UPPSALA
UNIVERSITET

X 14 013
Examensarbete 30 hp
April 2014

Induction kinetics of the lac operon

Studied by single molecule methods

Arvid Hedén Gynnå

UPPSALA UNIVERSITET

Molecular Biotechnology Programme

Uppsala University School of Engineering

UPTEC X 14 013	**Date of issue 2014-04**

Author

Arvid Hedén Gynnå

Title (English)

Induction kinetics of the *lac* operon studied by single molecule methods

Title (Swedish)

Abstract

The repression of the *E. coli lac* operon seems to be more efficient than the current theoretical model allows for. Specifically, it is more quiet than expected during the replication of the chromosome. I have induced cells during short periods and counted the number of protein products from the operon to determine if there is a delay in activation of transcription that could account for the discrepancy. The results are compatible with a delay of 10-20 s, but the delay could not be conclusively proven. Furthermore, it has been investigated if the mechanism behind the delay might be differential localization of the *lac* operon with and without induction. It is shown that the *lac* operon is more often located in the periphery of the cell and in the internucleoid region when induced. These might be regions where genes are higher expressed, giving a delay in expression after de-repression before the gene is transported there.

Keywords

Lac operon, repressor, transcription, single molecule fluorescence.

Supervisors

Prof. Johan Elf, Dr. Petter Hammar & Dr. Prune Leroy

Uppsala University

Scientific reviewer

Dr. Pia Lindberg

Uppsala University

Project name	Sponsors
Language **English**	Security
ISSN 1401-2138	Classification
Supplementary bibliographical information	Pages **33**

Biology Education Centre Biomedical Center Husargatan 3 Uppsala

Box 592 S-75124 Uppsala Tel +46 (0)18 4710000 Fax +46 (0)18 471 4687

INDUCTION KINETICS OF THE *LAC* OPERON

Studied by singe molecule methods

Arvid Hedén Gynnå

POPULÄRVETENSKAPLIG SAMMANFATTNING

Mikrobers genreglering har studerats under lång tid. Det kanske allra mest välstuderade systemet är *lac*-operonet, som uttrycker proteiner för nedbrytning av laktos. För att spara energi och resurser ska de proteinerna endast produceras när laktos finns tillgängligt. Den grundläggande mekanismen för regleringen är att ett protein kallat *lac*-repressorn binder till ett bindningsställe på DNA:t, och hindrar transkription när inget laktos finns. Om allolaktos tillförs binder det repressorn, får den att ändra form och därmed att släppa från DNA-strängen, vilket tillåter avläsning av genen. Trots att detta varit känt sedan 60-talet finns fortfarande många obesvarade frågor kvar.

Operonet är tystare utan induktion än det teoretiskt borde kunna vara, ty repressorn slängs av varje gång kromosomen kopieras innan celldelning och behöver sedan 20-30 sekunder för att hitta tillbaka. Under den tiden borde operonet vara öppet för avläsning. Jag har genom att inducera operonet under mycket korta perioder och sedan räkna antalet protein producerade försökt mäta om det möjligen finns en okänd fördröjningsmekanism som hindrar uttryck under korta perioder. Resultaten är kompatibla med att en fördröjning på 10-20 sekunder finns, men kan ännu inte helt bekräfta att så är fallet. Dessutom har en hypotes för den bakomliggande mekanismen undersökts: att genen när den aktiveras fysiskt flyttas i cellen från en plats där den är i vila till en plats med mer aktiv avläsning. Resultaten visar att aktiverade operon oftare befinner sig i cellens utkanter, vilket skulle kunna vara områden där det uttrycks högre. Vad som ligger bakom denna förflyttning är fortfarande kvar att klarlägga.

Resultaten är intressanta eftersom de antyder att nya effekter fortfarande kan upptäckas i ett mycket välutforskat system. Det är inte osannolikt att liknande effekter finns för många andra gener i *E. coli* och andra arter. Dessutom visar de återigen på att organisationen inuti bakterier är viktig, och att de inte är några enkla ”enzympåsar” som vi tänkte en gång in tiden.

Examensarbete 30 hp

Civilingenjörsprogrammet i Molekylär Bioteknik

Uppsala universitet, april 2014

TABLE OF CONTENTS

ABBREVIATIONS

AOTF	acoustic-optical tuneable filter
cAMP	cyclic adenosine monophosphate
CAP	catabolite activator protein
EMCCD	electron multiplier CCD (charge coupled device)
DPSS(L)	diode-pumped solid-state (laser)
IPTG	isopropyl-beta-D-thiogalactopyranoside
Lac	lactose, the lactose operon
lacA	gene coding for lactose transacetylase
lacI	gene coding for the lactose repressor
lacO	*lac* operator site
lacZ	gene coding for beta-galactosidase
lacY	gene coding for lactose permease
malI	gene coding for the maltose repressor
ONPG	o-nitrophenyl-beta-D-galactoside
PDMS	polydimethylsiloxane
RNAP	RNA polymerase

BACKGROUND

THE LAC *OPERON*

How microorganisms regulate their production of proteins is a field that has been studied for a very long time. In *E. coli* for example, the enzymes necessary for import and digestion of lactose are only produced in the presence of lactose. Such a regulatory system is of large importance for upholding the competitiveness of bacterial strains in an ever evolving environment, since they benefit from not wasting energy and nutrients on producing useless enzymes. On the other hand, the bacteria benefit from being able to quickly start using lactose if it suddenly becomes available. The regulatory system for lactose thus needs to be finely tuned to achieve the optimal balance between energy conservation and ability to use lactose. Due to the enormous evolutionary potential of intestinal bacteria achieved by combining a large population with short generation times, the system can also be expected to be very close to optimum for the common growth conditions of the species.

The *E. coli lac* operon is perhaps the most well studied prokaryotic system of gene regulation. The basic mechanism was deduced more than 50 years ago[1] by a combination of genetic and biochemical evidence. In the following decades, the operon has been studied in much more detail and is now well characterized. Both the classic mechanism and the long process behind its discovery have been described in book form by one of the pioneers in the field[2].

The organization of the *lac* operon is shown in Figure 1 with its respective sections to scale. It contains three genes: *lacZ*, *lacY* and *lacA*, coding for beta-galactosidase, lactose permease and lactose transacetylase, respectively. These are under transcriptional control by the P_{lac} promoter, which also contains the $lacO_1$ operator. The $lacO_1$ operator is a binding site for the *lac* repressor, coded for by the *lacI* gene. Binding of the repressor to the $lacO_1$ site inhibits the RNA polymerase from initiating transcription, which is the case in absence of lactose when enzymes should not be made. However, the binding of the repressor to the operator can itself be inhibited by binding of allolactose or an analogue thereof. This is the case when lactose is present and the enzymes of the operon should be produced. However, the regulation is more intricate than this. Two auxiliary binding sites, with lower affinity for the repressor, are positioned 401 base pairs downstream (named $lacO_2$) and 92 base pairs upstream ($lacO_3$) of the main operator site. The repressor forms a tetramer in vivo, which is able to bind two operator sites at the same time by creating creates a loop of DNA[3]. The ratio between protein expression in un-repressed and repressed state is more than 1000-fold in the wild type operon, but only around 20-fold when the auxiliary sites are removed. This occurs by faster association of the repressor to $lacO_1$, since a seeking repressor can be caught at several positions[4] and then transferred to the main operator site, which is very fast due to the temporary confinement of the repressor to the region around the $lacO_1$ site, effectively creating a higher local repressor concentration around it[5].

