF 100 $\times$ /1.49 oil immersion objective) equipped with a dichroic mirror (Chroma t515.5rdc), an excitation filter (Chroma 514/10), an emission filter (Chroma 550/50) and an EMCCD camera (iXon EM+ DU-897 from Andor). The camera was cooled to  $-80$  °C, and the linearized electron-multiplying gain was set to 150. A 2 $\times$  magnification lens was placed in the emission path before the camera. Fluorescence was excited by a Coherent Innova-304 Ar<sup>+</sup>-laser at 514 nm. When measuring association and dissociation rates, the power was 15 W/cm<sup>2</sup> using 4-s exposures. For single-particle tracking, the power was 650 W/cm<sup>2</sup>, and, for overnight growth experiments, the power was <5 W/cm<sup>2</sup> (see the **Supplementary Note** for details). A second camera (Scion Corp) was used for external phase-contrast imaging. The microscope was enclosed in an Okilab cage incubator where the set temperature was maintained at 37  $\pm$  0.1 °C, 42  $\pm$  0.1 °C or 25.5  $\pm$  0.3 °C. Image acquisition was controlled by the open-source software  $\mu$ Manager<sup>19</sup> in combination with custom-written acquisition scripts.

**Spot detection.** We used  $\acute{a}$  Trounev wavelet three-plane decomposition<sup>20</sup> and detected the spots in the second wavelet plane. Significant wavelet coefficients were determined through scale-dependent  $k\sigma$  thresholding where  $\sigma$  is the s.d. of the second wavelet plane, estimated by the MAD estimate<sup>21</sup>, and  $k = 3$  (association experiments) or  $k = 4$  (dissociation experiments).

**LacI-Venus kinetics using automated switching of medium.** Experiments were started when cells had grown to fill the whole microfluidic traps. For a fast and well-defined switch of medium, the medium reservoirs were connected to linear actuators and controlled from the computer in parallel with  $\mu$ Manager-run imaging acquisition.

For the analysis of operator-bound single LacI molecules in fusion with the fluorescent protein, YFP-derived Venus<sup>22</sup> (LacI-Venus), spots were detected as described above. Because the traps of the microfluidic chip were full with densely packed cells, we normalized the number of spots per trap by total cell area.

**Association with a single operator.** The principle of the experiment was essentially as presented previously<sup>5,23</sup> with the exception that the experiment was performed in the microfluidic device to allow for direct comparison with the corresponding dissociation experiment at 37 °C. The experiment was started by switching the medium for the induced cells from one containing IPTG to one containing the competitor ONPF at a 1 mM concentration. The addition of ONPF at high concentration was used to ensure that the association rate was not limited by the time it took for IPTG to leave the cell. Cells were imaged with 4-s exposures with a frame rate of  $\sim$ 0.18 frames/s. Fluorescent spots were counted as described above, and binding curves with data from the same strain were fitted (Igor Pro (v6.12A)) to the single exponential function  $y = a(1 - be^{-kt})$ , where  $a$  and  $b$  were independent for each series and  $k$  was the same for all series. Experiments were repeated to generate sufficient statistical power to test the hypothesis. For visualization in **Figure 2b**, the  $a$  and  $b$  parameters were used to normalize the data points in individual series before calculating the average and s.e.m. for each time point and plotting together with the fitted curve.

In **Supplementary Figure 2c**, the rate of LacI-Venus association is plotted as a function of the added ONPF concentration, and the plot shows that a 1 mM concentration is saturating. It also shows that LacI binds 1 min faster with the addition of ONPF at a saturating concentration, which suggests that it takes up to 1 min for the intracellular IPTG concentration to drop to a level where LacI can bind the operator. This timing is important for the dissociation assay described below. The relative difference in LacI-Venus concentration between strains is described in the **Supplementary Note** and **Supplementary Figure 8**.

**Chase assay for the measurement of dissociation rates.** In the *in vivo* chase experiment, LacI-Venus molecules are first bound to individual, single operator sites; then, through competition with non-fluorescent wild-type LacI in excess, they can be seen to dissociate as the number of fluorescent spots decreases. The chase experiment relies on the possibility of inducing binding of non-fluorescent LacI in a well-defined timeframe while LacI-Venus is already bound. To accomplish this, a single point substitution was introduced in the *lac* repressor gene (encoding LacI p.Asp274Asn), which causes more than a 1,000-fold reduction in IPTG affinity without changing operator binding strength<sup>7</sup>. The gene (referred to as *lacI<sub>c</sub>*) was expressed in fusion with Venus, resulting in a chromosomally expressed LacI<sub>c</sub>-Venus that does not dissociate, even in the presence of 1 mM IPTG (**Supplementary Fig. 1c**). Wild-type LacI was expressed from an arabinose-inducible promoter on the plasmid pBAD24.

The ratio between LacI<sub>c</sub>-Venus and wild-type LacI monomers when the plasmid was uninduced is seen at time 0 in **Supplementary Figure 1a**. When the plasmid was fully induced for a long time, the competitor copy number became so high that either 1 mM IPTG did not saturate LacI to prevent operator binding or the LacI<sub>c</sub>-Venus-LacI heterodimers, which naturally form (and are dominant when LacI is overexpressed) and bind one IPTG molecule, did not bind the operator. When instead XylR was expressed from pBAD24, LacI<sub>c</sub>-Venus was unaffected by IPTG (**Supplementary Fig. 1c**).

Before the switch, with IPTG present, LacI<sub>c</sub>-Venus homodimer bound the operator. When IPTG was removed at  $t = 0$ , there was a short (1-min) period of increased binding (**Supplementary Fig. 1b**). This increased binding is probably due to the association of heterodimers (in competition with non-fluorescent wild-type LacI) to available operator sites. Because of this initial association and the time delay required to reduce the intracellular IPTG concentration to a level where non-fluorescent LacI bound (see below and **Supplementary Fig. 2**), we fit the dissociation process from 1.5 min after switching to medium without IPTG to an exponential decay process that also took into account the fact that the transcription factor was displaced once per generation owing to replication. The implications of the approximation are quantified below and in the **Supplementary Note**. Experiments were repeated to generate sufficient statistics to test the hypothesis.

**Time-dependent excess of non-fluorescent LacI.** We induced the expression of non-fluorescent LacI at time 0 (medium containing 1 mM IPTG was switched to medium containing 0.2% arabinose). This switch resulted in a time-dependent increase in the concentration of the non-fluorescent LacI chase molecules (**Supplementary Fig. 1a**). This time-dependent increase motivated us to calculate how this would influence the measured dissociation kinetics.



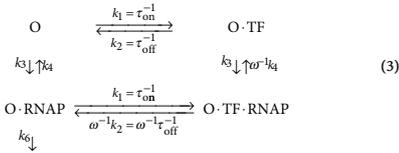
The equations that describe the probabilities that an operator is initially bound by a fluorescent molecule ( $P_F$ ), that it is empty ( $P_E$ ) or that it is occupied by a non-fluorescent molecule ( $P_N$ ) were

$$\begin{aligned} \frac{dP_F(t)}{dt} &= \tau_{\text{on}}^{-1}P_E - \tau_{\text{off}}^{-1}P_F \\ \frac{dP_E(t)}{dt} &= \tau_{\text{off}}^{-1}(P_F + P_N) - \tau_{\text{on}}^{-1}(1 + q(t))P_E \\ \frac{dP_N(t)}{dt} &= q(t)\tau_{\text{on}}^{-1}P_E - \tau_{\text{off}}^{-1}P_N \\ \frac{dP_F(0)}{dt} &= 1, \quad \frac{dP_E(0)}{dt} = 0, \quad \frac{dP_N(0)}{dt} = 0 \end{aligned} \quad (2)$$

Here  $q(t)$  is the fold excess of non-fluorescent transcription factor, which was measured directly by protein blot (Supplementary Fig. 1a) and is closely approximated by  $q(t) = 4 + t^2$ , where  $t$  is the time in minutes after the addition of IPTG.

For an infinitely high  $q$ ,  $P_F$  will decay as a pure exponential with rate  $\tau_{\text{off}}^{-1}$  starting from  $t = 0$ . For a finite  $q$ , the observed dissociation process is slightly slower. When fitting a single exponential function to the solution of  $P_F(t)$ , using parameters from Table 1, starting from 1.5 min and ending at 20 min, the dissociation rate is underestimated by up to 11% for  $lacO_1$  and by up to 9% for  $lacO_{\text{sym}}$  owing to the finite concentration of non-fluorescent Lacl. This underestimation would change the predicted repression ratios (based on the simple operator occupancy model) to 10.2 for  $lacO_1$  and 19.4 for  $lacO_{\text{sym}}$ , which do not alter the conclusions drawn when assuming large excess of non-fluorescent Lacl.

**Models. Cooperative Lacl binding.** Consider the scheme in Figure 3a,b written in further detail.



Here Lacl and RNAP bind  $\omega$  times longer when they are binding at the same time. The repression ratio in this non-equilibrium scheme is

$$\text{RR} = 1 + \frac{k_1(k_1 + k_2 + k_3 + k_4 + k_6)(k_3 + k_4\omega^{-1})}{k_2\omega^{-1}(k_1 + k_2 + k_3 + k_4)(k_3 + k_4 + k_6)} \quad (4)$$

If we assume that transcription initiation is slow where  $k_6 \rightarrow 0$  (equilibrium case),

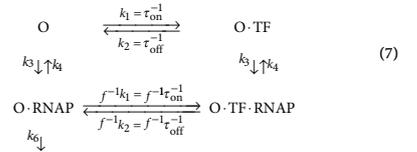
$$\text{RR} = 1 + \frac{\omega k_1(k_3 + k_4\omega^{-1})}{k_2(k_3 + k_4)} \xrightarrow{k_3 \gg k_4} 1 + \frac{\omega k_1}{k_2} = 1 + \frac{\omega \tau_{\text{off}}}{\tau_{\text{on}}} \quad (5)$$

If we assume that transcription initiation is fast where  $k_6 \rightarrow \infty$  (far from equilibrium case),

$$\begin{aligned} \text{RR} &= 1 + \frac{k_1(\omega k_3 + k_4)}{k_2(k_1 + k_2 + k_3 + k_4)} \xrightarrow{k_3 \gg k_4} 1 + \frac{\omega k_1}{k_2} \frac{k_3}{k_1 + k_2 + k_3} \\ &= 1 + \frac{\omega k_1}{k_2} \frac{k_3}{k_1 + k_3} \end{aligned} \quad (6)$$

These are the limiting approximations given in the main text (Fig. 3a,b). To see what we obtained with specific numbers, we used the measured  $\tau_{\text{on}}$  and  $\tau_{\text{off}}$  values and assumed that  $\omega = 1.5$ ,  $k_3 = 1 \text{ min}^{-1}$  and  $k_4 = 0.1 \text{ min}^{-1}$ . These numbers gave  $\text{RR} = 10.0$  and an induced transcription initiation rate of  $0.61 \text{ min}^{-1}$  (refs. 24,25) when  $k_6 = 1.7 \text{ min}^{-1}$  for  $lacO_1$  and  $\text{RR} = 28.2$  and an induced transcription initiation rate of  $0.61/5.4 \text{ min}^{-1}$  for  $lacO_{\text{sym}}$  when  $k_6 = 0.14 \text{ min}^{-1}$ . The value of 5.4 is the measured difference in expression between the induced  $lac$  operon controlled by  $lacO_1$  and  $lacO_{\text{sym}}$ .

**Non-equilibrium model with roadblock.** Consider the scheme in Figure 3c written in further detail.



The repression ratio  $\text{RR}$  is

$$\text{RR} = \frac{k_1(k_3 + k_4)(k_4 + k_6)}{k_2(k_3 + k_4 + k_6)} + k_4 + f^{-1}(k_1 + k_2 + k_3) \left( \frac{k_1(k_3 + k_4)}{k_2(k_3 + k_4 + k_6)} + 1 \right) \quad (8)$$

Assuming that the system is far from equilibrium, such that  $k_6 \gg k_3 + k_4$ , and that the transcription initiation rate is fast enough, such that  $k_1(k_3 + k_4)/(k_2k_6) \ll 1$ , then the repression ratio is

$$\text{RR} = 1 + \frac{k_1(k_3 + k_4)}{k_2(k_4 + f^{-1}(k_2 + k_3 + k_1))} \quad (9)$$

Further, assuming that RNAP binding is strong, such that  $k_3 \gg k_4$ , that the turnover of RNAP is faster than the turnover of the transcription factor, such that  $k_3 \gg k_1 + k_2$ , and that  $f$  is not very much smaller than 1, then the repression ratio is

$$\text{RR} = 1 + \frac{fk_1}{k_2} = 1 + \frac{f\tau_{\text{off}}}{\tau_{\text{on}}} \quad (10)$$

16. Datsenko, K.A. & Wanner, B.L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* **97**, 6640–6645 (2000).
17. Link, A.J., Phillips, D. & Church, G.M. Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli*: application to open reading frame characterization. *J. Bacteriol.* **179**, 6228–6237 (1997).
18. Ullman, G. et al. High-throughput gene expression analysis at the level of single proteins using a microfluidic turbidostat and automated cell tracking. *Phil. Trans. R. Soc. Lond. B* **368**, 20120025 (2013).
19. Edelstein, A., Amodaj, N., Hoover, K., Vale, R. & Stuurman, N. Computer control of microscopes using microManager. *Curr. Protoc. Mol. Biol.* **Chapter 14**, Unit 14.20 (2010).
20. Olivo-Marin, J.C. Extraction of spots in biological images using multiscale products. *Pattern Recognit.* **35**, 1989–1996 (2002).
21. Sadler, B.M. & Swami, A. Analysis of multiscale products for step detection and estimation. *IEEE Trans. Inf. Theory* **45**, 1043–1051 (1999).
22. Nagai, T. et al. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* **20**, 87–90 (2002).
23. Elf, J., Li, G.W. & Xie, X.S. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191–1194 (2007).
24. Taniguchi, Y. et al. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010).
25. Yu, J., Xiao, J., Ren, X., Lao, K. & Xie, X.S. Probing gene expression in live cells, one protein molecule at a time. *Science* **311**, 1600–1603 (2006).

## Supplementary information

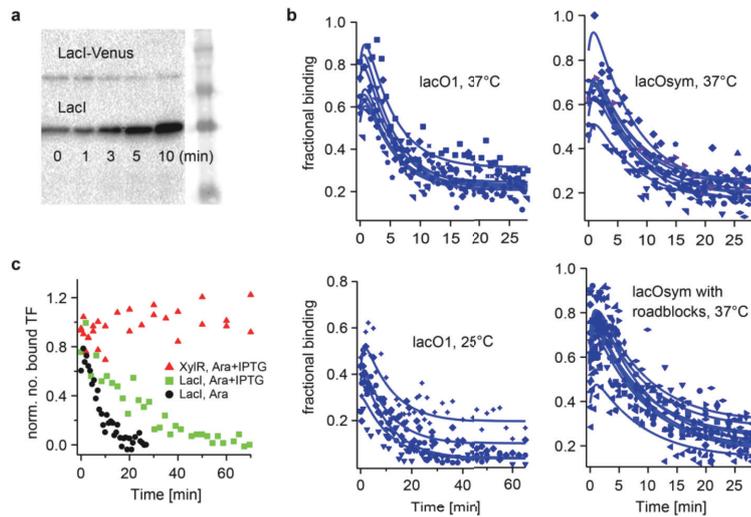
Direct measurement of transcription factor dissociation excludes a simple operator occupancy model for gene regulation