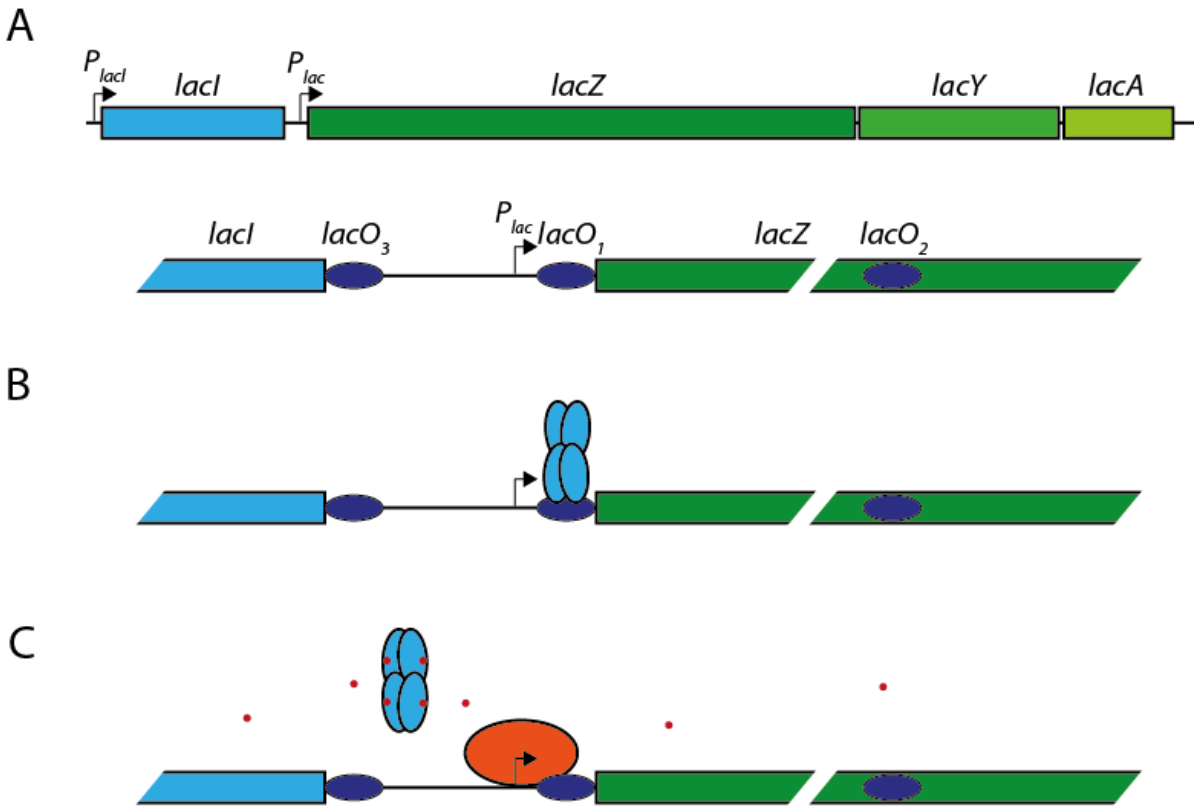

Figure 1: The organization of the lac operon A. The whole operon (above) with a close up of the area around P_{lac} (below) B. When lactose is not present, the lac repressor tetramer is bound to one (as shown) or several of the operators by DNA looping (not shown, see figure 1 in [5]). C. When lactose (red) is present, the repressor changes to a non-binding conformation, releases the operator(s) and allows access for RNA polymerase (orange).

In addition to this, the operon is positively regulated via the catabolite repression system: Since glucose is the preferred carbon source, operons for utilization of alternative sugars, such as lactose, are only highly transcribed when glucose is absent. This occurs by binding of a complex of catabolite activator protein (CAP) and cyclic AMP (cAMP) to binding sites present between the P_{lac} promoter and the *lacO₃* operator.

Although the lac operon is usually considered very well characterized, there are still fundamental issues that are unknown and certain discrepancies between predictions and measurements that need to be reconciled. For example, it is known that the circa 5 *lac* repressors of the cell[6] have an average association time of 20-30 seconds in vivo if they are dimers[7], while *E. coli* has a generation time of circa 2 000 seconds under laboratory conditions. Since on average one round of chromosome replication occurs each generation, one unrepressed copy of the lac operon is created every 2 000 seconds, even if the repressor is assumed to never disassociate without induction. Most likely, two unrepressed copies exist after replication, since it is hard to imagine a mechanism where a resident repressor would not be kicked off while the strands are separated and the DNA polymerase passes by. Combining the search time and the replication time tells us that the operon is by necessity unrepressed at least 20 seconds every 2 000 seconds, i. e. 1 % of the time. However, since an induced operon expresses about 1000-fold more protein than in the repressed case, either one of the measurements used in the above calculation must be wrong, or our equilibrium model for the regulation of the *lac* operon is still incomplete.

There is little reason to question any of the measurements we used, but one possibility is that the *lac* repressor tetramer associates considerably faster than the dimer. Since fluorescence labeling of the repressor with current methods disables its tetramerization domain, this is not experimentally testable given the current state of the art. It has in fact been speculated that a combination of effects can make its association as fast as six seconds[8] when at native concentration, but this is still too slow to achieve the measured repression.

Another option (although possible to combine with the alternative above) is that there is another mechanism that represses the operon, either during replication or generally when no repressor is bound during short periods. This would require a modification of the current model for regulation of the operon.

THIS STUDY

I have investigated if there is an unknown mechanism silencing the gene during short induction windows, by measuring the protein output after different induction periods down to 5 seconds. Overall, a linear response is expected up to a point where the dilution by cell growth becomes significant as shown in Figure 2. What is interesting is however what happens close to the origin. The operon is expected to have a shortest possible "on state" since the *lac* repressors need time to bind back after removal of the inducer. This pushes the response curve up making it intersect the vertical axis. However, if the response intersects the horizontal axis instead, this could be interpreted as that the start of expression after induction is delayed by a yet unknown mechanism. The delay has to be on the order of tens of seconds to account for the possible repressor search times.

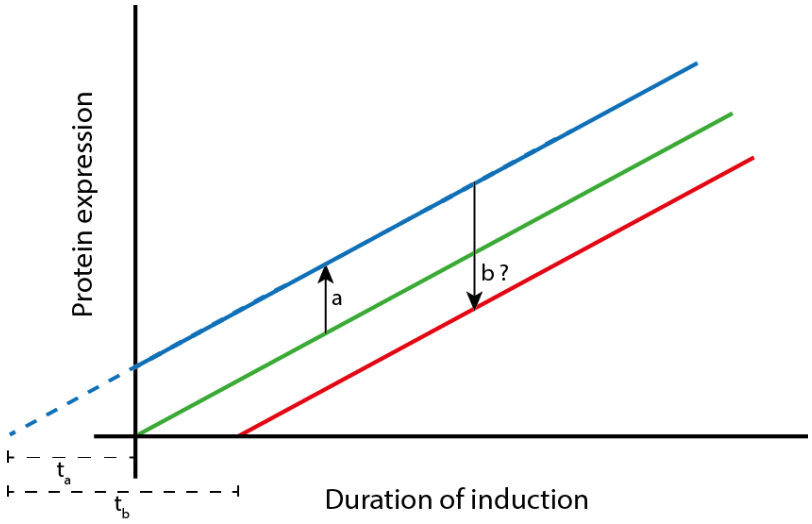

Figure 2: Response of the *lac* operon to an inducer. The linear response of the lac operon (green line) is expected to be shifted (a) by a time that corresponds to the average search time t_a needed for the lac repressors to bind the operators after the inducer has been removed. A mechanism to inhibit expression during short periods would shift the response back (b) and repress the operon during a time t_b, that is similar to or longer than t_a.

To resolve the behavior over a timescale of seconds, first, a system is required where the inducer can be quickly added and removed from the growth medium. Second, a method to detect and quantify minute expression levels is needed. The first requirement can be fulfilled using a microfluidic system where cells can be kept continuously growing for many generations, kept in place by physical restraining, while the growth media is continuously replaced and can be changed at any moment[9]. To measure the expression it might be tempting to use one of the most

common and sensitive reporters in molecular biology: the beta-galactosidase already encoded by the operon (see page 13 for description). However, there is no feasible way to measure beta-gal activity in living cells in the microfluidic system while continuously exchanging the medium since any produced reporter compound will be carried away with the stream. Instead a reporter in the form of a yellow fluorescent protein was inserted in the operon, to be excited by a strong laser and detected with a sensitive camera in otherwise naturally growing cells. This will be described in the transient induction chapter.

The mechanism of the delay

In the literature, there is actually a phenomenon documented that could be a mechanism of the delay theoreticized above. The *E. coli* chromosome, although not confined in a nucleus, is condensed into an elongated bundle that occupies the middle of the cell. In fast-growing cells this is more commonly seen as two lobes representing the segregating or already segregated daughter chromosomes. It has been observed[10] that the *lac* operon seems to change physical location in the bacteria when induced compared to when it is repressed. While repressed, the operon seems to be confined to a small region along the centre line of the cell in the middle of the nucleoid, compared to when the operon is induced and the operon is distributed over the width of the cell. The effect is claimed to be dependent on active translation taking place, but not to be related to the expression of the membrane protein of the operon, lactose permease. In another study[11], a similar displacement of the lac operon was observed which was concluded to depend on membrane insertion of permeases still in translation, which in turn pull the chromosome out to the cell periphery due to coupled transcription-translation. These views agree in that the operon location moves out of the nucleoid when active, but disagree on the cause thereof. Another aspect is what the consequences of a locus being present in the nucleoid periphery compared to the centre would be. It has been observed that RNA polymerase (RNAP) also seems to be located mainly at the edge of the nucleoid. This localization has been seen in a transmission electron microscopy study[12] where the bacteria were fixed and sliced before observation, and in 3D super resolution fluorescence microscopy[10]. A contradicting opinion has also been presented: that RNAP is evenly distributed through the nucleoid. This has been observed in a non-superresolution microscopy study[13] and with superresolution[14,15]. These three works however observes the whole cells in one plane, in essence observing a 2D projection of the cylindrical cell. An even distribution on a 2D image could also be interpreted as the presence of a RNAP envelope around the nucleus. The difference could also possibly be accounted for if the RNAPs imaged are in different states such as transcribing, searching or unspecifically DNA-bound and the localization differs between those states. As described below, subtle differences in microscopy techniques can make different partitions of the molecules to be visible in the image.

It is not difficult to imagine that a locus present near the nucleoid edge or out in the cytosol would be higher expressed than one buried deep in the chromosome if RNAP is more prevalent there, especially since ribosomes are known to reside between the nucleoid and the membrane. In fact, the prokaryotic paradigm of coupled transcription-translation combined with the fact that ribosomes are excluded from the nucleoid already suggests that the gene localization would matter. There are several mechanisms that could lead to a gene location dependence of the expression level. A gene located where the local RNAP concentration is higher would have a higher transcription initiation rate. The proposed effect could also act on a translational level: an mRNA produced where there are no ribosomes has a risk of being degraded before any

translation occurs while one made in or near the ribosome-rich regions of the cell can be protected by a train of ribosomes before any nuclease reaches it.

To investigate if this phenomenon exists under our experimental conditions and how it might affect the expression of the lac operon, the lac locus was labeled with fluorescence in living E. coli and the locations of thousands of loci were resolved as fluorescent spots under laser illumination with or without induction of the operon, as will be described in the location chapter starting on page 23.

GENERAL METHODS

The questions above are conceptually rather simple, but the methods needed to answer them have only recently become available. I have used a microfluidic system combined with advanced fluorescence microscopy and computational methods for automated image analysis.

CONSTRUCTION OF STRAINS

Strain for expression assay

For the expression rate assay, two strains were constructed based on *E. coli* BW25993[16] with the gene for the yellow fluorescent protein YPet[17] preceded by a Shine-Dalgarno sequence inserted between *lacZ* and *lacY* or replacing *lacY* (Figure 3). YPet is optimized to be one of the brightest yellow fluorescent proteins today and is also quite resistant to bleaching by light[18]. The location after *lacZ* was chosen to achieve an as strong fluorescence signal as possible due to polarity effects, which create a higher expression for early genes in the operon[19]. It cannot precede or replace lacZ since that would remove $lacO_2$ and change the regulatory characteristics of the operon. The permease coded for by *lacY* is an active transporter of lactose and the analogue IPTG[2021]. Since LacY synthesis will also be induced by IPTG, it could create a positive feedback loop that would introduce non-linear effects in the induction response of the operon. The effect is expected to be limited due to the induction time in most cases being shorter than the time needed for protein synthesis, and the high level of IPTG used induce the operon fully even without permease. However in order to avoid any effects that are difficult to incorporate in the model, strains both with and without *lacY* were constructed.

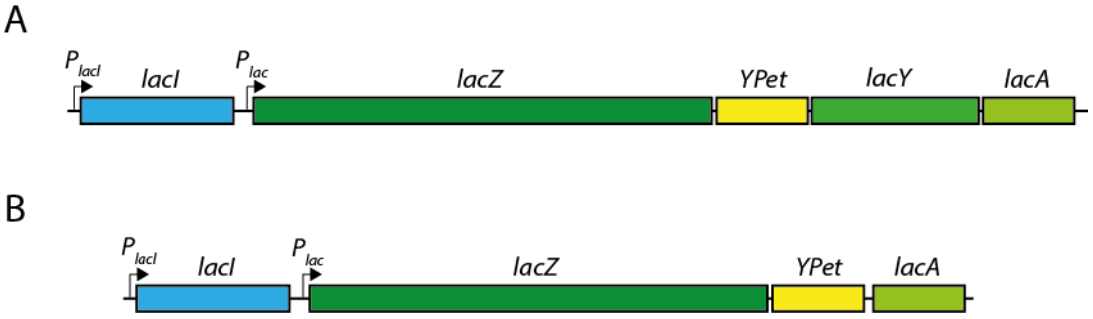

Figure 3: A. The *lac* operon in the lacZY::YPet strain. B. The *lac* operon in the *ΔlacY::YPet* strain.

After an initial attempt to insert the YPet gene by lambda-Red-mediated recombination[16] failed due to difficulties in producing a clean PCR product joining the YPet gene with a chloramphenicol cassette before integration in the chromosome, I turned to integration using the pKO3 suicide vector[22]. The pKO3 vector contains both positive and negative selection markers in the form of chloramphenicol resistance (*cat*) and sucrose lethality (*sacB*), as well as a temperature sensitive origin of replication (*pSC101(ts)*). If a chromosomal homology is inserted into pKO3, spontaneous integrants can be selected for using chloramphenicol in combination

with elevated temperature (42 ^{0}C). After isolation of clones containing the whole linearized plasmid in the chromosome, a counterselection for excisions of the vector backbone can be done using sucrose. This leaves a scarless integration of the originally homologous part including any changes or additions. In this case, the YPet gene was amplified from a chromosomal dnaQ-YPet fusion and inserted in a pGEMT vector with circa 570 base pairs of *lacZ* homology upstream and 460 base pairs of *lacY* homology downstream. This sequence was subcloned into pKO3 and inserted into BW25993 using the described protocol, leaving *YPet* inserted between *lacZ* and *lacY* (Figure 3A). This strain will be referred to as *lacZY::YPet*.

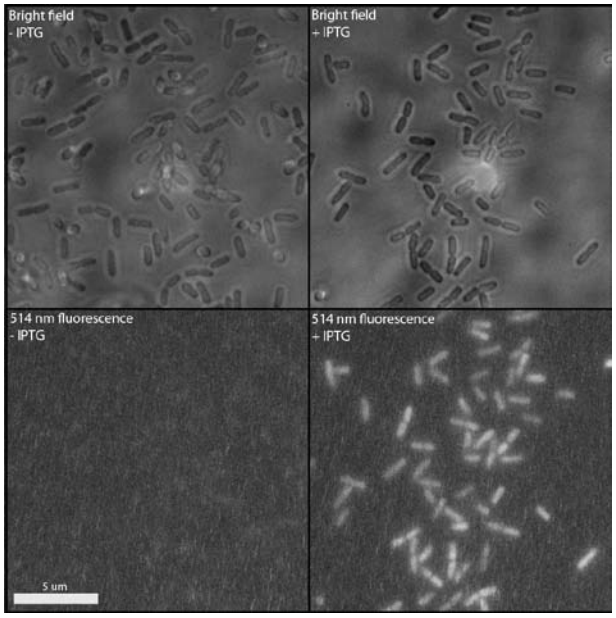

Figure 4: BW 25993 lacZY::YPet E172D incubated with and without 1 mM IPTG, fixed by polylysine and illuminated by 514 nm light. The brightness of the fluorescence images is scaled identically.