Hammar, Walldén *et al.* 2014

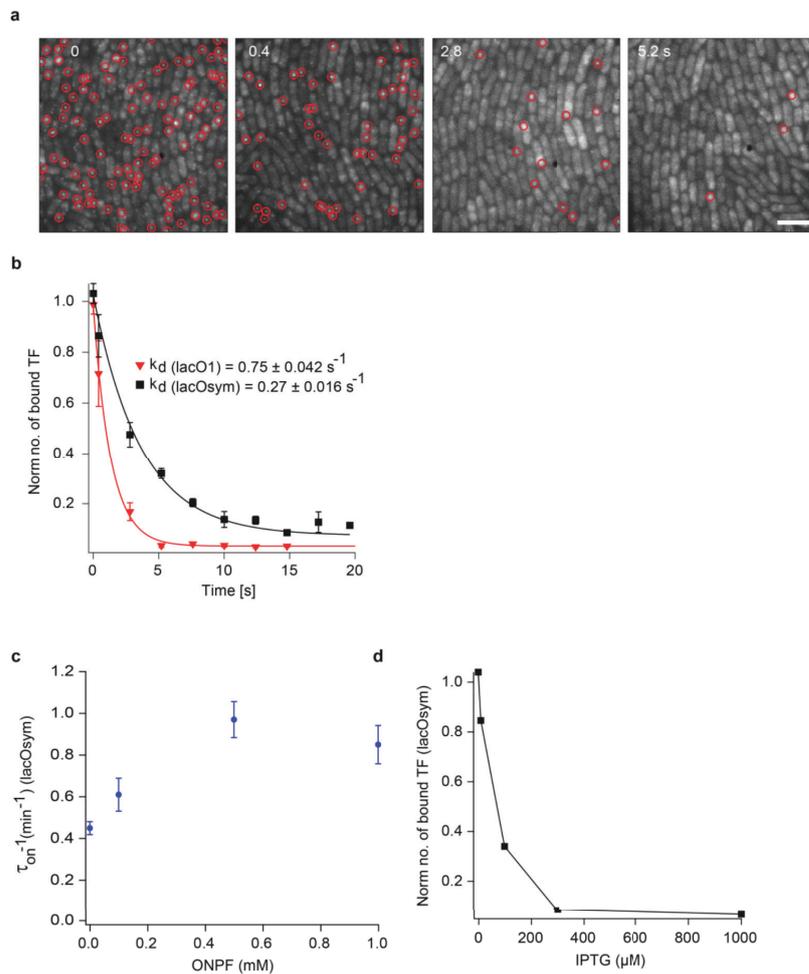
### Table of contents

Supplementary Figures .....	2
Supplementary Tables .....	12
Supplementary Note .....	13
1. The operator occupancy model at equilibrium .....	13
2. IPTG induction .....	14
3. Replication-dependent dissociation .....	15
4. Relative repressor concentrations .....	16
5. Are there spatial gradients of LacI in rapidly growing cells? .....	17
Results .....	17
Experimental procedures .....	18
6. Strain construction .....	21
7. Growth conditions and expression assays .....	22
8. Supplementary references .....	23

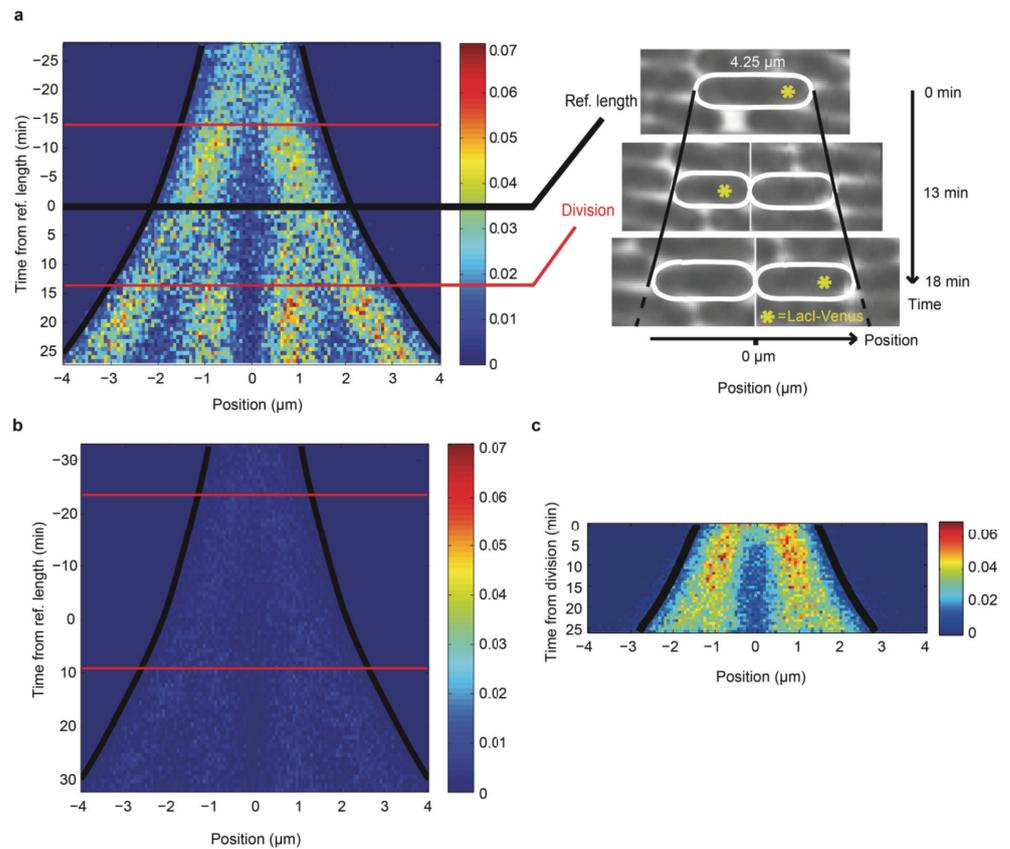
## Supplementary Figures



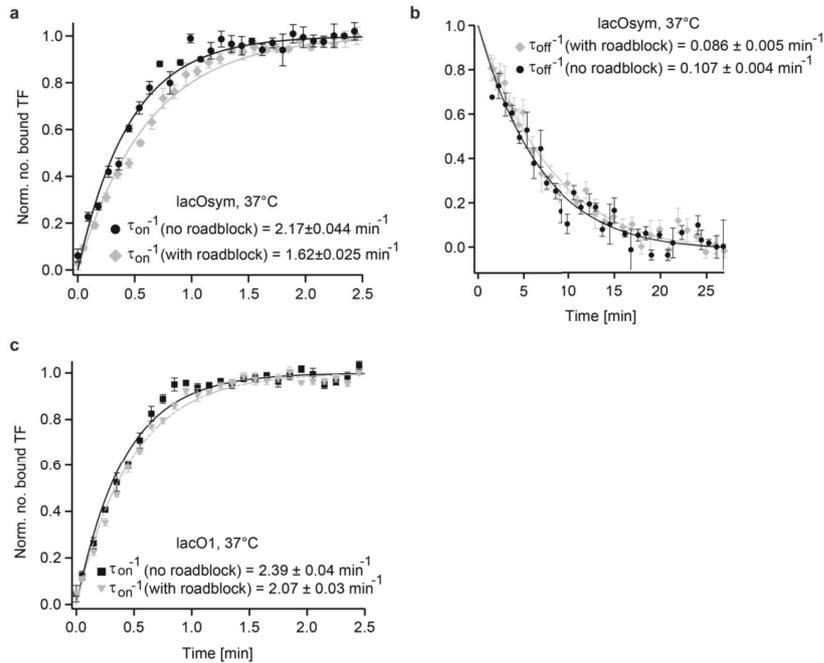
**Supplementary Figure 1. Single-molecule chase assay for dissociation.** (a) Western blot of strain JE127 harboring pBAD24-*lacI*. While plasmid-expressed LacI was immediately upregulated by the addition of arabinose, chromosomal LacI<sub>s</sub>-Venus remained unaffected. Using ImageJ (supplementary reference 2) the excess of competitor (LacI over LacI<sub>s</sub>-Venus) was quantified for each time point to be 4× (at 0 min), 5× (1 min), 12× (3 min) and >20× (5 and 10 min). (b) Raw data of the different dissociation rate measurements shown in Figure 2a and Supplementary Figure 4b—each data set is fitted including an initial association of fluorescent molecules as described in the Online Methods. For each strain/condition all series are fitted to a fixed generation time (i.e., time between divisions, which is 26 min at 37 °C (Supplementary Fig. 3a) and 70 min at 25 °C (data not shown)), a common  $\tau_{\text{off}}$  and individual parameters for start and end levels. (c) LacI<sub>s</sub>-Venus binding to *lacO*<sub>sym</sub> at 37 °C; in black circles (as in Fig. 2a)—pBAD24-*lacI*, switch from IPTG to arabinose; in green squares—pBAD24-*lacI*, switch from IPTG to IPTG + arabinose; in red triangles—pBAD24-*xyIR*, switch from IPTG to IPTG + arabinose.



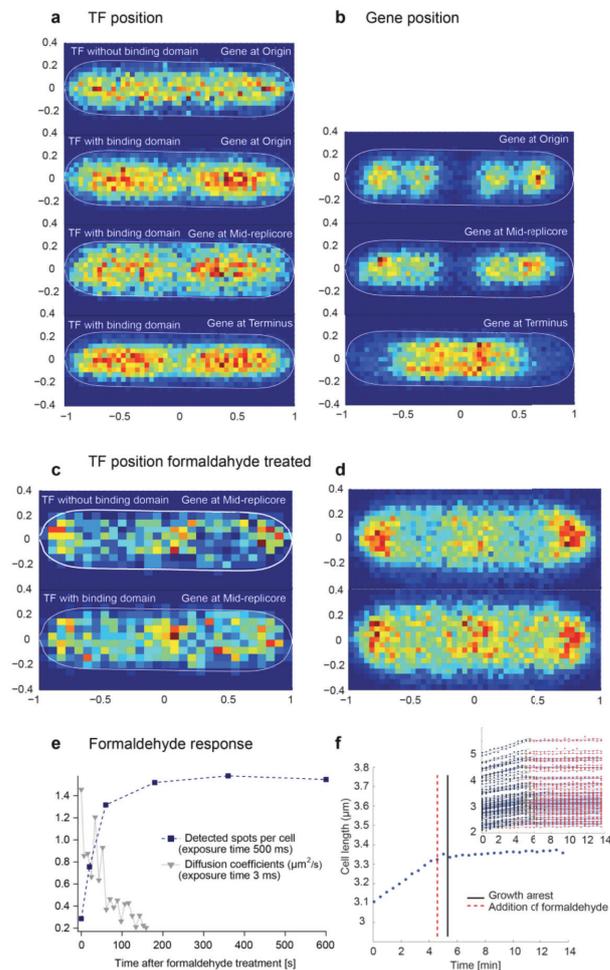
**Supplementary Figure 2. Fast switching kinetics as demonstrated by IPTG response. (a)** Images show binding of LacI-Venus to *lacO*<sub>1</sub> (at 37 °C), with automatically detected localized spots encircled, before (0) and 0.4, 2.8 and 5.2 s after removal of IPTG. Scale bar = 4  $\mu\text{m}$ . As described in the Supplementary Note, a 2-s delay was subtracted from the time points to account for the time delay from the medium switch. **(b)** The corresponding dissociation curves from *lacO*<sub>1</sub> and *lacO*<sub>sym</sub> operators. **(c)** Association rates to *lacO*<sub>sym</sub> measured with increasing competing concentration of the anti-inducer ONPF. **(d)** Steady-state binding (number of spots per cell after 15-min incubation with IPTG at different concentrations) to *lacO*<sub>sym</sub>.



**Supplementary Figure 3. Average number of operator sites.** The number of *lac* operator sites per cell increases at replication, decreases at division and can be detected as specifically bound LacI-Venus molecules. **(a)** The observed frequency (number of molecules  $\text{cell}^{-1} \text{min}^{-1} \mu\text{m}^{-1}$ ) of specifically bound LacI-Venus molecules mapped onto the major axis (horizontal) of the cells over time (vertical, from top to bottom). The exponentially expanding black lines indicate the average apex position of the poles of the cells, and the red horizontal lines indicate the average division times. Cells are synchronized at the time they achieve a reference length (4.25  $\mu\text{m}$ ). **(b)** The contribution of spots detected in a strain in which all operator sites have been removed (same scale as in a). **(c)** Same as (a) but with cells synchronized according to time of division.

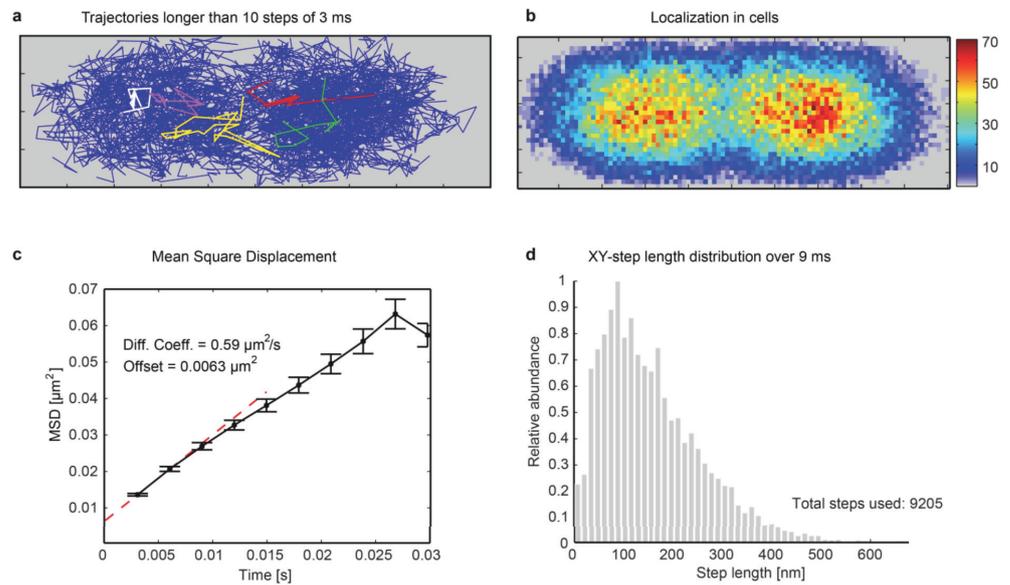


**Supplementary Figure 4. Roadblock effect for association to  $lacO_{sym}$  and  $lacO_1$  and dissociation from  $lacO_{sym}$  (as described in Fig. 1a and summarized in Table 2). (a)** Association to  $lacO_{sym}$  with and without roadblock on one side of the operator; the roadblock effect is  $f = 1.35 \pm 0.04$ . **(b)** Dissociation with and without roadblock; the same data as in Supplementary Figure 1b (bottom right panel) and fitted as described in the Supplementary Note; the roadblock effect is  $f = 1.25 \pm 0.09$ . **(c)** Association in a strain with two  $lacO_1$  operators with binding sites for two TetR roadblocks in between. Association curves with and without TetR expressed are shown. The roadblock effect is  $f = 1.16 \pm 0.03$ . (a–c) We verified that the TetR roadblock binds >99% of the time by comparing  $\beta$ -galactosidase activity  $\pm$  aTC in a strain (JE125; Supplementary Fig. 7b) where  $lacO_1$  is replaced by the *tet* operator *tetO*<sub>2</sub> and TetR is chromosomally and highly expressed as in the kinetic experiments. The repression ratio is  $220 \pm 26$  (mean  $\pm$  s.d.). As a control the experiment was repeated without TetR expressed (strain JE112), and here the repression ratio is  $1.0 \pm 0.35$  (mean  $\pm$  s.d.). Comment: the rates for binding to two operator sites in (c) are not a factor of two higher as compared to Figure 2b as under these experimental conditions, i.e., exponential growth at 37 °C in the microfluidic device, the fraction of mature Venus is on the order of 0.25 and many sites are occupied by non-fluorescent LacI-Venus.