Lambda-Red-mediated recombination was used to remove the *lacY* gene from the *lacZY::YPet* strain. The recombination is performed by expressing the Red recombinase from a temperature sensitive helper plasmid, and electroporating in fragments with 40-50 nucleotides sequence homology and a resistance marker. The fragment will be integrated in the chromosome and can be selected for using a suitable antibiotic. To avoid any unwanted recombinations elsewhere in the genome by Red recombinase, the modified *lac* allele was transferred back to BW25993 by phage P1 transduction. The chloramphenicol resistance marker inserted by lambda-Red is surrounded by Flippase recombinase targets (*FRT*), and could easily be removed by expressing the Flp site-specific recombinase from a temperature sensitive plasmid followed by curing of the plasmid by incubation at 42 ^{0}C. This leaves a 42 base pair *FRT* scar between *YPet* and *lacA*, indicated by the gap between the genes in Figure 3B. This strain is designated *ΔlacY::YPet*.

Sequencing of the *YPet* gene and its close environment confirmed the expected sequence, except that glutamic acid 172 was changed to aspartic acid in the *YPet* sequence (E172D). Measurements in plate reader and fluorescence microscopy (Figure 4) revealed that cells incubated with 1 mM IPTG despite this mutation emitted strong yellow fluorescence spectrum when excited with 514 nm light.

Strain for labeling the *lac* locus

The lac locus was fluorescently labeled using a fusion of the maltose repressor MalI[23] and the yellow fluorescent protein Venus[24], which binds to maltose operator sites inserted in the *lacI* position. The strain (Figure 5) is also based on *E. coli* BW25993, and has the following

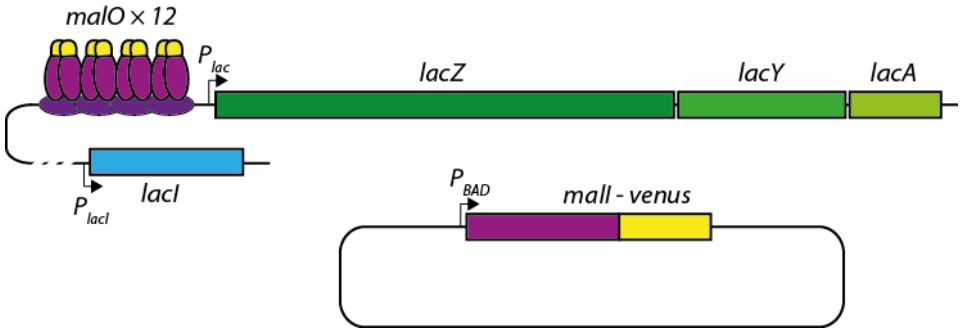

Figure 5: The *lac* locus labeling strain. LacI has been move to a mirroring position on the other arm of the chromosome, and twelve *malO* binding sites have been inserted in its location (top, only four shown). A MalI-Venus fusion is expressed from a plasmid (bottom) and binds the operators marking the location of the operon.

modifications: The P_{lacI} promoter and *lacI* gene have been deleted and reinserted in the *ygaY* gene (*ygaY::P_{lac}I-lacI-CmR*), 12 maltose operators have been inserted in place of *lacI* (*lacI::malO × 12*), and the maltose repressor gene on the chromosome has been deleted (*ΔmalI::FRT*). As a reporter, a pBAD24 plasmid expressing a C-terminal MalI-Venus fusion and carrying an ampicillin resistance marker has been inserted (*pBAD24-malIvenus*). MalI is not well studied biochemically, but is portraited here as a dimer due to the close functional and sequence similarity to LacI, which would be a dimer when fused in a similar way. Each *lac* loci can be labeled by up to 24 fluorophores. In reality, the number varies since some operators will be unoccupied and some will be occupied by immature and bleached molecules, creating spot to spot variations as seen in Figure 6. A number of fluorophores will also be unbound and freely diffusing contributing to increased fluorescence background.

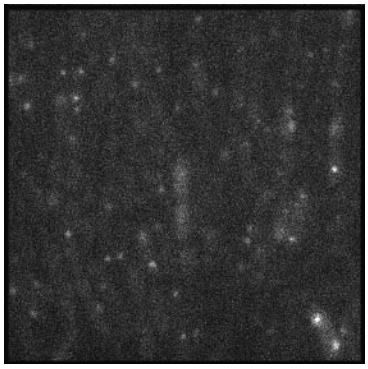

Figure 6: Fluorescent spots of different intensity labeling lac loci in a part of a microcolony on a microfluidic chip. Exposure time is 500 ms.

Regulation of the lac operon

A beta-galactosidase (Miller) assay[25] in was performed on all constructed strains to show that they all still had an efficient regulation of the lac operon after the genetic changes made (Table 1). The assay is done by harvesting cells in log phase, grown with and without IPTG, lysing the cells and adding the lactose analogue o-nitrophenyl-β-D-galactoside (ONPG). ONPG is degraded by beta-galactosidase to galactose and yellow o-nitrophenol that can be quantified using a spectrophotometer. If ONPG is available in excess, its degradation rate is essentially linear and dependent on the beta-gal concentration. By correcting each sample for reaction time and amount of cells added the activity can be calculate d in normalized Miller units. The assay was done in M9 media with 0.4 % glucose and supplemented RPMI amino acids. As a rule of thumb, non-induced wild type *E. coli* have about 1 and induced cells about 1000 Miller units, with variations depending on exact strain and growth conditions. The ratios achieved in the modified

strains indicate that their regulations were not significantly impaired by the genetic modifications.

Table 1: Expression of beta-galactosidase in the reporter strains and their common ancestor strain. The regulation ratio is the quota of activity after maximal induction (1mM IPTG) and after no induction.

Strain	Beta-gal activity (Miller units)		Regulation ratio
	- IPTG	+ IPTG	
BW25993	1.62	1125	695
lacZY::YPet	0.48	431	904
ΔlacY::YPet	0.52	479	928
lacI::malO×12 ygaY::PlacI-lacI ΔmalI	1.60	1310	821

MICROFLUIDICS

A microfluidic chip was used to study the cells while they grow exponentially under normal lab conditions. The design[7] is very similar to what has been described in detail earlier[9] but with two media ports allowing for two different types of media to reach the cells depending on the pressure in the respective ports. The setup is shown in Figure 7.

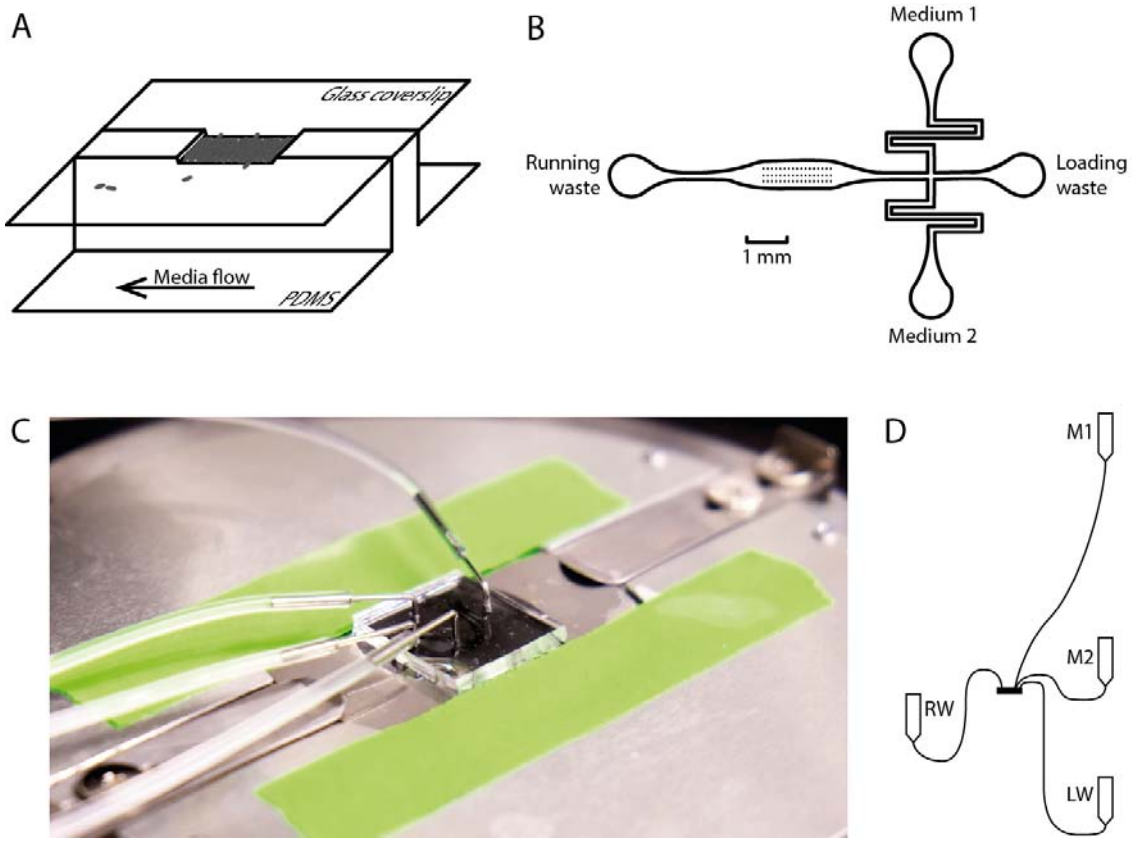

Figure 7: The microfluidic chip used in this study. A. A trap is surrounded on two sides by walls and open to the media flow on two sides. Cells grow and are continuously pushed over the edges and carried away by the flow. B. The flow channels in the chip. Media and waste reservoirs are connected by piercing the PDMS and inserting connectors. The traps are indicated by the three rows of squares in the wider channel. C. The chip as mounted for incubation and microscopy. The plastic tubing leads to reservoirs at different heights, providing continuous medium renewal. The metal clip in the top right is 8 mm wide. D. Typical configuration of reservoir heights. In this case, M1 will flow to both RW and LW, while M2 is fully diverted to LW and never reaches the cells. A & B are adapted from[9].

Briefly, the chip used is cast out of polydimethylsiloxane (PDMS), a soft clear silicon-like polymer, on a microfabricated mold containing microstructures that are imprinted on the PDMS cast. After solidifying, the cast is cleaned and attached to a glass cover slip creating channel structures between the glass and the PDMS. This chip has 51 traps, each 40 µm square and 900 nm deep, where bacteria can grow in a monolayer forming a microcolony of circa 250 cells. The traps are open on two lateral sides to deeper channels where the growth medium of choice flows by orthogonally to the openings. Diffusion of small molecules over width of the traps is efficient and cell growth rate does not seem to depend on position in the trap[9].

The chip is connected to two media reservoirs and two waste reservoirs via plastic tubing. The flow is gravity driven and the relative pressure between each reservoir and the chip is determined by adjusting reservoir height. Any combination of growth media can be used, provided a surfactant such as Plurionic F108 (normally at 0.85 g/l) is added to prevent cells from sticking to the PDMS or glass surfaces. It is also desirable to have media with low autofluorescence to lower the background during fluorescence microscopy, which often excludes the use of LB and similar media, and to add polystyrene beads of 0.5 or 2 µm diameters to one or both of the media for visualization of flow directions and velocities under the microscope.

MICROSCOPY

All experiments were performed on a Nikon Eclipse Ti-E inverted microscope with a Nikon Plan Apo TIRF 100×/1.49 oil immersion objective. The microscope was equipped with a motorized stage and had relevant parts enclosed in a cage incubator for temperature control. For imaging, the microscope was equipped with a highly sensitive Andor iXon EMCCD camera (resolution 512 × 512 pixels) for fluorescence imaging and either an auxiliary Scion or Luminera camera for phase contrast imaging. To avoid losing any fluorescence signal, only the light to the auxiliary camera was routed through a phase plate by using external phase contrast. Fluorescence excitation was provided by a green 514 nm laser, either a 150 mW Cobolt Fandango DPSS laser or a 2 W Coherent Genesis CX STM laser, and either an acoustic-optical tunable filter (AOTF) or a mechanical shutter, both depending on which experimental station that was used. An AOTF is a device to divert selected light wavelengths extremely fast using diffraction, without any mechanical parts. The laser beams were filtered through a pinhole, expanded and shaped for even sample illumination. All laser intensities given are back-calculated from measurements taken in the expanded beam and are approximate only. All optical devices, including the lasers and microscopes, were mounted on a stabilized optical table.

All cellular components are to some extent mobile and diffuse at different rates. This can be used to distinguish between bound and non-bound fluorophores. A freely diffusing small protein will cover the majority of the cell over a second long time span. When in complex with a larger molecule, such as a RNA or the chromosome, it will diffuse much slower. By choosing a long exposure time, unbound particles will be blurred into the background while more stationary molecules are still detected as single fluorescent spots. On the other hand, by taking very short exposure images each fluorophore is visible as a diffraction limited spot and can be individually counted as long as it is well separated from others[26].

In most cases three images were taken of each trap and at each time point: A phase contrast image to record the cell contours, a fluorescent image with laser illumination to find the fluorophores and a bright field image acquired with the EMCCD camera but under white light.

The purpose of the third image is to correlate the two other images as described below, and it is taken slightly out of focus to give more distinguishable cell contours.

Microscope operation and image acquisition were computer controlled using the open source Micro-Manager software[27]. For the shortest exposure time used, 1 ms, the timing of the computer control is not accurate enough. Instead the laser light shutter and the camera acquisition time are set to 5 ms. An external signal generator is triggered by the EMCCD camera at the start of acquisition of each frame and controls a second faster shutter in the laser light path exposing the sample for 1 ms only.

IMAGE ANALYSIS

The microscopy images were automatically analysed using MATLAB with a combination of pre-written routines and custom scripts. The data presents three problems for image analysis: segmentation of the phase contrast images to identify the cell outlines, detection and localization of spots on the fluorescence images, and correlation of the phase contrast and fluorescence images to each other. Of these, the cell segmentation is very computing power intense and can only be performed with reasonable efficiency on parallel high performance systems. The other steps can be done within reasonable time on an ordinary workstation or laptop. The process is illustrated schematically in Figure 8.

The phase contrast images were segmented using a combination of methods. The algorithm is based on the microbeTracker software[28], but modified and expanded as has been described[9]. An initial estimate of the cell outlines is made, which is refined using an active contour model. A grid is applied to each cell and is used to define the cell outline. Images from the two different phase contrast cameras need different algorithms for optimal segmentation. For the Scion camera 10 different parameter sets are applied separately and the result is combined to give one consensus segmentation. For the Luminera camera, one of the parameter sets has been found to work well enough that it can be used directly. This results in a computation time on a modern 12-core computer of around 8 minutes per frame for Scion images and 1 minute per frame for Luminera images.

As described above, fluorophores that are well localized over the timescale of the exposure are imaged as diffraction limited spots. To identify the spots automatically *á Trous* wavelet transform was applied[29]. In brief, this includes repeatedly applying a coarser and coarser smoothing filter to the image, in this case three times. In each iteration the resulting image is compared with the previous one and the difference is calculated, giving three different wavelet planes. The first plane has information about small structures which are mainly noise, while the latter planes include larger and larger structures such as diffraction-limited spots. The spot detection was done in the second wavelet plane, where areas over a threshold of 4 estimated standard deviations were recognized as spots. The actual location of the protein was determined to subpixel resolution by computing the weighted centroid of the original image pixel values of the spots. The implementation was the same as has been used in a previous study[7]. An alternative method of determining the location is by fitting a 2D Gaussian profile to the spot and use its peak for position. Gaussian fitting was evaluated since it gives a more accurate location in theory, but was discarded because it in this case gave unrealistic results in certain cells with unusually high background fluorescence.