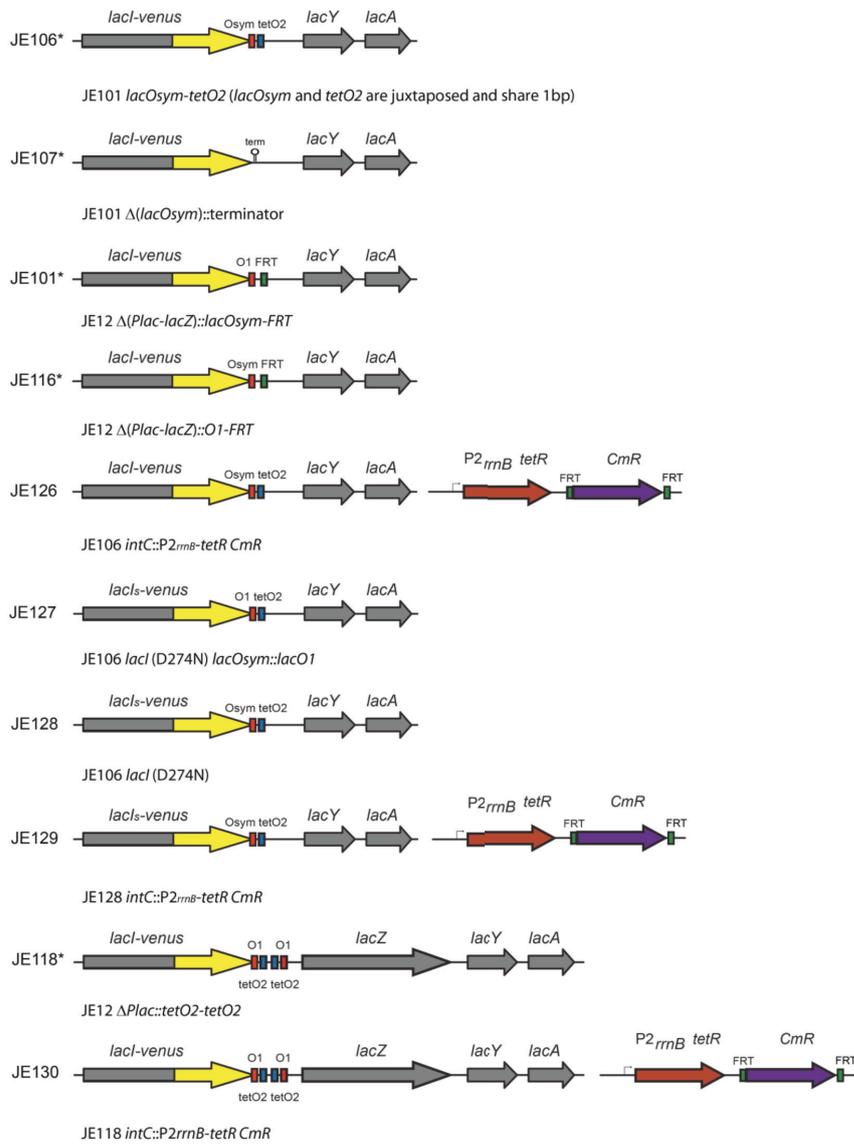
**Supplementary Figure 5. Spatial distribution of non-specifically bound LacI-Venus molecules in cells between 4.6  $\mu\text{m}$  and 5.1  $\mu\text{m}$ .** (a) The distribution of LacI-Venus molecules observed at 2-ms exposures in strains without specific operator sites. The top panel shows the distribution of LacI42-Venus ( $n = 20,863$  cells with  $\langle s \rangle = 0.26$  detected spots per cell), in which the DNA binding domain of LacI has been deleted. The following panels (from top to bottom) show strains expressing LacI-Venus from loci proximate to the origin of replication ( $n = 15,538$ ,  $\langle s \rangle = 0.77$ ), mid-replicore ( $n = 13,572$ ,  $\langle s \rangle = 0.65$ ) and the terminus ( $n = 19,023$ ,  $\langle s \rangle = 0.63$ ), respectively. (b) Localization of chromosomal gene loci as determined by the specific binding to operator sites. Top, origin seen as the binding of plasmid-expressed LacI-Venus to an array of *lacO*<sub>1</sub> operators ( $n = 5,453$ ,  $\langle s \rangle = 1.55$ ). Middle, mid-replicore seen as the binding of chromosomal LacI-Venus to single *lacO*<sub>1</sub> ( $n = 10,942$ ,  $\langle s \rangle = 0.5$ ) (data the same as in Supplementary Fig. 3a). Bottom, terminus seen as the binding of chromosomal Mall-Venus to two colocalized *malO* sites ( $n = 12,648$ ,  $\langle s \rangle = 0.8$ ). (c) Distribution of LacI42-Venus without (top) ( $n = 517$ ,  $\langle s \rangle = 1.35$ ) and LacI-Venus with (bottom) ( $n = 1,054$ ,  $\langle s \rangle = 1.7$ ), DNA binding domain; expressed from the mid-replicore region and imaged 40 min after fixation with formaldehyde. (d) Same as (c), although with data pooled from the whole data set and thus a broader range of cell lengths (2.3–5.6  $\mu\text{m}$ ) (top) ( $n = 17,579$ ,  $\langle s \rangle = 0.96$ ) (bottom) ( $n = 15,175$ ,  $\langle s \rangle = 1.38$ ). (e) Time lapse

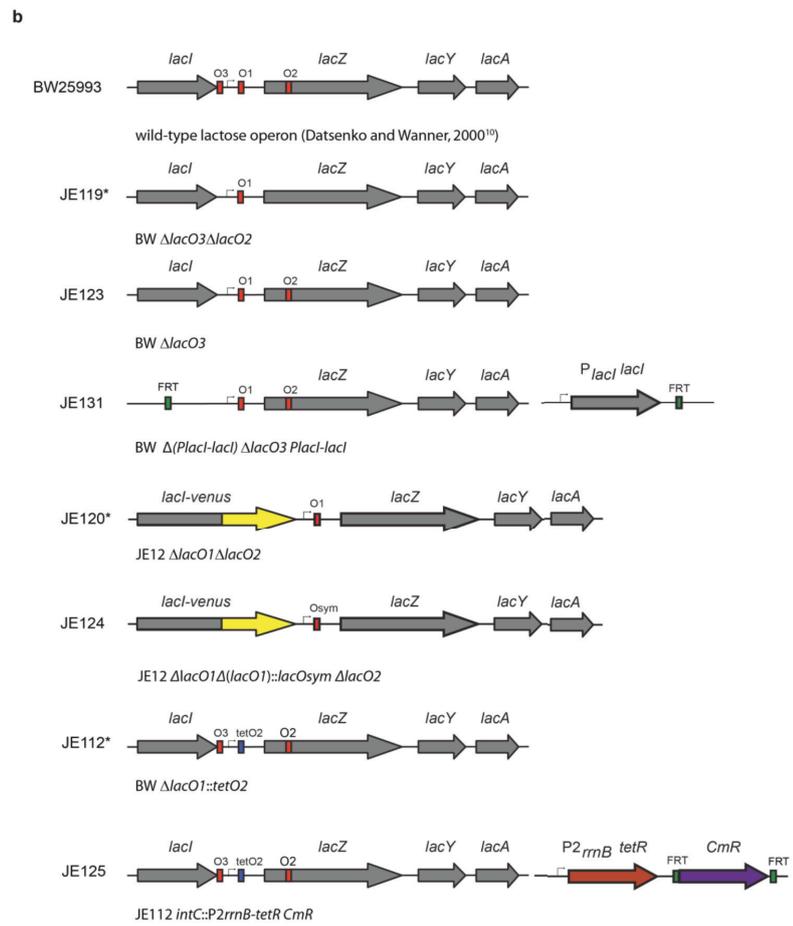
showing how non-specifically binding LacI-Venus molecules are immobilized by formaldehyde treatment; the detection of near-diffraction limited spots increases with time when imaging is done using long exposure times (blue squares), and the rate of diffusion decreases on the same timescale as measured by fast tracking (3.5 ms between frames). **(f)** The growth of single cells as a function of time (25 s between frames) before and after (vertical red dashed line) formaldehyde treatment. The data show the average of 230 individual cell traces. The vertical black continuous line marks the point of growth arrest as found by regression analysis. Inset, a subset of the cell traces (from a single trap in the microfluidic chip) and the fitting to each one of them is shown with the same axis labels as in the main figure.



**Supplementary Figure 6. Tracking of LacI-Venus in cells between 2 and 3  $\mu\text{m}$ .** (a) Trajectories longer than 30 ms at a sampling rate of 330 Hz with 1-ms exposures in a strain without specific operator sites, mapped onto a unit cell (425 trajectories from approximately 500 cells). Five trajectories were randomly chosen and highlighted. (b) The distribution of detected spots mapped onto a unit cell (117,891 points from approximately 500 cells). (c) MSD analysis for the two-dimensional trajectory data presented in (a). The analysis was performed prior to mapping coordinates onto the unit cell. The offset indicates a localization accuracy of 40 nm. (d) A histogram showing the distribution for step lengths in two dimensions for the diffusing molecules during a time of 9 ms, i.e., for trajectories subsampled by a factor of 3.

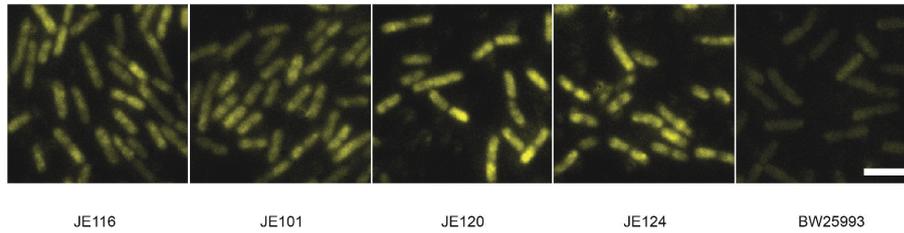
a



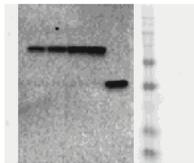


**Supplementary Figure 7. Strains used in this study.** Those marked with an asterisk were previously published (supplementary reference 1). **(a)** Strains used in microscopy experiments and **(b)** strains used in  $\beta$ -galactosidase experiments.

a



b



**Supplementary Figure 8. Calibration of LacI-Venus expression levels.** (a) From left to right: total fluorescence in strains (Supplementary Fig. 7) having LacI-Venus being autorepressed at the sequence *lacO*<sub>1</sub> (JE116) and the sequence *lacO*<sub>sym</sub> (JE101) then in two strains having LacI-Venus and no autorepression (JE120, JE124) and a strain lacking Venus (BW25993). Scale bar = 4  $\mu$ m. (b) Western blot using antibody against LacI in the corresponding strains. From left to right: JE116, JE101, JE120, JE124 and BW25993.

## Supplementary Tables

**Supplementary Table 1.** Operator sequences used in this work. Sequences that are mutated such that their protein binding specificity has been removed are labelled with a “-”.

Operator	Sequence
<i>lacO<sub>sym</sub></i>	AATTGTGAGCGCTCACAATT
<i>lacO<sub>1</sub></i>	AATTGTGAGCGGATAACAATT
<i>lacO<sub>1</sub>-</i>	AATTGTGATGAAGCTTGTCAG
<i>lacO<sub>2</sub></i>	GGTTGTTACTCGCTCACATTT
<i>lacO<sub>2</sub>-</i>	GGCTGCTATAGCTTGACGTTT
<i>lacO<sub>3</sub></i>	GGCAGTGAGCGCAACGCAATT
<i>lacO<sub>3</sub>-</i>	GGCAGTGATGAAGCTTGTCAG
<i>tetO<sub>2</sub></i>	TCCCTATCAGTGATAGAGA
<i>malO<sub>1</sub></i>	AGGTAAAACGTTTTATCT
<i>malO<sub>2</sub></i>	AGATAAAAACGTTTTATCA

## Supplementary Note

### 1. THE OPERATOR OCCUPANCY MODEL AT EQUILIBRIUM

Consider a situation where there exists a number of TF free operator states,  $C_i^{free}$ , and a number of TF bound operator states,  $C_j^{bound}$ . For example, in Figure 3a of the main text, there are two TF free states, operator without RNAP bound and operator with RNAP bound. Assume further that all TF free states are at equilibrium with each other as are the TF bound states. A TF binds the free state  $i$  with rate constant  $k_a^i$  and dissociates from the bound state  $j$  with rate constant  $k_d^j$ . Transcription is initiated from TF free state  $m$  with rate constant  $k_m$ .

The total concentration of free states is given by  $[C]_{free} = \sum_i [C]_i^{free}$  and the total concentration of bound states is given by  $[C]_{bound} = \sum_j [C]_j^{bound}$ . The total concentration of all states is constant  $[C]_{tot} = [C]_{free} + [C]_{bound}$ . In absence of TF  $[C]_{free} = [C]_{tot}$ . At equilibrium the repression ratio is

$$RR = \frac{\sum_m k_m P(m) [C]_{tot}}{\sum_m k_m P(m) [C]_{free}} = \frac{[C]_{tot}}{[C]_{free}} \quad \text{Eq. S1}$$

where  $P(m)$  is the probability to be in the state  $m$  that can initiate transcription given that the operator is free. At equilibrium the flux from bound to unbound states and vice versa are related by

$$\sum_i k_a^i [TF]_{free} P(C_i^{free}) [C]_{free} = \sum_j k_d^j P(C_j^{bound}) [C]_{bound} \quad \text{Eq. S2}$$

Using the above relation together with the mass conservation criterion gives  $RR = (\tau_{off} + \tau_{on}) / \tau_{on}$  where  $\tau_{on} = 1 / \sum_i k_a^i [TF]_{free} P(C_i^{free})$  and  $\tau_{off} = 1 / \sum_j k_d^j P(C_j^{bound})$ .

In the case where  $k_a^i = k_a$  for all  $i$  and  $k_d^j = k_d$  for all  $j$  we arrive at the simple non-cooperative model with  $\tau_{off} = 1 / k_d$  and  $\tau_{on} = 1 / (k_a [TF])$  defined from what we can measure for dissociation and binding on DNA without RNAP. The expression is also true for the cooperative model at equilibrium but  $\tau_{off}$  would in this case be another value than  $\tau_{off}$  that we measure by dissociation from a state without RNAP bound. For this reason it is easier to describe the cooperative models independently (see Online Methods).

## 2. IPTG INDUCTION

The steady state occupancy of the operator at different isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) concentrations is shown in Supplementary Figure 2d. At 300  $\mu$ M IPTG the number of detected bound molecules are down at the background level of about 0.05 detected spots per cell. We estimate how fast IPTG goes in to the cell by measuring the temporal response of switching from 0 to 300  $\mu$ M IPTG in the microfluidic chip (Supplementary Fig. 2b). We measure that it takes  $\sim 2$  s until the medium surrounding the cells is replaced once the switch-command is executed. Therefore, the fraction of bound molecules before induction is plotted at time 0, and the subsequent data points are plotted at the time of acquisition minus a 2 s delay. The curves thus show the time from having un-induced to IPTG-bound LacI including membrane transport and intracellular medium exchange. 1 s exposures were used at a frame rate of  $0.4 \text{ s}^{-1}$ . A time trace of the kinetic response is illustrated in Supplementary Figure 2a. LacI-Venus dissociates from the operator within a few seconds from the switch to IPTG and the steady state binding level of 0.05 spots per cell is reached within 10 s.

### 3. REPLICATION-DEPENDENT DISSOCIATION

In the chase experiment there are two possible ways by which the TF can go from its DNA bound form to its DNA unbound form. The first is by spontaneous dissociation and the second is by the replication machinery passing the TF binding site. The time for spontaneous dissociation is exponentially distributed while the time until dissociation by replication is given by a uniform distribution starting at zero and ending at the generation time,  $\tau_G$ . The rationale for the uniform distribution is that any random cell chosen for analysis can be anywhere in its replication cycle.

When there are  $n$  numbers of DNA-bound TFs per cell they change according to

$$\{n\} \xrightarrow{\lambda_{diss}(t)} \{n-1\} \quad \text{Eq. S3}$$

for spontaneous dissociation and according to

$$\{n\} \xrightarrow{\lambda_{rep}(t)} \{0\} \quad \text{Eq. S4}$$

at replication.

When starting with  $N$  DNA-bound TFs at time zero, the probability of having a certain number of DNA-bound TFs at time  $t$  evolves according to

$$\begin{aligned} \frac{dP(0,t)}{dt} &= \lambda_{diss}(t)P(1,t) + \sum_{i=1}^N \lambda_{rep}(t)P(i,t) \\ \frac{dP(n,t)}{dt} &= -\lambda_{diss}(t)nP(n,t) + \lambda_{diss}(t)(n+1)P(n+1,t) - \lambda_{rep}(t)P(n,t) \\ \frac{dP(N,t)}{dt} &= \lambda_{diss}(t)NP(N,t) - \lambda_{rep}(t)P(N,t) \end{aligned} \quad \text{Eq. S5}$$

For exponentially distributed dissociation times the reaction propensity  $\lambda_{diss}(t) = \tau_{off}^{-1}$  and for uniformly distributed dissociation times in the case of replication the reaction propensity  $\lambda_{rep}(t) = (\tau_G - t)^{-1}$ . Note that the time  $t$  now only is defined up to the generation time  $\tau_G$ .

The ensemble average of the number of DNA-bound TFs,  $\langle n(t) \rangle$  is given by

$$\frac{d\langle n(t) \rangle}{dt} = -\left(\tau_{off}^{-1} + (\tau_G - t)^{-1}\right) \langle n(t) \rangle \quad \text{Eq. S6}$$

which has the solution

$$\langle n(t) \rangle = \langle n(0) \rangle \frac{e^{-t/\tau_{off}} (\tau_G - t)}{\tau_G} \quad \text{Eq. S7}$$

In the chase experiment we initially also see a rapid increase in the number of DNA-bound TFs. We have approximated the effect of binding by adding a second exponential to the expression used for regression of the average number of DNA-bound TFs to the chase experiment as

$$\langle n(t) \rangle = C_1 \frac{e^{-t/\tau_{off}} (\tau_G - t)}{\tau_G} + C_2 (1 - e^{-t/\tau_{on}}) \quad \text{Eq. S8}$$

With the parameters gained by regression of the chase experiment to Eq. S7 above (Supplementary Fig. 1b) the contribution of the second exponential term,  $C_2 e^{-t/\tau_{on}}$  is less than 5% after 1.5 minutes. For the chase experiments fluorescent spots were counted as described in the Online Methods and dissociation curves with data from the same strain were thus fitted to the function  $y = a((\tau_G - t)/\tau_G) e^{-t/\tau_{off}}$ , where  $a$  is independent for each series,  $\tau_{off}$  shared between series and  $\tau_G$  set to the generation time of 26 min. Experiments were repeated to get sufficient statistics to test the hypothesis.

#### 4. RELATIVE REPRESSOR CONCENTRATIONS

The reduction in LacI-Venus expression due to auto-repression by *lacO<sub>1</sub>* in the *lacO<sub>3</sub>* position was previously estimated to be a factor of 1.7<sup>1</sup>. The measurement is based on comparing the integrated fluorescence intensity between the different strains after subtraction of background fluorescence from BW25993. The measurements were here repeated where JE101 with *lacO<sub>sym</sub>* in the *lacO<sub>3</sub>* position, and JE116 with *lacO<sub>1</sub>* in the *lacO<sub>3</sub>* position, are respectively compared to strains JE124 and JE120 without auto-repression and with *lacO<sub>sym</sub>* or *lacO<sub>1</sub>* in the actual *lacO<sub>1</sub>* position in order to determine the relative *lac* repressor concentrations between strains. Cells were grown at 37°C, immobilized on poly-L-Lysine coated culture dishes and induced with 1 mM IPTG just prior imaging at 4 s exposures. Data was analyzed using ImageJ<sup>2</sup>. There is a 1.8±0.2-fold (mean ± s.d.) decrease in expression for JE116, with auto-repression at *lacO<sub>1</sub>*, as compared to JE120 and a 2.1±0.2-fold (mean ± s.d.) decrease for JE101 as compared to JE124 where binding to *lacO<sub>sym</sub>* causes a slightly stronger auto-repression. We find no difference in expression between JE120 and JE124, which is expected since none of them are auto-repressed. Representative images (of a subset of cells) together with a corresponding western blot of the same strains are shown in Supplementary Figure 8.

When accounting for these concentration differences we find that the measured repression ratios (with the β-galactosidase assay) at 37°C, which is 62±3.5 (mean ± s.e.m.; n=8) for *lacO<sub>sym</sub>* and 18±1.3 (mean ± s.e.m.; n=9) for *lacO<sub>1</sub>*, corresponds to 29.7±3.4 and 10.0±1.3 in the respective strain with auto-repression. The actual β-galactosidase activity in the fully induced states, which are assumed to be proportional to the expression levels, are 215±13 (mean ± s.e.m.; n=8) for *lacO<sub>sym</sub>* and 1150±90 (mean ± s.e.m.; n=9) for *lacO<sub>1</sub>*. The ratio between the values for the two operators is ~5.4 and could for example be explained by a difference in abortive transcription initiation rate at the two different sequences<sup>3</sup>.

## 5. ARE THERE SPATIAL GRADIENTS OF LACI IN RAPIDLY GROWING CELLS?

A locally higher LacI gradient around the position of the gene can be maintained if the diffusion away from this region is slow as compared to the rate of protein degradation and cell growth. Such gradients have been reported for *E. coli* in stationary phase or growing under poor growth conditions<sup>4</sup>. We set out to test if we can find evidence for gradients in our conditions with rapidly growing cells.

### Results

In order to study if there is any spatial dependence of the position of the gene for the fluorescent LacI-Venus we used three strains kindly provided by the Cox lab: LacI-Venus is expressed from the origin, the mid replicore or the terminus. The three strains were imaged to determine the location of LacI-Venus in more than 20000 cells per strain using the microfluidic chip device (see the section below for experimental details). Molecules that are non-specifically bound are visible due to short laser exposures (2 ms). The results were compared to a strain where the DNA binding domain of LacI-Venus has been removed. In Supplementary Figure 5a we show the intracellular distribution of observed LacI-Venus expressed from the three different positions on the chromosome in cells with length 4.6-5.1  $\mu\text{m}$ . In this length interval we have more than 10000 cells per strain. We conclude that under our experimental conditions there are no significant differences in intracellular localization of fluorescent LacI-Venus based on where on the chromosome the protein is encoded. For comparison of the LacI position to the corresponding *lacI* gene position we determined the position of the origin, mid replicore and terminus under the same conditions as above using fluorescent proteins binding specifically to these regions. The result (Supplementary Fig. 5b) shows a clear difference between terminus and mid-replicore. For origin the result is similar to mid-replicore although a tendency towards separation of the origin chromosomal regions from 2 into 4 copies can be seen. For slightly longer cells 4 clearly separated origin positions are visible (not shown).

In summary we show that independently of where the protein is expressed we observe the same nucleoid localization of LacI while at the same time LacI without DNA binding domain is evenly distributed and the genes have distinct localizations.

It can be argued that the reason why we do not observe spatial gradients is that we only can observe the molecule after which the fluorophore has matured, a process which takes longer than the time of diffusional mixing. One way to circumvent this is to treat the cells with formaldehyde, which would stop the molecules from diffusing and allow them time to mature at a stationary position before imaging<sup>4,5</sup>. We thus repeated the experiments above for LacI positioning after flowing formaldehyde into the microfluidic chip. To determine how fast this stops the molecules from diffusing we developed an assay to measure the diffusion rate of individual LacI-Venus as a function of time after adding the formaldehyde. This is done by temporally spaced rounds of single molecule tracking where a new field of view of cells is imaged (200 frames at 3.5 ms time resolution) every 4.5 s following the switch to formaldehyde. In Supplementary Figure 5e it is seen that most molecules stop moving within 1 min. Another approach to study the immobilization of molecules is to use long exposure times (500 ms). In formaldehyde untreated cells only a few spots are detected when there is no specific operator on the chromosome. When formaldehyde is added the number of spots increase as molecules become immobilized. Also here the time scale is 1-2 minutes (Supplementary Fig. 5e). A third assay is to follow the growth of individual cell histories in the microfluidic device before and after the addition of formaldehyde. It is observed that the cells stop growing apparently instantaneously after change of medium, that is in less than a minute given the time resolution in the experiment (Supplementary Fig. 5f). Altogether these experiments show that cross-linking starts within seconds of formaldehyde treatment and that most fluorescent protein seem to be completely immobilized after 1-2 minutes.

To get the 1.5 times higher excess concentration around the operator the non-fluorescent molecules would need to stay in the synthesis region about half a generation time. This means that the measured

fixation is sufficiently fast to preserve spatial gradients which are caused by molecules slowly leaving the region of synthesis. When analyzing the localization of LacI-Venus in formaldehyde treated cells they display distinct polar and mid cell localization both with and without DNA binding domain (Supplementary Fig. 5c, d). Since the pattern is the same for LacI that cannot interact with DNA, we consider this to be the result of formaldehyde treatment, meaning that no conclusions can be drawn from the spatial localization we observe in formaldehyde treated cells.

Another line of evidence against spatial gradients come from previous single molecule tracking of LacI-Venus<sup>6</sup> where it is seen that the fluorescent molecules move through the cells on the order of seconds, which would make it hard to maintain any region of high local concentration. These experiments were done in a strain with specific binding sites present, and to avoid binding to these sites IPTG was used. To test how LacI-Venus diffuses without IPTG present we have repeated the experiment in a strain without specific binding sites. Using the microfluidic device and segmentation of cells makes it possible to overlap 425 trajectories from approximately 500 cells of length between 2 and 3  $\mu\text{m}$  (Supplementary Fig. 6a). The average localization in the tracking experiment shows that most molecules are found in the nucleoid (Supplementary Fig. 6b); the mean square displacement (MSD) shows that the localization distribution broadens to 100 nm in <50 ms (Supplementary Fig. 6c); and the distribution of XY-step lengths after 9 ms shows that practically all molecules have moved more than 100 nm (Supplementary Fig. 6d). Taken together the tracking results imply that unless the newly synthesized non-fluorescent LacI-Venus interacts more strongly with DNA compared to a fluorescent LacI-Venus, it would not be possible to maintain a locally higher concentration around the site of synthesis. We find this possibility unlikely since the newly made LacI is a monomer that has to dimerize before it can bind DNA strongly.

We also made further controls, in which the *lacI* gene was moved to the mirror position on the other chromosome arm in order to express the LacI protein from another position in the cell without changing the gene dosage. The different chromosome arms have previously been shown to segregate to different parts of the cell<sup>7</sup>, and are therefore likely to be spatially separated. The repression ratio measured using a  $\beta$ -galactosidase assay is  $1410 \pm 200$  (mean  $\pm$  s.d.) when *lacI* is in the wild type position and  $1880 \pm 200$  (mean  $\pm$  s.d.) when *lacI* is moved to the opposite chromosome arm. i.e. there are no indications that spatial proximity would give a higher repression ratio which would be expected if there is a high local concentration of LacI around its gene expression site.

## Experimental procedures

### *Determination of chromosomal spatial locations in live cells*

To determine the localization of the *lacI* gene locus throughout the cell cycle, the strain JE116 (where LacI-Venus binds to a single *lacO<sub>I</sub>*) was studied using time-lapsed fluorescence microscopy. Phase contrast images, used for cell segmentation and lineage tracking, were taken every 30 seconds, while 4 s fluorescence images were taken once every 3 minutes to detect, count and localize individual specifically bound LacI-Venus molecules. The intensity of the excitation laser was set (<5W/cm<sup>2</sup>) so that the growth rate of the cells was not noticeably altered as compared to cells unexposed during the same experiment.

The tools used for image analysis have been extended from that previously described<sup>8</sup> to allow for totally automated cell segmentation and cell tracking of many thousand cells from phase contrast images (Fig. 1f, Supplementary Fig. 3). In addition, spots in the fluorescence images were automatically identified and localized as described in the Online Methods. In order to compare several cell lineages the cells need to be synchronized in time. A commonly used method to do this is to synchronize the cells from cell division to cell division. However, since there is a large variability in timing between cell divisions for individual cells, this causes a very broad distribution for the localization of individual *lacI* loci (Supplementary Fig. 3c). Instead, the positions of individual loci are

more structured when the cell cycles are synchronized by length (Supplementary Fig. 3a, b). We surmise that this may be related to a constant cell mass for initiation of DNA replication<sup>9</sup>.

The cell histories were analyzed in pairs of mother and daughter cells. The time when the mother cell reaches a reference length (4.25  $\mu\text{m}$ ) is set to 0 and the time the daughter cell reaches the same length is set to 1. The spots appearing between these two time points are assigned to a time within the interval [0, 1] (Supplementary Fig. 3a).

The locations of different loci were then determined in the following way. The chromosomal origin was detected using an array of 240 *lacO<sub>1</sub>* sites located close to the origin and LacI-Venus expressed from pBAD24-*lacI-venus*. Binding at terminus was detected by having two *malO* sites at their wild type positions and Mall-Venus chromosomally expressed (the strain BW25993 *malI::mall-venus* was a gift from the Xie lab). Binding at mid-replicore was detected using JE116 as described above, having a single *lacO<sub>1</sub>* site and with LacI-Venus being chromosomally expressed. The localization pattern was determined for all three loci by plotting the coordinates of detected spots in each case for the cell length interval 4.6-5.1  $\mu\text{m}$  (Supplementary Fig. 5b).

#### **Localization of LacI-Venus expressed from different positions in the chromosome in live cells**

Localization of LacI-Venus in cells without specific binding sites (Supplementary Fig. 5a) was performed in accordance to the protocol described previously<sup>8</sup> for imaging of non-specifically binding transcription factors, i.e. combining cell segmentation in phase contrast with short and intense fluorescence excitations every 3 min (2 ms exposures at 1.4kW/cm<sup>2</sup>). The strains encoding *lacI-venus* at different positions in the chromosome, MG1655 $\Delta$ lac *atpI::lacI-Venus* (origin), MG1655 $\Delta$ lac *atpI::lacI42-Venus* (origin), MG1655 $\Delta$ lac *ybbD::lacI-Venus* (mid replicore) and MG1655 $\Delta$ lac *essQ::lacI-Venus* (terminus), were kind gifts from Dr. Edward Cox and are described in<sup>4</sup>.

#### **Localization of LacI-Venus using fixation experiments**

In experiments detecting LacI position in formaldehyde treated cells we compare the strain MG1655 $\Delta$ lac *ybbD::lacI-Venus* (mid replicore) with the corresponding strain without DNA binding domain on LacI. The cells were grown in our switching microfluidic chip and fixed by changing medium to M9 glucose containing 1% formaldehyde. After 10 minutes of incubation, medium was changed back such that formaldehyde was removed and cells continuously washed for 30 minutes. Each trap was then imaged once in the fluorescence channel (500 ms exposure time) and once in phase contrast. Cells were segmented and spots were detected and mapped on a coordinate system as described above (Supplementary Fig. 5c, d).

To investigate the time scale of formaldehyde induced fixation (cross-linking) we made three different experiments (Supplementary Fig. 5e, f). In the first, different traps were imaged at different time points following the start of fixation (4.5 s interval). At each position a 3 ms exposure every 3.5 ms for a total of 200 frames provide movies of how the molecules stop diffusing in real-time. For each position spots were detected, trajectories built, and diffusion coefficients calculated as described below for live cell tracking. Data was pairwise binned from different time points to improve statistics in the MSD analysis. In the second experiment the cells were imaged for 500 ms before and at different time delays (20 s, 1, 3, 6 and 10 min) following the start of formaldehyde treatment. We also made time laps of growing cells, where they were imaged every 25 s before and after formaldehyde treatment. Using automated segmentation, each cell was analyzed as to see how long time after formaldehyde treatment it stopped increasing in length.

#### **Live cell tracking of LacI-Venus**

We performed live cell tracking of LacI-Venus at 25°C using an approach similar to the one described in<sup>6</sup>. However there the strain did have a specific *lac* operator and IPTG was added to remove specific binding. Here we use the strain JE107 (no operator site) (Supplementary Fig. 7a) where only non-

specific binding is possible and no IPTG is present. In addition we use microfluidics and cell segmentation to carefully quantify the xy-plots of the trajectories. For each group of cells (different field of view) 4-6 datasets were taken, in total 117. Each dataset correspond to a phase contrast image (used for cell segmentation), a bright field image (used to correlate the phase contrast image and fluorescence images) and a 500 frames long movie with a sampling time between images of 3 ms and laser exposure times of 1 ms. The power in the sample plane was  $650 \text{ W/cm}^2$ . Spots were detected and localized using the previously described spot detection and a weighted centroid calculation (see Online Methods). Spots were connected into trajectories if they were closer than 312 nm between subsequent frames (corresponding to include jumps from a  $3\mu\text{m}^2\text{s}^{-1}$  fast diffusing particle with a confidence of 90%) and if there was only one possible connection within this distance. The trajectories were then assigned to a cell from the cell segmentation and only data from within cells with lengths in range 2-3  $\mu\text{m}$  and widths of 1.1-1.2  $\mu\text{m}$  were included in the further analysis (Supplementary Fig. 6).

***Repression as a function of lacI position on the chromosome***

We did  $\beta$ -galactosidase experiments where wt LacI regulates operon expression through the binding at its operators (*lacO*<sub>1</sub> and *lacO*<sub>2</sub>) at their native positions just downstream of the *lacI* gene itself (strain JE123), or where the *lacI* gene has been moved to the other chromosome arm (strain JE131) (Supplementary Fig. 7b).

## 6. STRAIN CONSTRUCTION

Strains JE123 (BW  $\Delta lacO_3::lacO_3^-$ ) and JE124 (JE12  $\Delta lacO_1::lacO_1^- \Delta lacO_1::lacO_{sym} \Delta lacO_2::lacO_2^-$ ) were constructed by  $\lambda$  Red recombination of a PCR fragment<sup>10</sup> containing the desired modified *lac* promoter sequence generated by overlapping PCR into, respectively, BW  $\Delta(P_{lac-lacZNter})::Tet^R$  (JE111, cf<sup>d</sup>) or JE12  $\Delta(P_{lac-lacZNter})::Tet^R$  (JE113, cf<sup>d</sup>), which both contain the pKD46 plasmid. The two resulting strains were selected on M9 lactose plates. The constructs were moved into a clean background (JE111 or JE113) by P1 phage transduction and new selection on M9 lactose.