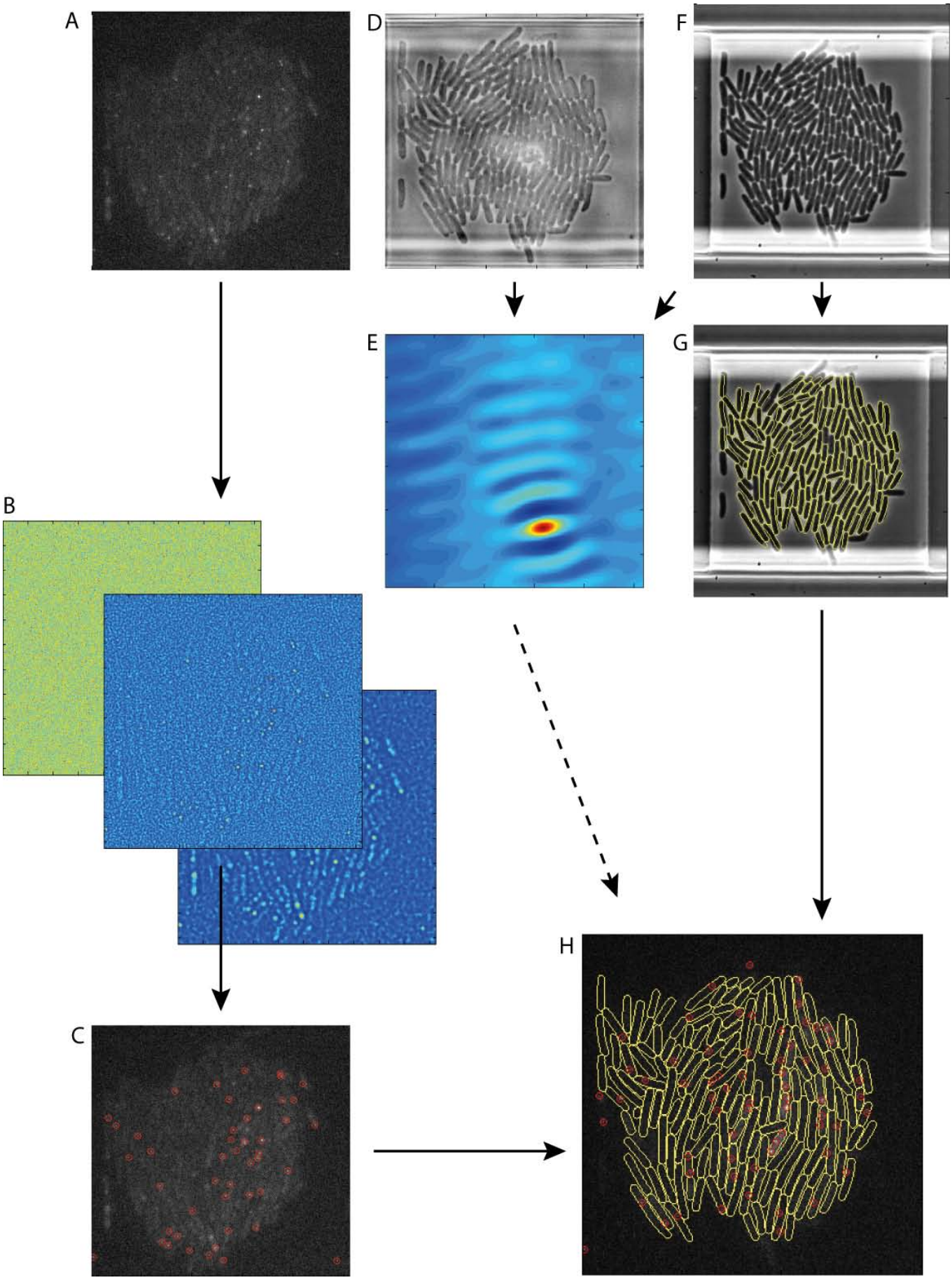

Figure 8: Overview of image analysis. A. Fluorescence image with spots from localized fluorophores and background from freely diffusing ones. B. 1st, 2nd and 3rd wavelet plane decompositions (left to right) C. Spots detected from the 2nd wavelet plane. D. Bright field image taken by the same camera as A. E. Correlation matrix between D and F. The red peak indicates the best image alignment. F. Phase contrast image. G. Cell outlines detected from F. H. Cell outlines superimposed on C based on alignment from E.

To align the computed cell outlines with the detected spots, the bright field images were rotated and scaled up to fit the size and orientation of the phase contrast images, before calculating the 2D correlation matrix for each image pair and using the peak correlation to determine appropriate translation. Then, the outlines calculated from the phase contrast images is overlaid with the spot locations from the fluorescence images and used to determine which cell a molecule belongs in and its location within the cell with superresolution accuracy.

TRANSIENT INDUCTION OF THE *LAC* OPERON

EXPERIMENTAL PROCEDURES

The strains were loaded in the 4-port microfluidic chip and supplied with M9 minimal media supplemented with 0.4 % glucose and RPMI amino acids with and without 1 mM IPTG. The medium used in microfluidics is normally degassed under vacuum to avoid formation of air bubbles in the microfluidic channels. However, oxygen is necessary for maturation of the YPet fluorophores and we want a strong and reproducible signal. Since only 21 % of air is oxygen, it was considered degassing the media and then aerating it with oxygen gas at atmospheric partial pressure. This would require specialized equipment not available so instead I decided to set up the media flow and load the chip with cells using degassed media, and then diluting the media in the reservoirs circa 1:10 with non-degassed media before the measurements were started. In most cases air bubbles do not form when a stable and rapid flow has already been established.

For inducing the operon, the media that was fed to the cells was temporary changed from the media without IPTG to the media with, and then back again after the desired induction time. For this purpose the media reservoirs were mounted on a vertical linear actuator with two holders moving in anti-parallel (i. e. one moves downwards when the other moves upwards), increasing the pressure in one reservoir and synchronously decreasing the other's pressure. The actuator was computer controlled for timing accuracy, reproducibility and automatic synchronisation with image acquisition. The movement of the actuator between the two states takes less than two seconds. To be sure that the induction times are accurate, the shortest time used was set to five seconds. The chip was incubated for six hours, corresponding to approximately twelve cell generations, without IPTG between each induction to renew the cells and dilute the YPet.

When using proteins as a reporter of a transcriptional control system, there is of course a delay between induction and when we can measure the protein concentration due to the time needed for transcription and translation. But fluorescent proteins are also not born fluorescent. They need to undergo spontaneous chemical reactions in their fluorophores, termed maturation, before they are ready to emit any fluorescence. The maturation rate of YPet under our in vivo conditions is difficult to predict from available in vitro measurements so we do not know how long we have to wait until the majority of proteins have matured. On the other hand, the concentration of the protein is constantly diluted by the division of the cells. Since we do not know when the strongest fluorescent signal can be observed, the cells were imaged at several time points up to one hour after the induction. Fluorescence imaging is destructive since the fluorophores are bleached by light exposure. Due to bleaching, different traps of the chip were imaged at the different time points. This will however lead to an increased noise over the time course as background fluorescence varies between the positions of the chip.

Imaging conditions

The clear cell shapes visible in the lower right of Figure 4 suggests that the cells are strongly fluorescent, but note that these have been induced for many cell generations and are imaged fixed to a glass surface by polylysine. In the actual experiment we need to measure much lower YPet concentrations, down to a few mature proteins per cell. Imaging continuously growing cells in the microfluidic chip also adds significant background fluorescence emitted from the PDMS.

There are two alternative approaches to measuring expression of fluorescent protein with a microscope: either use a long exposure time to collect as much fluorescence as possible while the proteins diffuse over the cell, or use a very short time combined with a high laser intensity to image the proteins as diffraction limited spots and count them individually.

When using the first method laser intensity and exposure time needs to be optimized to increase signal to noise ratio. There are different kinds of noise that behave differently: there is background fluorescence from the PDMS, there is cellular autofluorescence and camera sensor noise. Sensor noise is negotiable in this experiment, and the chip was poured using half the normal volume of PDMS, giving less than half the normal thickness to reduce PDMS autofluorescence. However, the noise level was still high.