Strains JE127 and JE128 were generated from JE106 in the following way: the relevant sequence of the chromosomal *lac* region of JE106 (*lacI-venus*, *lacO<sub>sym</sub>-tetO<sub>2</sub>* and ~1kb homology on both sides) was cloned into the pGEMt vector and modified by PCR to introduce the mutation (GAC->AAC) in the *lacI* gene at position 820-822. The resulting strain expresses the LacI D274N variant (LacI<sub>s</sub>, repressor mutant with diminished IPTG-affinity)<sup>11</sup>. The construct was further modified by replacing the *lacO<sub>sym</sub>* sequence with *lacO<sub>1</sub>*. The two constructs containing *lacO<sub>sym</sub>* and *lacO<sub>1</sub>* respectively, were then transferred to the temperature-sensitive plasmid pKO3 in between two NotI restriction sites and chromosomally integrated<sup>12</sup> to replace the corresponding sequences in JE106.

Strains JE126, JE129, JE130 and JE125 are derivatives of strains JE106, JE128, JE118 and JE112, respectively, overexpressing TetR from the chromosome. A PCR fragment, containing the *tetR* gene under control of the P<sub>2<sub>rrnB</sub></sub> promoter, the FRT-Cm<sup>R</sup>-FRT selection marker, and *intC* homology sequences on both ends; was integrated using  $\lambda$  Red recombination<sup>10</sup> in the strain JE12<sup>6</sup>, which contains the pKD46 plasmid. The resulting strains were selected for positive integrants on LB-chloramphenicol plates. The constructs were finally moved into strains JE106, JE128, JE118 and JE112 by P1 phage transduction and new selection on LB-chloramphenicol plates.

Strain JE131 was constructed by deleting the *lacI* gene together with its promoter P<sub>lacI</sub> in JE123 background (*lacO<sub>3</sub>*-) to generate  $\Delta(P_{lacI-lacI})::Tet^R$  and inserting P<sub>lacI-lacI</sub> Cm<sup>R</sup> on the opposite side of the chromosome relative to *oriC*, in *ygaY'* pseudogene (60.5'). Both antibiotic cassettes were removed by FLP recombinase, leaving FRT scars<sup>10</sup>.

The *lac* operator array (*zie-3627::lacZo(240X)::kan<sup>R</sup>*) located near the origin was transduced with phage P1 from the ILO1<sup>13</sup> strain (ordered from the *E. coli* Genetic Stock Center) into BW25993 background. The final strain was then transformed with the plasmid pBAD24-*lacI-venus* (see below).

pBAD24-*lacI* was made by inserting the *lacI* gene under the control of the L-arabinose-inducible P<sub>BAD</sub> promoter<sup>14</sup>, between NcoI and HindIII restriction sites in the low copy number pBAD24 vector. pBAD24-*xyIR* was previously generated in the same way<sup>1</sup>. pBAD24-*lacI-venus* was made in a similar way by inserting the *lacI-venus* fusion between EcoRI and HindIII restriction sites.

All strains are depicted in Supplementary Figure 7. For the operator sequences used, see Supplementary Table 1.

## 7. GROWTH CONDITIONS AND EXPRESSION ASSAYS

In all experiments other than those described in the Online Methods section (i.e. Western blot,  $\beta$ -galactosidase assay and fluorescence microscopy using poly-L-Lysine coated dishes) saturated cultures were diluted 1:500, collected after 4-5 hours growth (at  $OD_{600}\sim 0.15-0.2$ ), and concentrated by centrifugation. For expression studies at room temperature, saturated cultures were diluted 1:100 and collected after 10-12 hours growth.

Western blots were done using standard procedures with primary antibody Anti-LacI (1:1000) (Millipore) and secondary antibody peroxidase-conjugated Goat Anti-Mouse IgG and (Jackson Immuno Research).

To measure how the expression of *lacI* from pBAD24 is up-regulated by arabinose, the strain JE127 harboring pBAD24-*lacI* were grown to  $OD_{600}\sim 0.2$ , kept in a 37°C water bath and induced with 0.2% arabinose. Protein synthesis was blocked at different time points by removing aliquots and mixing with chloramphenicol in pre-chilled tubes which were kept on ice.

$\beta$ -galactosidase activity assays were performed according to <sup>15</sup>. For induction of LacI or TetR regulated promoters, 1mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) or 100ng/ml anhydrotetracycline (aTc) (Sigma) were added to the cultures upon dilution of saturated cultures.

In fixation experiments a fresh solution of 1% formaldehyde in culture media was prepared at the same day as the experiment, using a 37% formaldehyde stock solution from Sigma.

## 8. SUPPLEMENTARY REFERENCES

- 1 Hammar, P. *et al.* The lac repressor displays facilitated diffusion in living cells. *Science* **336**, 1595-1598, doi:10.1126/science.1221648 (2012).
- 2 Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature methods* **9**, 671-675 (2012).
- 3 Goldman, S. R., Ebright, R. H. & Nickels, B. E. Direct detection of abortive RNA transcripts in vivo. *Science* **324**, 927-928, doi:10.1126/science.1169237 (2009).
- 4 Kuhlman, T. E. & Cox, E. C. Gene location and DNA density determine transcription factor distributions in Escherichia coli. *Molecular systems biology* **8**, 610, doi:10.1038/msb.2012.42 (2012).
- 5 Little, S. C., Tkacik, G., Kneeland, T. B., Wieschaus, E. F. & Gregor, T. The formation of the Bicoid morphogen gradient requires protein movement from anteriorly localized mRNA. *PLoS biology* **9**, e1000596, doi:10.1371/journal.pbio.1000596 (2011).
- 6 Elf, J., Li, G. W. & Xie, X. S. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191-1194, doi:10.1126/science.1141967 (2007).
- 7 Wang, X., Liu, X., Possoz, C. & Sherratt, D. J. The two Escherichia coli chromosome arms locate to separate cell halves. *Genes & development* **20**, 1727-1731, doi:10.1101/gad.388406 (2006).
- 8 Ullman, G. *et al.* High-throughput gene expression analysis at the level of single proteins using a microfluidic turbidostat and automated cell tracking. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **368**, 20120025, doi:10.1098/rstb.2012.0025 (2013).
- 9 Hill, N. S., Kadoya, R., Chatteraj, D. K. & Levin, P. A. Cell size and the initiation of DNA replication in bacteria. *PLoS genetics* **8**, e1002549, doi:10.1371/journal.pgen.1002549 (2012).
- 10 Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 6640-6645, doi:10.1073/pnas.120163297 (2000).
- 11 Chang, W. I., Barrera, P. & Matthews, K. S. Identification and characterization of aspartate residues that play key roles in the allosteric regulation of a transcription factor: aspartate 274 is essential for inducer binding in lac repressor. *Biochemistry* **33**, 3607-3616 (1994).
- 12 Link, A. J., Phillips, D. & Church, G. M. Methods for generating precise deletions and insertions in the genome of wild-type Escherichia coli: application to open reading frame characterization. *Journal of bacteriology* **179**, 6228-6237 (1997).
- 13 Lau, I. F. *et al.* Spatial and temporal organization of replicating Escherichia coli chromosomes. *Molecular microbiology* **49**, 731-743 (2003).
- 14 Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of bacteriology* **177**, 4121-4130 (1995).
- 15 Miller, J. H. *A short course in bacterial genetics : a laboratory manual and handbook for Escherichia coli and related bacteria.* (Cold Spring Harbor Laboratory Press, 1992).



Paper IV 



# Fluctuations in replication initiation determine the generation time and size distributions of *E. coli* cells

Mats Walldén<sup>a</sup>, David Fange<sup>a</sup>, Gustaf Ullman<sup>a</sup>, Erik G. Marklund<sup>a,b</sup>, Johan Elf<sup>a,c</sup>

- a. Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, 75124 Uppsala, Sweden
- b. Current affiliation: Department of Chemistry, University of Oxford
- c. Corresponding author: johan.elf@icm.uu.se

***E. coli* can grow and divide faster than the time required to replicate its genome. This is accomplished by using overlapping replication cycles where multiple parallel replication processes are initiated in the mother or even grandmother generation. In addition, cells living under constant growth conditions will vary considerably in their cell cycle times and division sizes. In spite of these fluctuations, cells have to terminate one round of replication during each cycle. It has been unclear how cells accomplish this without cell cycle checkpoints which, in turn, would be inconsistent with overlapping replication cycles. Using single molecule fluorescence microscopy to localize and track the replisome-protein DnaQ throughout thousands of bacterial cell cycles, we find that initiation of replication occurs at a relatively fixed volume per number of origins independently of the time from division. This ratio is also constant for different growth conditions. Further, we find that a simple model of the bacterial cell cycle, where division occurs a constant time after replication initiation, accurately reproduces the variations in timing and sizes at divisions.**

## Significance

Replication of DNA in *Eukaryotic* cells is tightly coordinated with the cell division cycle via checkpoints that cannot be passed until previous steps have been completed. In contrast, rapidly growing bacteria cannot have this type of checkpoints since they need to use overlapping replication cycles in order to achieve a generation time shorter than the time it takes to complete one round of replication. Using single molecule imaging of individual replication machineries throughout the lifespan of thousands of bacterial cells, we identify a simple regulatory principle that coordinates DNA replication and cell division. The scheme allows each cell to grow exponentially without checkpoints and ex-

plains the large variation observed in the generation time and cell size for genetically identical bacteria.

## Introduction

In the Cooper-Helmstetter model the *E. coli* cell cycle is regarded as a deterministic process, where all cells in a population grow at the same rate, initiate replication at a same time after birth and divide after the same time, as determined by the growth rate [1]. Replication of the *E. coli* chromosome proceeds bi-directionally from the origin of replication locus, *oriC*, to the terminus locus, *ter*, and the duration of one round of replication, the C-period, is reported to last ~40 minutes independently of growth conditions [1]. The period between the termination of one round of replication and the completion of the next division event is referred to as the D-period and is reported to last ~20 min [1]. This model conceptualizes how exponentially growing cells can maintain a stable cell cycle which is shorter than the C-period by initiating and terminating one round of replication per cycle. The key to accomplishing this is by maintaining parallel rounds of replication which are initiated and terminated during different cycles.

Studies of individual bacteria have revealed that isogenic cells under constant growth conditions vary considerably in size and division timing [2, 3]. How stable cell cycles are maintained in the midst of these fluctuations is not yet well understood. Assuming that each cell grows as fast as possible given the available nutrients, the molecular control systems responsible for establishing the timing of replication initiation and division must be able to match the synthesis of new chromosomes with the growth rate so that there is at least one chromosome in each daughter cell after division. Given the cell to cell variability, the question arises whether initiation of DNA replication is triggered at a certain cell size or time after division, or if it is division that is triggered at a certain time after replication?

From the deterministic view of the Cooper-Helmstetter model, cell size and time since the last division are perfectly coupled and a distinction between a timer or a sizer to control replication or division is not required. In 1968 Donachie reported that the initiation of replication in *E. coli* occurs at a critical level of cell mass per number of origins [4]. However, more recently it was concluded that replication was initiated according to the time after last division based on experiments in slowly growing cells that were synchronized using the baby cell column [5, 6]. Later, Hill *et al.* proposed that replication is initiated according to size. This statement was based on studies of size-altered mutants of *E. coli* under conditions of fast growth [7]. Further, Osella *et al.* suggested that division timing in cells growing in a microfluidic device under conditions of fast growth was best described as a combination of a sizer and a timer and defined a phenomenological model to this end [2, 8].

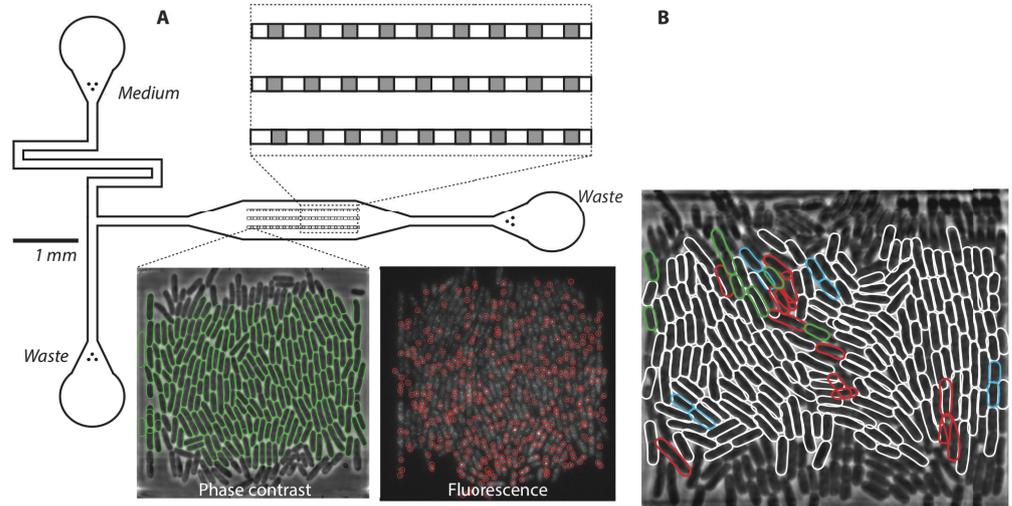
In order to reconcile how the *E. coli* cell cycle is coordinated at different growth rates, we have studied the replication machinery of *E. coli* using time-lapsed single molecule fluorescence microscopy in fast and slow growing cells. We clarify the spatiotemporal organization of the *E. coli* replisomes by superimposing the locations of detected replisomes on the intracellular coordinate system spanned by each cell outline at each stage of the cell cycle. The integrated method allows for the automated acquisition of an arbitrary number of obser-

vations from a single experiment.

## Results

**Cell segmentation and tracking.** To study exponentially growing cells we used a microfluidic device [3] (Fig 1A) in which bacterial micro-colonies are maintained under constant conditions for days at a time. In order to outline and track cells in time-lapsed images we used a modified version [3] of the software package MicrobeTracker [9], which we have extended for this study. Machine learning methods are used to enhance the algorithms for cell tracking and outlining by detecting incorrectly outlined cells. The algorithms have been trained with examples of correctly and incorrectly outlined cells that have been identified manually (Materials and Methods). The tracking method now identifies and removes incorrect outlines and refills the corresponding gaps (Fig. 1B), which substantially improves both the throughput and accuracy. We estimate that roughly 90% of all cells in an imaged micro-colony are both outlined and tracked from one frame to next such that >60% of all complete cell cycles are captured (Supplementary Text).

**Precision of cell division timing and growth rate determination.** To validate the accuracy of our division classifier based on phase contrast we have benchmarked the method against division timing obtained by studying cells carrying a fluorescently labelled division protein FtsQ-GFP [10]. FtsQ is required for the con-



**Fig. 1: Overview of the microfluidic setup and cell tracking with error handling.** (A) Media is continuously flowed through the chamber containing 51 traps aligned in 3 rows. Each trap can harbor a micro-colony of up to about 300 bacterial cells for the growth conditions used in this study. Cells are lost in the media-flow above and below the trap when the micro-colony expands in size. (B) Cells in a trap together with outlines of identified cells using the automatic cell-tracking algorithm. In each step of the cell-tracking method, cell-outlines may be identified as incorrect (red) and removed. Empty space due to removal of incorrect cell outlines is re-filled with new estimates of outlines (green). The cell tracking algorithm also identifies divided cells in each frame (blue).

traction of the bacterial membrane that ultimately leads to division [10]. When studied using fluorescence microscopy, FtsQ-GFP can be observed to form a band at the septum, which contracts during division. As the contraction ceases, the foci vanishes abruptly (Fig 2A and B) [10]. The difference between divisions as defined by the disappearance of FtsQ-GFP foci and division according to our division classifier based on phase contrast images has a standard deviation of 2.7 minutes (Fig. 2C), which is sufficiently precise for our purposes.

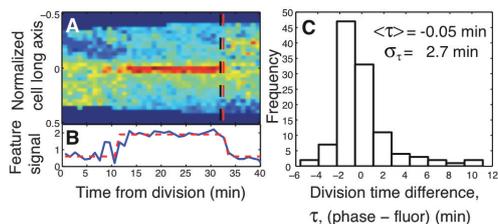
**Spatiotemporal organization of replisomes.** Actively replicating DnaQ-Ypet molecules are sufficiently immobile to form diffraction-limited spots when imaged using exposures of 1 second (Fig. 1A) [11]. For each time point we obtain the cell size and the intracellular locations of detected replisomes. To understand the organization of replisomes we organize our data according to time since last division or according to cell size. In the first case, we construct a spatiotemporal distribution of the time from the last division until the detection of the replisome and the intracellular locations of the replisome as measured along the major axis of the cell. In the second case, we organize our data according to cell size. Here we construct a joint distribution of cell volumes at replisome detection and the intracellular replisome location.

**Replication initiation under conditions of slow growth.** To study replication initiation under conditions without overlapping rounds of DNA replication, we grow cells in M9 minimal medium supplemented with 0.4% succinate and amino acids. Further, we maintain the microfluidic chip at 30°C. We find that cells grow with the average rate  $0.59 \text{ h}^{-1}$  (SEM:  $0.003\text{h}^{-1}$ ) and that the standard deviation of the cell growth rate is  $0.17 \text{ h}^{-1}$  (Fig. S3). The cell length and width are on average  $3.36 \mu\text{m}$  and  $0.92 \mu\text{m}$  respectively (Fig. S3) implying a cell volume of  $2.4\mu\text{m}^3$ . We find that for short or young

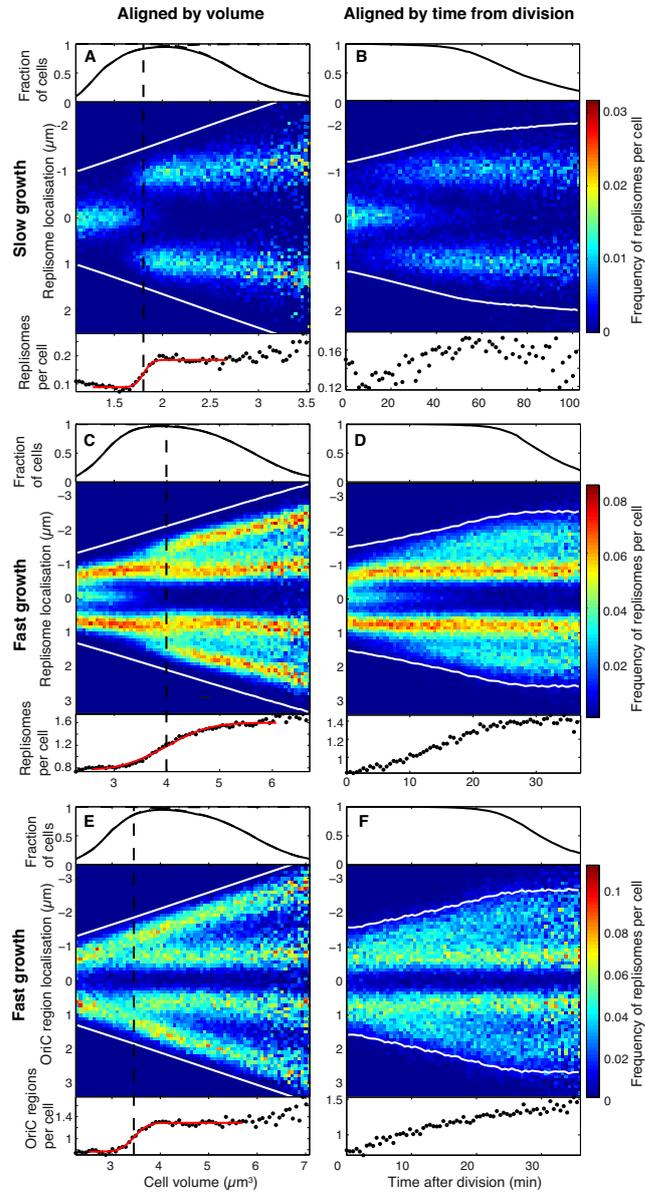
cells, replisomes are localized to a band at mid-cell (Fig. 3A and B). We interpret this band as two replication forks approaching termination at which point no spots can be detected. In these replication sites, the two replisomes are spatially too proximate to be resolved. After essentially all cells have terminated replication, a new round of DNA replication initiates at  $\frac{1}{4}$  and  $\frac{3}{4}$  of the cell length. Although it is possible to identify DNA replication initiation from the figure where different cells have been aligned based on the absolute time from the last division (Fig 3B), the increase in the mean number of detected replisomes per cell (Fig. 3A, bottom) is substantially more distinct if the cells are aligned by cell size (Fig. 3A and B). Also note that when aligning by size, the identified replisomes from the different cells will fall on the same intracellular position paths although the different cells are born and divided at very different sizes (Fig. 3A, top). These observations suggest that initiation of DNA replication is coordinated by cell size rather than time from division and that this initiation occurs as cells reach a size of about  $1.8 \mu\text{m}^3$  where the average number of detected replisomes per cell has its half max value.

**Replication initiation under conditions of fast growth.** To study the more complicated case of cells with overlapping rounds of replication we used M9 minimal medium supplemented with amino acids, 0.4% glucose and kept at 37°C. These cells grow with the average rate  $1.29 \text{ h}^{-1}$  (SEM:  $0.003\text{h}^{-1}$ ) and that the standard deviation of the cell growth rate is  $0.26 \text{ h}^{-1}$  (Fig. S3). The cells have an average length and width of  $4.23 \mu\text{m}$  and  $1.15 \mu\text{m}$  respectively, corresponding to an average volume of  $4.9 \mu\text{m}^3$ . By again aligning according to size or time from last division, we find that more distinct paths emerge only when cells are aligned according to size as in the case of slowly growing cells (Fig. 3C and D). This size dependence shows that initiation of DNA replication is regulated according to size, independent of growth rate. After aligning by size we find that for short cells, there are three replication sites in which there are high replisome density. We interpret this as two replication forks close to termination in the middle of the cell as was also the case for slow growing cells and two additional forks on either side of the middle one. At around  $4 \mu\text{m}^3$  there is an appearance of two new replication sites (Fig. 3C) coupled with an increase in the mean number of identified replisomes per cell (Fig. 3C, bottom). The fact that only two new sites appear suggests that this is due to close spatial proximity between the already replicating forks with one of the newly initialized such that they cannot be spatially resolved, which is also in line with previous observations [12].

**Origin localization in rapidly growing cells.** The replisome data for rapidly growing cells is not as clear-cut in terms of the initiation size as was the case for the slow growing cells. Another way of studying replica-



**Fig. 2: Timing of disappearance of FtsQ signal.** (A) Example of FtsQ fluorescence average along long-axis of a single cell at different time-points. Identified drop in fluorescence is shown as dashed red vertical lines. Division based on phase contrast is shown in dashed black vertical lines (B) Time trace of long-axis average (C) Histogram of difference of timing between fluorescence drop and division identified by phase contrast.



**Fig. 3: Intracellular positions along the long-axis for replisomes in slow (A, B) and fast (C, D) growing cells together with positions for oriC regions (E, F) in fast growing cells.** Cells that have been tracked for a complete generation are included in the analysis and are here aligned by their volume (A, C and E) or by the time from when they last divided (B, D and F). In all panels the (*top*) part shows the fraction of all analyzed cells that are included in the volume (A, D, and G) or time (B, C, E) segment. The total number of analyzed cells is 5352 for replisomes at slow growth (A, B), 13003 for replisomes at fast growth (C, D) and 5379 for oriC region at fast growth (E, F). The (*middle*) part of each panel show frequencies of cells in certain intracellular positions for given cell volume (A, C, and E) or for given times (B, D and F). The frequencies are normalized such that when each volume or time segment is summed for all intracellular positions it gives the average number of detected replisomes (A-D) or oriC regions (E, F) as shown on the (*bottom*) of each panel. Red curve is a fit to a Gauss error function over the data in the region covered by the red curve. Dashed vertical lines show the half max of the error function fit. White solid lines indicate average cell long-axis length.

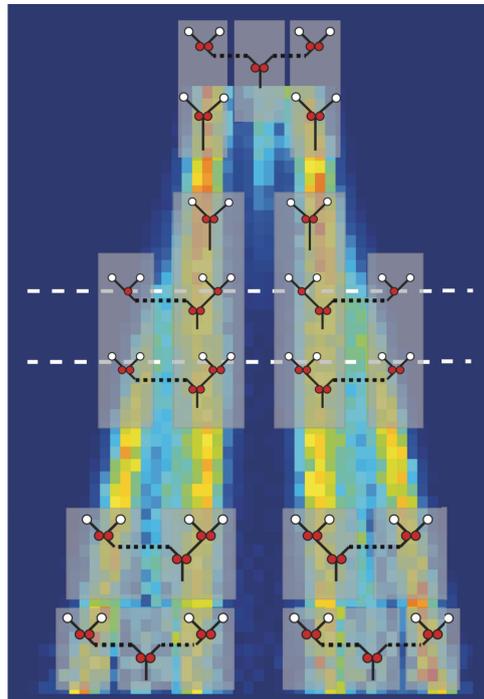
tion initiation is to monitor the copy number and position of the origin region of the chromosome. We do this by introducing two *malO* binding-sites for the Mall transcription factor in the *bgLG* locus close to *oriC*. The advantage of using a single operator pair as compared to an array of sites that was used in earlier studies is that it does not obstruct replication elongation [13]. The locus is detected by using a fluorescently labelled Mall (Mall-Venus) expressed from the *bgLG* locus. The spatial and temporal positions of paths of high origin density (Fig. 3E and F) are consistent with the replisome data for fast growing cells, since there needs to be at least one origin where replication initiates. However, there is a sharp increase in the average number of detected origins and splitting of the paths already at  $3.5 \mu\text{m}^3$ . This is surprising considering that the replisome should assemble before the origins are replicated (Fig 3E). A tentative solution to this apparent inconsistency is presented in the discussion.

## Discussion

We have presented data supporting the idea that initiation of DNA replication is regulated according to cell size [4] rather than by time from last division, as have been previously suggested [5, 6]. We find that in slowly growing cells ( $0.59 \text{ h}^{-1}$ ) the initiation of DNA replication occurs at  $1.8 \mu\text{m}^3$ . In fast growing cells ( $1.29 \text{ h}^{-1}$ ) we find that initiation of replication can occur either at  $3.5 \mu\text{m}^3$  or at  $4.0 \mu\text{m}^3$ . To explain the seeming contradiction between the replisome and origin data, we propose a model which reconciles these observations.

In our model (Fig 4), initiation of replication occurs when the origins split at  $3.5 \mu\text{m}^3$ . The reason for not observing a strong increase in the number of replisomes at  $3.5 \mu\text{m}^3$  is that replication in different directions from the same origin start at different times. This time difference was previously estimated to 7 minutes using qPCR at different chromosomal loci [14]. The *bgLG* locus, which we use as a proxy for *oriC*, is located on the chromosome arm that is replicated first. Based on this model a 50% increase in the number of replisomes detected is expected at  $3.5 \mu\text{m}^3$ , *i.e.* assembly of replisomes necessary for replication in one direction only. Accordingly, we find an increase in the average number of replisomes at smaller sizes than  $4.0 \mu\text{m}^3$ . In addition the split of the replisome paths and the increase in average number of replisomes per cell seen at  $4.0 \mu\text{m}^3$  is, in this model, due to initiation of DNA replication in the other, time-delayed, direction.

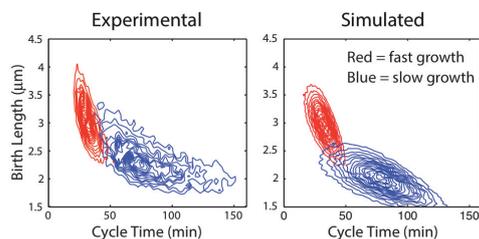
Based on the result presented for the volumes at which initiation occurs we may ask if this can be explained by the classical model suggested by Donachie [4] in which initiation of replication occurs at a fixed size to DNA content ratio. By dividing the volumes at which initiations occur with the expected copy number of *oriCs* we find that this ratio is similar for cells under conditions supporting fast and slow growth. For the fast growth



**Fig. 4: Cartoon model of initiation of DNA replication.** White circles show origins. Red circles show replication forks, where a single red circle show that replication fork is progressing in one direction and pair of red circles show replication forks progressing in two directions. Black lines are chromosomes. Dashed lines are drawn at  $3.45$  and  $4 \mu\text{m}^3$ . Gray boxes indicate regions of spatial co-localization of *oriC* and replisomes.

condition the ratio is  $3.5/4 = 0.87$  and for the slow growth condition it is  $1.8/2 = 0.9$ , which is in line with Donachie's original model in which initiation of DNA replication occurs at a constant volume to origin ratio.

It has recently been discussed whether cell division is based on cell size or time after last division [8]. It was concluded that cells use a combination of both a timer and sizer in order to decide when to divide. The simplest model consistent with the data presented here assumes that initiation of DNA replication occurs at a constant cell volume per number of origins and that cell division occurs at a subsequent constant time delay where the time delay is given by the sum of the C and D periods. This would correspond to a sizer, which is subsequently followed by a timer. Using this simple model we can accurately describe the variation in cycle times and birth sizes of cells using only the variability in the cell size at replication initiation. Using our estimates of initiation sizes and their uncertainties together with estimated for the C and D periods (Materials and Methods in Supplementary Information), the model



**Fig. 5: Length at birth as function of cycle time for both experiments and simulations.**

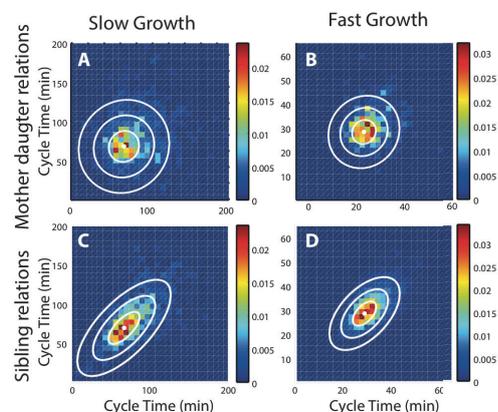
reproduces the observed relation between growth rate and size at division (Fig. 5). The reason for the difference in cell cycle variability between the slow and fast growth conditions is that slow growing cells spend a considerably longer time in any given size interval. Variability in initiation, which is size-dependent, thus turns into larger variability in initiation time in the cells growing slow as compared to the fast growing cells.

A further prediction from our model is that siblings are expected to divide at the same time. In agreement with this prediction we find a greater correlation in the time from division to division between siblings as compared to mothers and daughters (Fig. 6). The reason for this is that variability in initiation timing is shared by siblings but not by consecutive generations. In cases where the initiation and termination of the same round of replication occurs during the same cell cycle, *i.e.* at very slow growth, the model predicts that the time from division to division should be uncorrelated also for siblings.

In conclusion we have shown that initiation of chromosome replication in *E. coli* is triggered at a fixed cell size per number of origins and not at a specific time after cell division. This is most easily explained by a titration mechanism where an inhibitor of replication is diluted below a threshold at which point replication is initiated. The inhibitor can for example be the *datA* locus on the chromosome that sequesters the replication activator DnaA. If we further assume a constant time period between initiation of replication and cell division, the small variation in initiation size propagates to establish the distributions of division size and generation time observed for different growth rates. Taken together this explains how individual *E. coli* can grow exponentially as fast as is supported by the growth media despite dividing at very irregular time intervals and sizes.

## Materials and Methods

**Strains and Media.** For replisome detection an *E. coli* strain JJC5350, in which DnaQ is genetically fused to a fluorophore, Ypet [15] was studied. To detect and localize the origin region of the chromosome a strain was constructed in which a gene encoding a fluorescent reporter construct, Mall-Venus, and two operator sites to



**Fig. 6: Correlations in cycle time for fast and slow growth conditions.** Correlations between cells in two subsequent generations (A-B) and between siblings in the same generation (C-D). The two dimensional cycle time frequency histogram are fitted to two dimensional normal distributions.

which this construct will bind, was introduced at the *bgLG* locus. In the same strain the native operator sites and the native *malI* gene were deleted. For timing divisions a strain carrying a reporter construct FtsQ-GFP was studied [10].

Cells were grown in M9 minimal medium with a supplement of amino acids (RPMI 1640 sigma Aldrich). Also a surfactant, Pluronic F108 (Sigma Aldrich), was added to a final concentration of  $0.85 \text{ gL}^{-1}$ . For conditions fast growth, the carbon source in the medium was glucose (0.4%) and during conditions of slow growth the carbon source was succinate (0.4%). A temperature regulated incubation hood maintained the cells at  $37^\circ\text{C}$  for fast growth and  $30^\circ\text{C}$  for slow growth.

**Microfluidic setup.** Fabrication and setup of the microfluidic devices was performed as in Ullman-Walldén 2012 [3].

**Image acquisition.** The imaging was performed using an inverted microscope (Nikon Ti-E) similar to that described in [3]. Our Nikon Ti-E has external phase contrast and thus the phase contrast images were detected on another camera (Scion CFW-1312M, Scion Corp.) than fluorescence and brightfield images (Ixon EM+, Andor). We imaged 6 micro-colonies containing exponentially growing *E. coli* cells in phase contrast every 30s and in fluorescence/brightfield every 3 min for at least 8.3 hours. Cells were illuminated using a 514 nm laser (Coherent Genesis CX 514-2000 STM) (Venus and Ypet) or a 488 nm laser (GFP) (Cobolt MLD™).

**Cell segmentation and Tracking.** The classifiers used in segmentation and tracking are outlined in Figure S1. The algorithms corresponding to blocks 2-6 in the flow chart in Figure S1 are described in Ullman-Walldén et al 2012 [3] and blocks 1, 7 and 8 are described below.