To investigate the optimal imaging parameters, 50 fluorescent images with 0.1 s exposure time and three different laser intensities were taken quickly of traps filled with the lacZY::YPet strain previously induced for 30 minutes, corresponding to total illumination time of 5 s. The fluorescence decreases quickly down to a level where it stabilizes as seen in Figure 9A. Assuming that the background fluorescence bleaches slowly and that all YPet has bleached within the first 4.5 seconds of exposure, the background can be removed by subtracting the average of the last five measurements of each trap and the decay rate calculated by fitting an exponential function. Surprisingly, the bleaching rate seems to be constant regardless of laser power with a fluorophore half life of circa 7 s. Calculating the ratio between YPet and background fluorescence is helpful for deciding imaging parameters (Figure 9B) and shows that in this case the initial signal-to-background is 1:2 and decreases down to circa 1:2.5 after the first second of imaging. The ratio will be much lower in the experimental cases when induction is shorter but the decay rate will be the same. For this reason a 500 mW laser power was selected together with the reasonably short exposure time of 500 ms. The EMCCD camera was cooled to -80 ^{0}C and gain was set to 150. The experiment was performed with the lacZY::YPet strain and induction times up to four minutes.

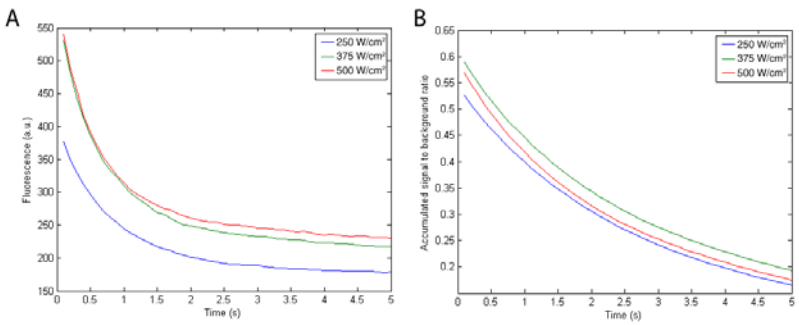

Figure 9: Decay of YPet fluorescence under laser illumination. A. Total area fluorescence for every 100 ms interval up to 5 s. B. The cumulative ratio between YPet and background fluorescence

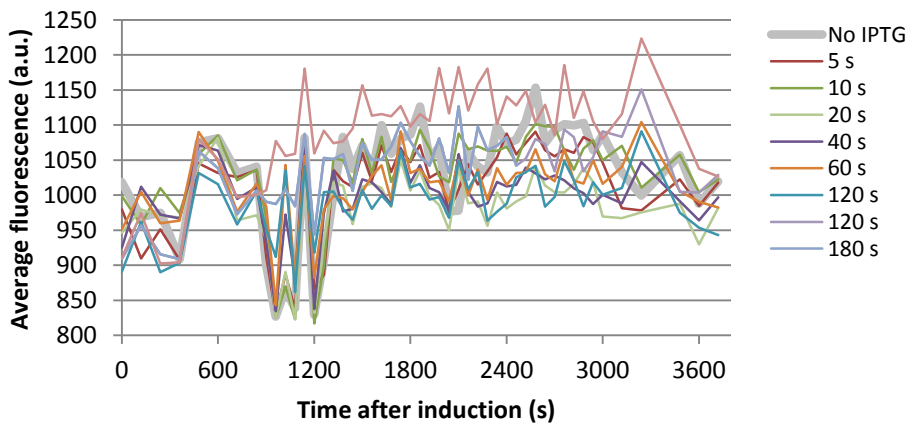

Figure 10: Time courses of fluorescence after different induction times, measured as the average pixel value of the area covered with cells.. All curves are averages of three image sequences at different traps.

The phase contrast images were segmented as described above, but since no spots were present due to the long exposure time the average fluorescence of the area covered with cells was measured using the camera pixel values instead, shown in Figure 10. Also, the total fluorescence of a selected cell-filled square, per-cell fluorescence and per-cell fluorescence normalized for cell size were calculated with a result similar to that of the cell covered area (data not shown). The background value was circa 1 000 units, while the maximum increase was measured to be circa 200 units. Note that this was after an induction time of 240 seconds. If the expression depends linearly on induction time, the expected maximum signal from a 5 second induction would be less than 0.5 percent of background value. A such small signal would be very difficult to measure accurately, so this measurement was complemented by actual counting of molecules for the shortest induction times.

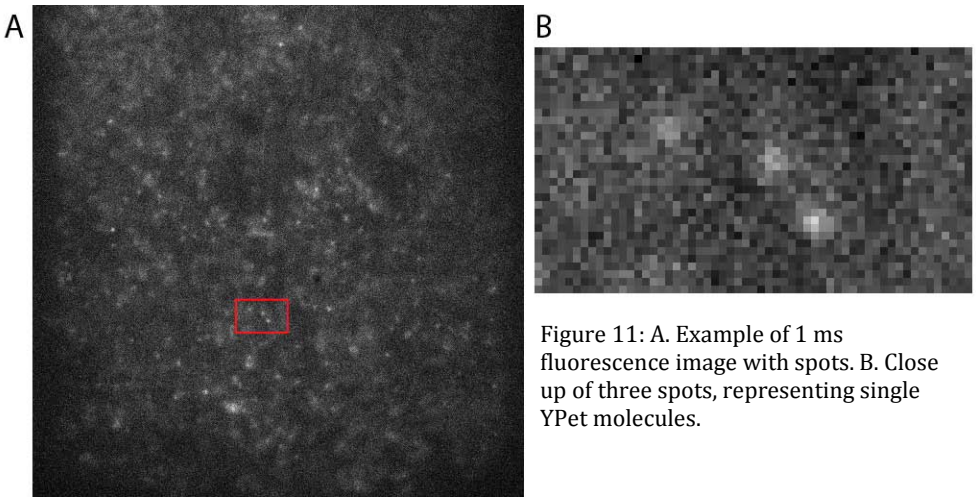

Figure 11: A. Example of 1 ms fluorescence image with spots. B. Close up of three spots, representing single YPet molecules.

To resolve single fluorescent molecules, a 1 ms exposure time was combined with the strongest laser power achievable. The 2 W Coherent laser was used at maximum power and the pinhole for light filtering was removed, increasing intensity more than 10-fold. The actual power of the beam could not be determined due to limitations in the measuring equipment. To detect the minute fluorescence the camera gain was set to 500. The camera was cooled by water cooling to -95 ^{0}C to decrease the noise caused by a high gain. As seen in Figure 11, images were very noisy but it is still possible to identify single molecules as diffraction limited spots, both by eye and by automatic detection. The experiment was done in the ΔlacY::YPet strain with induction times up to 60 seconds and four hour delays between each induction. All traps were imaged between 40 and 45 minutes after induction start where the signal, based on Figure 10, was expected to be strongest to maximise the amount of useful data collected.

Running microfluidic experiments for several days is associated with technical problems that limit the practical throughput. Common problems are contamination of media reservoirs, formation of air bubbles that stop the flow, pressure fluctuations causing the bacteria to fall out of the traps, growth of bacteria adhering to the flow chamber walls that blocks media flow, and software crashes. In many cases an experiment can be saved and restarted by proper intervention, but in other cases chips have to be given up (e. g. when there is major cell growth outside of the traps). The media pressures for switching have to be calibrated individually for each chip and since media flow cannot be monitored during a program, it should ideally be checked by manual microscopic inspection both before and after an experiment. Due to these constraints, each of these experiments has only been performed successfully once.

RESULTS

The time courses recorded in the first experiment were corrected for background by removing the value at time zero from the remainder of the series (Figure 12A). There is evidently also a clear trend that some traps are darker or brighter than the others, but in most cases these phenomena are not present after all inductions. The cause of this is presently not clear, but further correction of the individual traps based on their fluorescence in the control run did not improve the smoothness of the curves. A better method to cancel out trap-to-trap variations is to average the results of several traps. The strongest signal is observed around 40 minutes after induction start. When averaging the fluorescence at time points from 35 to 50 minutes (n = 16) as shown in Figure 12B, a strong linear trend is seen at least for inductions of 60 s or longer. Below that, there is a too large variation to draw any conclusions.