To detect cells in phase contrast images, the segmentation algorithm of the *microbeTracker* suite is applied using 10 different edge detection parameter sets. A classifier is used to identify incorrectly segmented cells and the different segmentations are combined into one. For classifying segmented cells, a coordinate system (Figure S2) of 44x17 points was created in order to define an image of a cell in a standardized framework, independent of cell length and curvature. The cell coordinate system uses a fixed grid of 8x17 points at both cell poles. The other 18x17 points are stretched depending on the cell length. In addition, 80 Fourier descriptors [16] are used to describe the shape of the cell contour. Also, the cell angle, with 0 degrees defined as parallel to the x-axis, is used as a feature. In total, 829 features are used for the segmentation error classifier. Dimension reduction using principal component analysis was tried, but yielded a worse result. A set of 3276 training examples of erroneously segmented cells together with a set of 7818 correctly segmented cells is chosen manually. The classification tree inherent in MATLAB is trained with these examples. The same cell can be detected and approved in more than one of the sets of segmented cells and it is necessary to decide with segmentation is the best. For this purpose the positive training examples are described as a multivariate Gaussian model. By using this model, cells are given a probability for belonging to the distribution of correctly segmented cells. The segmentation out of the 10, which is most probable of being correct, is then selected. When the complete set of trackers and error detectors (Blocks 2-6 in Figure S1) is finished, the segmentation algorithm (described in the paragraphs above) is re-executed in order to replace eventual cells lost in the tracking (Block 7 and 8, Figure S1). To save time the re-segmentation is performed only at positions where there are lost cells and cells disapproved by the segmentation error classifier. Again, the cell is selected that has highest probability according to the distribution of correctly segmented cells.

**Single molecule detection and localization.** Identification of single molecules was carried out as described previously [11] using a wavelet plane decomposition based method. In order to assign fluorescent molecule to cells, a coordinate system was established that maps a position in the phase contrast image to a corresponding position in the fluorescence image. For this purpose we used a brightfield image captured on the same camera as the fluorescence image. The brightfield image was aligned using normalized cross correlation with the phase image in order to align the fluorescence images with the phase image. With the global coordinate sys-

tem established, each dot could then be assigned to a local coordinate system inside the cell.

## References

1. Cooper, S. and C.E. Helmstetter, *Chromosome replication and the division cycle of Escherichia coli B/r*. J Mol Biol, 1968. **31**(3): p. 519-40.
2. Wang, P., et al., *Robust growth of Escherichia coli*. Curr Biol. **20**(12): p. 1099-103.
3. Ullman, G., et al., *High-throughput gene expression analysis at the level of single proteins using a microfluidic turbidostat and automated cell tracking*. Philos Trans R Soc Lond B Biol Sci. **368**(1611): p. 20120025.
4. Donachie, W.D., *Relationship between cell size and time of initiation of DNA replication*. Nature, 1968. **219**(5158): p. 1077-9.
5. Bates, D., et al., *The Escherichia coli baby cell column: a novel cell synchronization method provides new insight into the bacterial cell cycle*. Mol Microbiol, 2005. **57**(2): p. 380-91.
6. Bates, D. and N. Kleckner, *Chromosome and replisome dynamics in E. coli: loss of sister cohesion triggers global chromosome movement and mediates chromosome segregation*. Cell, 2005. **121**(6): p. 899-911.
7. Hill, N.S., et al., *Cell size and the initiation of DNA replication in bacteria*. PLoS Genet, 2012. **8**(3): p. e1002549.
8. Osella, M., E. Nugent, and M.C. Lagomarsino, *Concerted control of Escherichia coli cell division*. Proc Natl Acad Sci U S A, 2014. **111**(9): p. 3431-3435.
9. Sliusarenko, O., et al., *High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics*. Mol Microbiol, 2011. **80**(3): p. 612-27.
10. Soderstrom, B., et al., *Disassembly of the divisome in Escherichia coli: evidence that FisZ dissociates before compartmentalization*. Mol Microbiol, 2014. **92**(1): p. 1-9.
11. Hammar, P., et al., *Direct measurement of transcription factor dissociation excludes a simple operator occupancy model for gene regulation*. Nat Genet, 2014. **46**(4): p. 405-8.
12. Youngren, B., et al., *The multifork Escherichia coli chromosome is a self-duplicating and self-segregating thermodynamic ring polymer*. Genes Dev, 2014. **28**(1): p. 71-84.
13. Possoz, C., et al., *Tracking of controlled Escherichia coli replication fork stalling and restart at repressor-bound DNA in vivo*. EMBO J, 2006. **25**(11): p. 2596-604.
14. Joshi, M.C., et al., *Escherichia coli sister chromosome separation includes an abrupt global transition with concomitant release of late-splitting intersister snaps*. Proc Natl Acad Sci U S A, 2011. **108**(7): p. 2765-70.
15. Reyes-Lamothe, R., D.J. Sherratt, and M.C. Leake, *Stoichiometry and architecture of active DNA replication machinery in Escherichia coli*. Science, 2010. **328**(5977): p. 498-501.
16. Zahn, C.T. and R.Z. Roskies, *Fourier Descriptors for Plane Closed Curves*. Ieee Transactions on Computers, 1972. **C 21**(3): p. 269-&.



# Supplementary Information

## Materials and Methods

### Detection of FtsQ-GFP signal drop in single cells

To identify the time point of the decrease in FtsQ-GFP signal the following steps were taken. To obtain an intensity profile along the long axis, the intensity values along the short axis of the cell outline were averaged for each point along the long axis of the cell. The standard score, *i.e.*  $z=(x-\text{mean}(x))/\text{std}(x)$ , was computed for this profile, where  $\text{mean}(x)$  and  $\text{std}(x)$  are the mean and standard deviation of the values of the profile. The long axis of the cell was then scaled to a relative length, *i.e.* to span from -0.5 to 0.5. The profile values at each time point were then added to an intensity distribution (Fig. 2A). This distribution was subsequently Fourier transformed along the horizontal direction as shown in (Fig 2A). The FtsQ-GFP bands were identified as the most pronounced peak in the resulting power spectrum. A weighted average of the intensity along the long-axis for each time point yields an initial time-trace. Here the weights were a Gaussian function with average at the detected mid of the band and standard deviation of 1 pixel on the camera. Onto the initial time-trace we multiply a Gaussian error function, which gives the final trace (Fig 2B of the main text, blue solid line). To find the drop in signal intensity a t-test score was calculated for each possible combinations of duration in the high and low state respectively. The combination with the highest score was then selected (Fig. 2B of the main text, red dashed line) and the time when this description goes from the high state to the low state is used as drop in FtsQ-GFP signal.

### Cell-growth simulations

Simulations of cell growth based on initiation of DNA replication and cell division was carried out according to Algorithm 1. Based on data from slow growing cells we set the average initiation length per oriC copy number to 1.5  $\mu\text{m}$  (Fig. 3 in the main text). For the fast growing cells the initiation length per oriC is chosen such that it has the same initiation volume per oriC. *i.e.*  $I_f = I_s * (r_s/r_f)^2$ , where  $I_f$  and  $I_s$  are the initiation lengths per oriC in the fast and slow respectively and  $r_f$  and  $r_s$  are the cell radii in the fast and the slow respectively. From our measurements  $r_s = 0.92 \mu\text{m}$  and  $r_f = 1.15 \mu\text{m}$  (Fig S3). We set the standard deviation of the initiation length per oriC region to 0.2  $\mu\text{m}$ . Based on our estimates (see below) we set the C period to 60 minutes for both the slow and the fast growing cases. For the slowly growing cells the D period is set to 50 minutes (see below) and the growth rate to  $0.0089 \text{ min}^{-1}$  (Fig. S3). For the fast growing cells the D period is set to 25 minutes (see below) and the growth rate to  $0.023 \text{ min}^{-1}$  (Fig. S5).

**Algorithm 1**

Constants:  $I$ : Average initiation length per origin copy.  
 $s$ : Standard deviation of the initiation length per oriC region copy.  
 $C$ : C-period.  
 $D$ : D-period.  
 $\mu$ : Growth rate.  
 $t_{max}$ : Maximum simulation time.

Variables:  $n_{ori}$ : Copy number of oriC regions.  
 $L_i^{\circ}$ : Initiation length per  $n_{ori}$ .  
 $L_i$ : Initiation length.  
 $\tau_D$ : Division time.  
 $\tau_i$ : Initiation time.  
 $t$ : Time.  
 $L$ : Cell length.  
 $\tau_{D,S}$ : Sorted heap

```

while  $t < t_{max}$ 
  if  $\tau_i < \tau_D$  OR  $n_{ori} = I$ 
     $t \leftarrow \tau_i$ 
    push  $(t+C+D)$  to  $\tau_{D,S}$ 
     $\tau_D \leftarrow \text{top of } \tau_{D,S}$ 
     $L \leftarrow L_i$ 
     $n_{ori} \leftarrow n_{ori} * 2$ 
     $L_i^{\circ} \leftarrow \text{normmd}(I, s)$ 
     $L_i \leftarrow L_i^{\circ} * n_{ori}$ 
     $\tau_i \leftarrow t + (1/\mu) * \log(L_i/L)$ 
  else
     $L = L * \exp(\mu * (\tau_D - t))$ 
     $t = \tau_D$ 
     $L = L/2$ 
    store  $t$  and  $L$ 
    if  $\tau_{D,S}$  is empty
       $\tau_D \leftarrow \infty$ 
    else
      pop  $\tau_{D,S}$ 
       $\tau_D \leftarrow \text{top of } \tau_{D,S}$ 
    end if
     $n_{ori} \leftarrow n_{ori} / 2$ 
     $L_i \leftarrow L_i^{\circ} * n_{ori}$ 
     $\tau_i \leftarrow t + (1/\mu) * \log(L_i/L)$ 
  end if
end while

```

cells per frame without tracking (Fig. S5, time point zero), which was roughly 700 % of the cells physically present in a single colony at a given time. In addition to tracking cells from the previous frame, empty spaces within the segmented colony are re-segmented for each frame (Fig. 1 or the main text). When combining re-segmentation and cell tracking, the average number of segmented cells increases to 200 cells per frame and micro-colony after reaching steady state (Figure S5). The combination of segmentation and tracking hence increased the number of identified cells per frame and micro-colony to ~90 % of the expected number of cells physically present.

**Error analysis**

To estimate the accuracy of our segmentation and tracking, a total of 236 cells were randomly selected with replacement from cycle time intervals cycle time intervals 10-15, 15-20, ..., 55-60 minutes. The accuracy as determined by the fraction of completely correctly segmented cell cycles was 76% for all cycle time intervals. Cells with cycles shorter than 20 minutes had a lower accuracy of 52%. The accuracy for cells with cycle times in the range 20-35 minutes 93% and for cells in the range 35-60 minutes, the accuracy was 76%.

**Estimates of the C and D periods**

Based on our observations on average initiation volumes,  $V_I$ , the average volume at division,  $V_D$ , and the average volume at which replication terminates,  $V_T$ , we can calculate the time the C-period. By converting the volumes to times using the exponential growth relation, the C-period is given by  $C = (1/\mu)\log(V_D/V_I) + (1/\mu)\log(V_T/V_B)$  for slow growth and as  $C = (1/\mu)\log(V_D/V_I) + (1/\mu)\log(V_D/V_B) + (1/\mu)\log(V_T/V_B)$  for fast growth (Fig. S4). We find that the C-period is estimated to 63 minutes for slow growth and 56 minutes for fast growth. By comparing the average termination time to the average division time in Figure S4 we estimate the D period to 25 and 50 minutes in the fast and slow growing cases respectively.

**Detection efficiency in segmentation and tracking**

The throughput of the image analysis was tested on fast growing cells alone. The segmentation algorithm detected on average 140

### Efficiency in acquisition of cell cycles

The number of cell cycles acquired (Fig. S6) is a linear expansion after a lag time corresponding to about a generation time. We fitted the model  $y(t) = H(t - t_0) \cdot k \cdot (t - t_0)$  to the data. Here,  $H$  is the Heaviside function,  $k$  is the accumulation rate per number of micro-colonies monitored in parallel and  $t_0$  is the lag time. For the fast growing cells,  $k = 3.18 \pm 0.0008 \text{ min}^{-1}$  (95% confidence interval),  $t_0 = 42.8 \pm 0.11 \text{ min}$  (95% confidence interval). For the slow growing cells  $k = 1.418 \pm 0.0027 \text{ min}^{-1}$  (95% confidence interval),  $t_0 = 108.8 \pm 0.64 \text{ min}$  (95% confidence interval). The expected maximum number of acquired cell cycles is estimated as

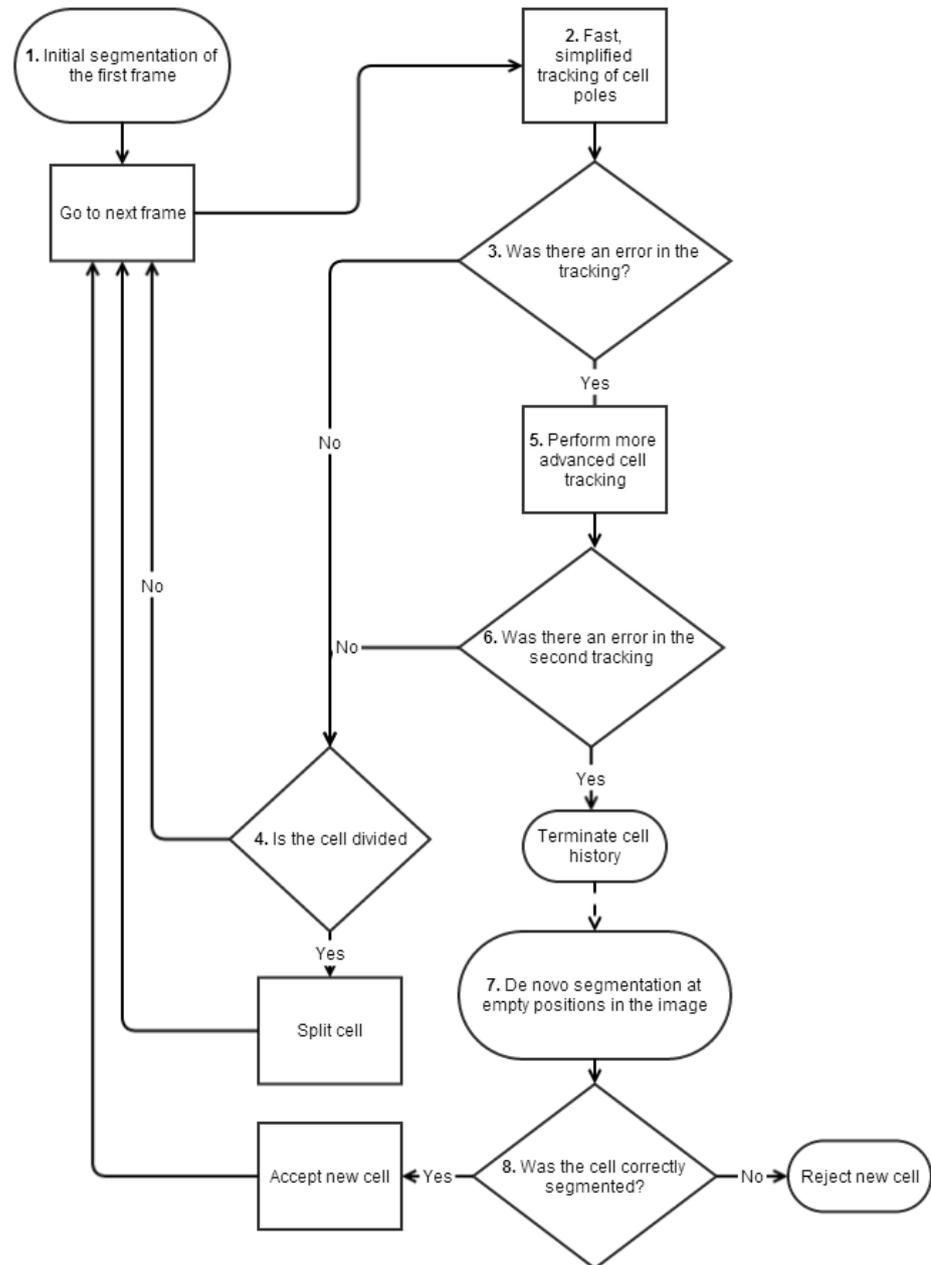
$y(t) = H(t - t_0) \cdot k_{max} \cdot n_{colony} \cdot (t - t_0)$  where  $k_{max}$  is the measured division rate per cell ( $0.0214 \text{ min}^{-1}$  for fast growth and  $0.0089 \text{ min}^{-1}$  for slow growth),  $n_{colony}$  is the estimated maximum number of cells per colony (220 for fast growth and 325 for slow growth, estimated by comparing the micro-colony area to the area of an average imaged cell). The maximum rate of cells completing their respective cell cycles is expected to be 5.12 and 2.35 cell cycles per minute for fast and slow growing cells respectively. This implies an acquisition of ~60% of the maximum number of cell cycles available.

### Characterization of cell growth

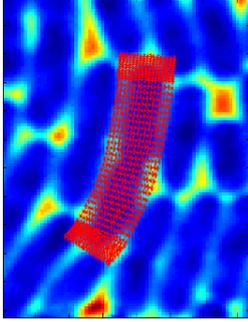
#### Growth and Morphology

As *E. coli* cells are rod shaped, the cell volume and mass is expected to expand proportionally to cell length as the variation between cells in width was small (Fig. S3) and therefore, the cell volume can be represented by cell length. During conditions of exponential growth the length evolves as  $L(t - t_B) = L_B \cdot \exp(\mu \cdot (t - t_B))$ , where  $L_B$  is the length at birth and  $t_B$  is the time of division. We found that cell length expansion during a cell life cycle under the conditions studied here was well described by an exponential function of time (Fig. S7). The exponential growth rate constant did not vary significantly between micro-colonies (Fig. S8) or over the course of the acquisition (Fig. S9), neither cycle times (Fig. S8). We find that siblings in general have more of the growth characteristics in common than mothers and daughters for both slow and fast growth (Fig. S10).

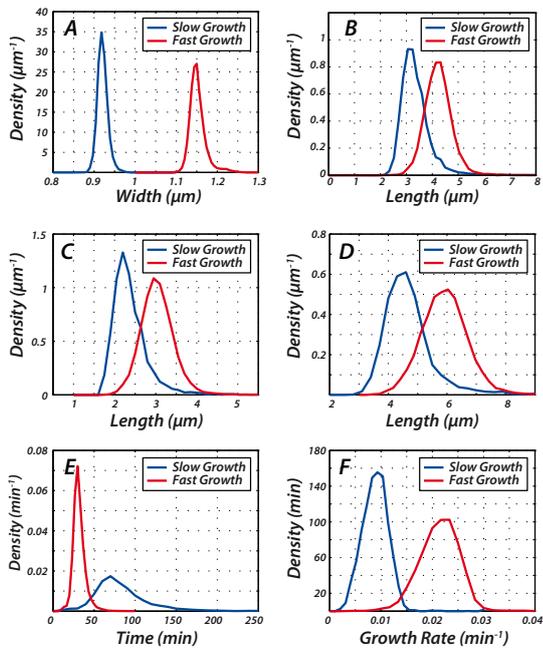
## Figures



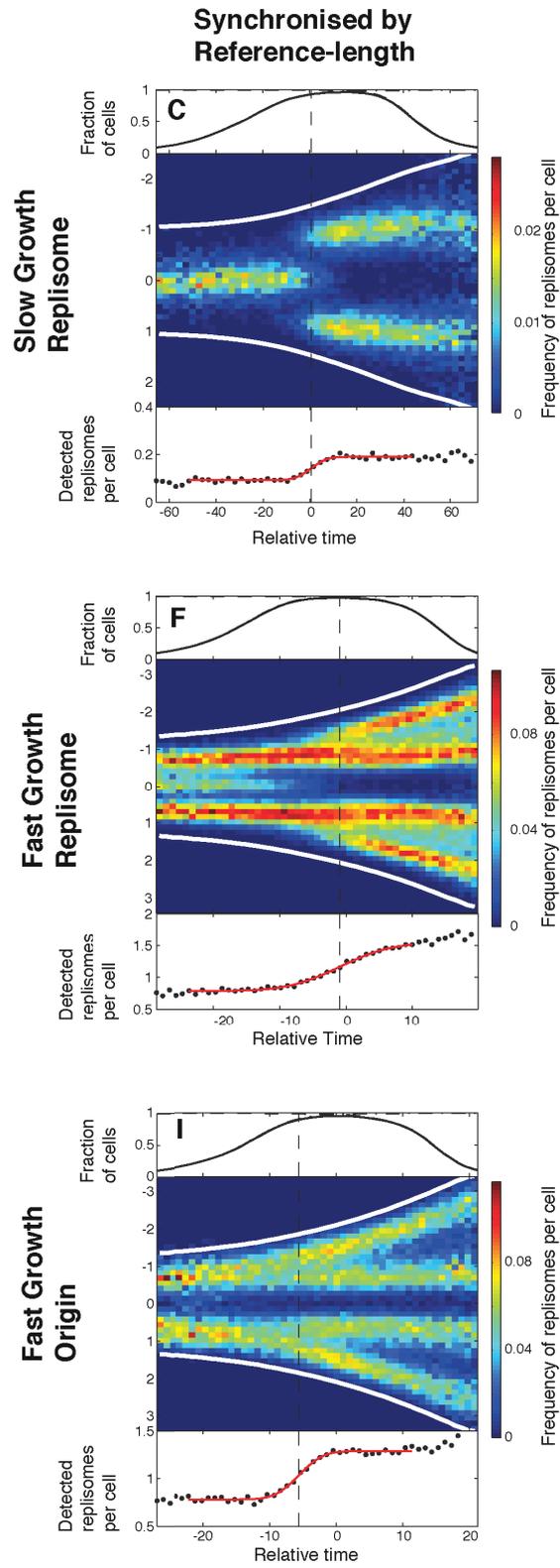
**Figure S1:** Flow chart of the cell segmentation and tracking algorithm.



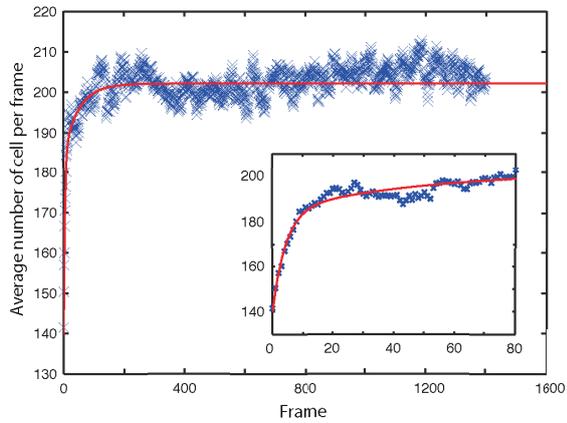
**Figure S2:** Grid of feature measures overlaid on phase contrast image.



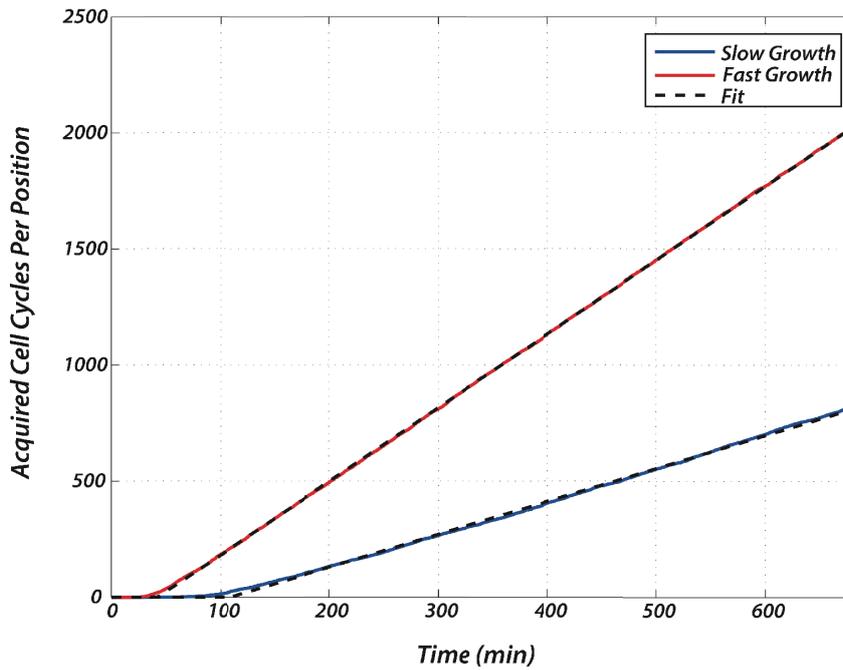
**Figure S3: Growth characteristics.** (A) Cell width average over cell cycle. (B) Cell length average. (C) Birth length. (D) Division length. (E) Cycle time. (F) Growth rate. Number of observed cases are 15190, 4091 for fast and slow growth respectively.



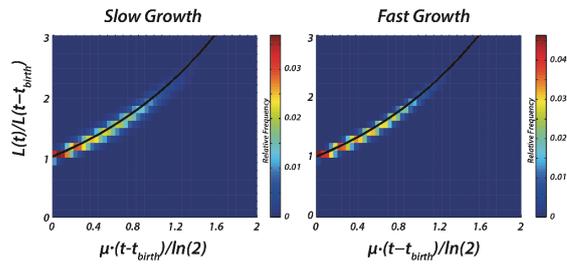
**Figure S4:** As Fig. 3 of the main text, but here cells are aligned based after time, where time point zero is chosen close to the initiation volume.



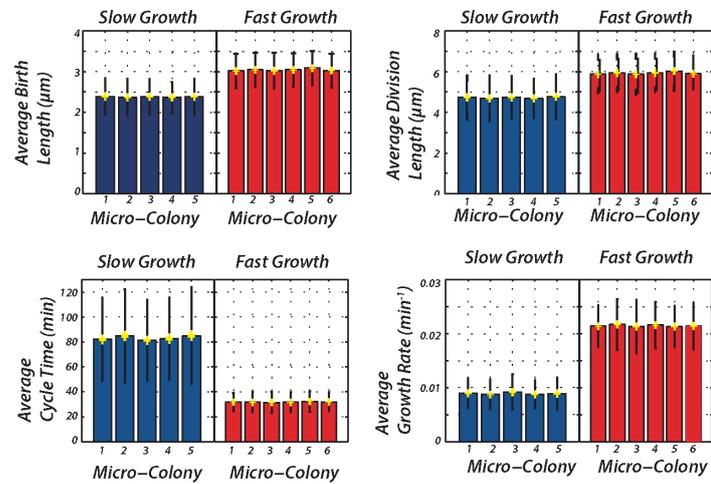
**Figure S5:** Number of detected cells per frame as a function of experiment duration. The number of detected cells is fitted to an exponential increase,  $y = y_0 + c_1(1 - e^{-k_1x}) + c_2(1 - e^{-k_2x})$  in which the coefficients (with 95% confidence bounds) converge to  $y_0 = 139.9$  (135, 144.7)  $c_1 = 17.75$  (14.34, 21.16),  $c_2 = 44.54$  (39.22, 49.87),  $k_1 = 0.02233$  (0.01689, 0.02777),  $k_2 = 0.2335$  (0.1732, 0.2938).



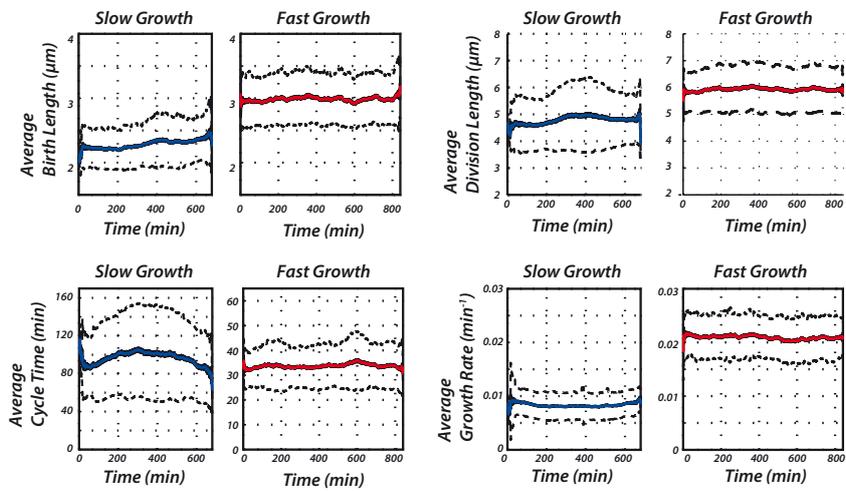
**Figure S6:** Number of acquired complete cell cycles as a function of the total time in the experiments. The data is fitted to a model  $y = H(t - t_0) \cdot k \cdot (t - t_0)$ , where  $k$  is the acquisition rate of new cycles and  $t_{lag}$  is a time-lag before the linear part of the curve and  $H$  is the Heaviside function. According to the best fit for slow growth  $k = 1.418 \text{ min}^{-1}$  (1.4153, 1.4207 95% confidence interval)  $t_0 = 108.8 \text{ min}$  (108.1644, 109.4356 95% confidence interval) and for fast growth  $k = 3.18 \text{ min}^{-1}$  (3.1792, 3.1808 95% confidence interval)  $t_0 = 42.8 \text{ min}$  (42.6868, 42.9132 95% confidence interval)



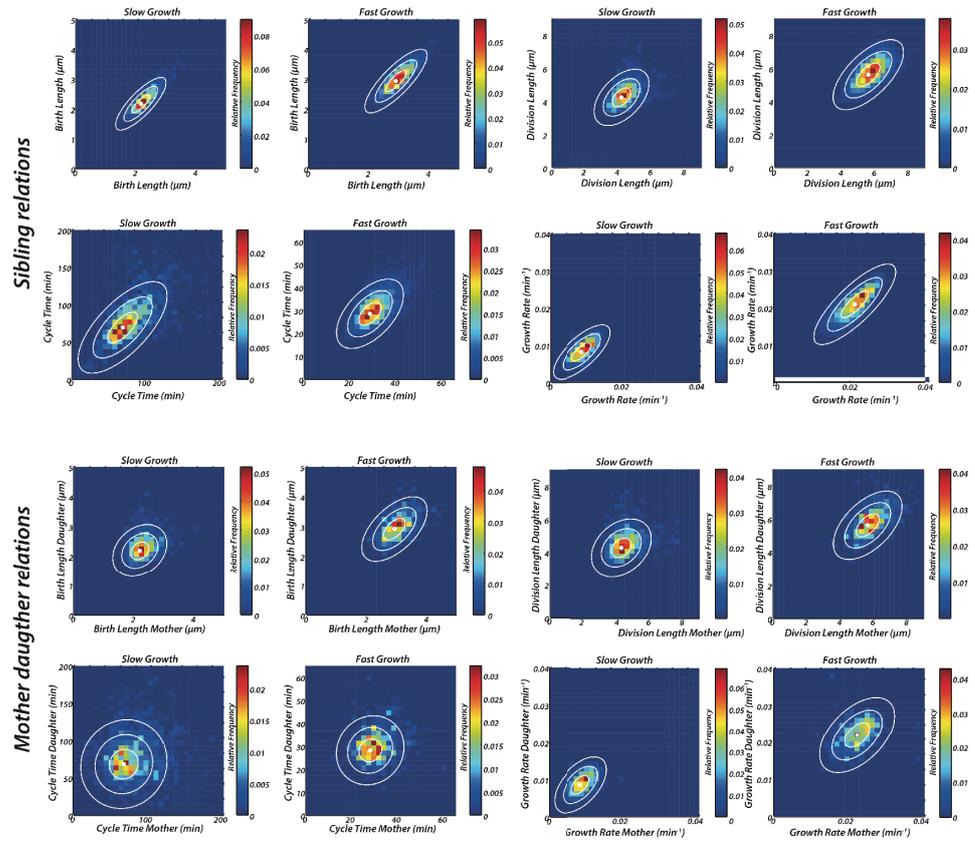
**Figure S7:** Cell length expansion from birth. Solid line is an ideal exponential function. Number of observed cases are 15190, 4091 for fast and slow growth respectively.



**Figure S8:** Growth characteristics in different traps of the same chip. Double sided standard deviation of the mean is shown as black bars and double sided standard error of the mean is shown as yellow bars. Number of observed cases are 15190, 4091 for fast and slow growth respectively.



**Figure S9:** Growth characteristics during an experiment. Each cell cycle contributes with the observed value for the time points within the cycle. The standard deviations from the mean are shown as dashed black lines and standard errors of the mean are shown as black solid lines. Number of observed cases are 15190, 4091 for fast and slow growth respectively.



**Figure S10:** Correlations between cell that have the same mother in the earlier generation (siblings) and correlations between mother and a daughter.



**US-AB**  
universitetsservice digitaltryck

Printed in Sweden 2014, [www.us-ab.com](http://www.us-ab.com)



341550  
Printed matter