The data for the number of single molecules gives a more detailed picture of what happens at short induction times. The 5 and 60 s inductions were excluded from the analysis due to microfluidic problems, but the remaining data is shown in Figure 12C. Little or no increase in the number of molecules is seen after a ten second induction, while it increases four-fold by induction during 20 s. Surprisingly however, it again decreases to an intermediate level of four times the background after a 40 s induction. It is expected that the number of detected spots will stop to increase or even decrease when the number of molecules reaches a certain lever. This is since the spots will blend into each other and increase background fluorescence making detecting of single molecules more difficult. To see if the dip at 40 s is caused by a too large number of fluorophores, I calculated the area fluorescence of the central 312 × 312 pixels of each fluorescent image. Even if molecules are colocalised, they should still contribute to overall fluorescence linearly to their number. However, the area fluorescence follows the same pattern as the number of spots (Figure 12D), taking us to the conclusion that Figure 12C represents the true number of fluorophores in the images.

What is notable is that the spread of the individual traps is much larger in the 40 s case than in the others. As it turns out, there seems that the one of the three rows of traps in the chip is more induced than the other two. This could mean that the laminar flow of the chip was somehow obstructed during the 40 s run and that not all traps were equally exposed to IPTG, putting some doubt in the result of the 40 s induction.

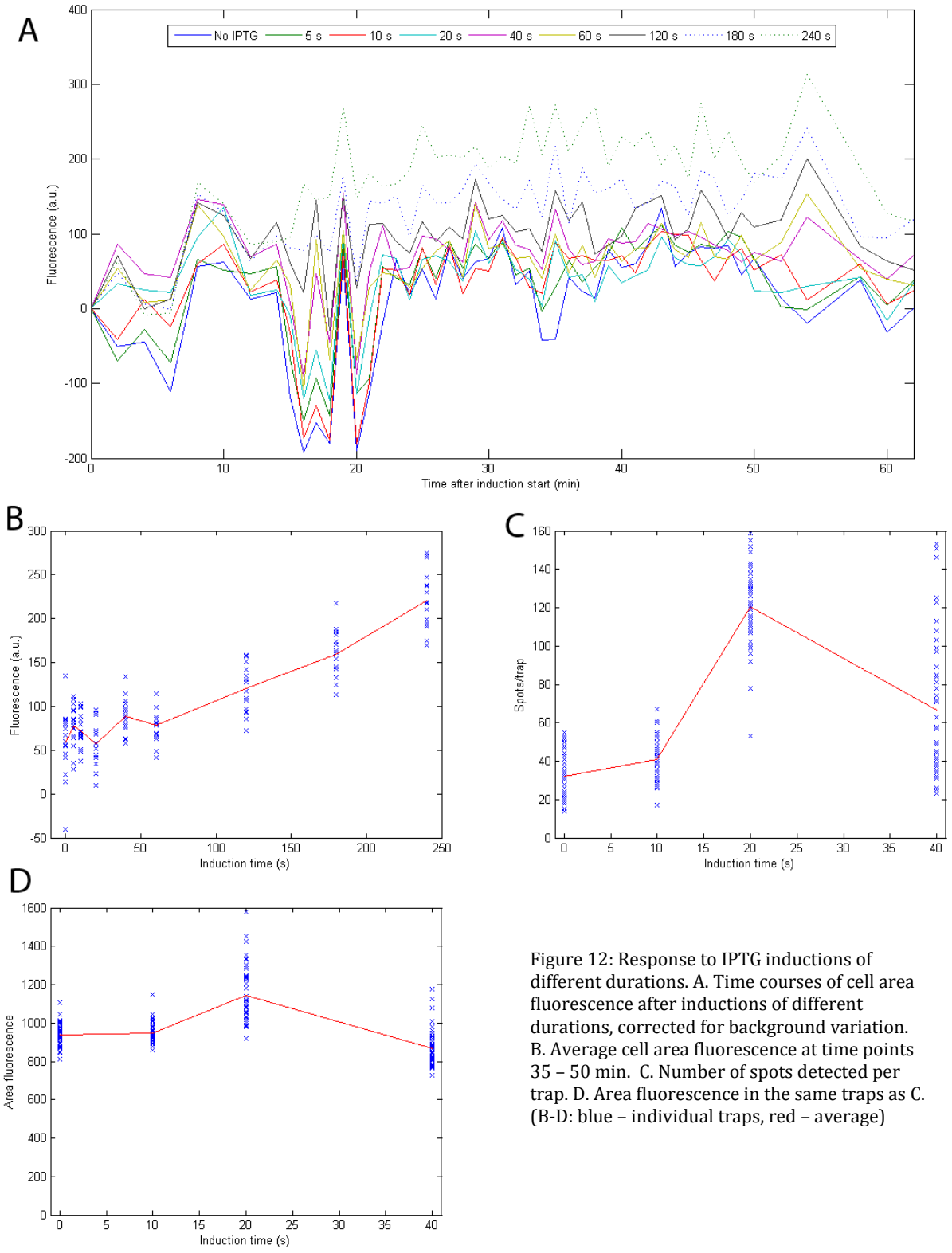

Figure 12: Response to IPTG inductions of different durations. A. Time courses of cell area fluorescence after inductions of different durations, corrected for background variation. B. Average cell area fluorescence at time points 35 – 50 min. C. Number of spots detected per trap. D. Area fluorescence in the same traps as C. (B-D: blue – individual traps, red – average)

LOCATION OF THE *LAC* OPERON

EXPERIMENTAL PROCEDURES

Even though the location of the operon has been labeled as described on page 13, we cannot just watch single loci and see how they move. The reason for this is location uncertainty of single molecules, and more importantly, that we can only observe a 2D projection of the actual spot position in the rod-shaped bacterium. A locus that moves in the z direction (i. e. normal to the imaging plane) will not be detected by these methods. Since we expect the movement to be equally probable in all directions orthogonal to the cell long axis, only a fraction of the actual position changes will be seen directly. Instead we need to record a large number of loci positions and see if their distribution over the cell changes depending on induction. Initially, this has been performed with the same growth conditions as the transient induction experiment (M9 with 0.4 % glucose at 37 ^{0}C) and with a medium more similar to an earlier study showing this effect [11] (M9 with 0.4 % glycerol). The strain described in Figure 5 was used. IPTG (1 mM) was added to medium 2 for induction, and carbenicillin (100 µg/ml) to both media for maintenance of the pBAD24 plasmid. As no arabinose was present in the medium, the reporter protein is only expressed at the background level of the P_{BAD} promoter. Four port chips were used with the reservoir heights manually controlled. The cells were fluorescently imaged using a laser power of circa 125 W/cm^2 and an exposure time of 500 – 1000 ms, with camera gain at 150 and the sensor cooled to -80 ^{0}C. All traps were imaged every three or six minutes for two to four hours at each growth condition, depending on the number of traps included in each experiment.

The images were analyzed as described in the methods section, resulting in at least 17 000 spots per growth condition that could be mapped to a segmented cell. Spots that were not located inside cell outlines were excluded from further analysis. Manual inspection revealed that most of those were actually inside cells that had not been detected by the segmentation algorithm, rather than false spots that had been found in an empty area of the trap. The spot locations inside the cells were recorded and transformed to standardized internal cell coordinates, where distance between the cell midpoint and each pole equals one. By this method, a large number of cells with varying size and shape can be overlaid and the distribution of locations can be accurately studied even without any 3D information. While it is technically possible to overlay any cells that have been converted for unit size, for biological reasons only cells from specific length brackets have been compared here since the internal organization of *E. coli* changes with cell length.

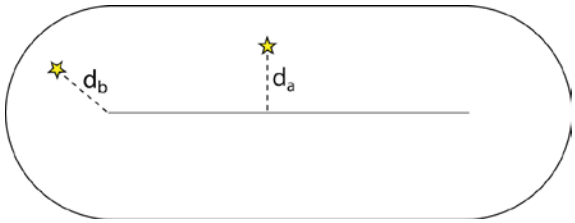

Figure 13: The off-centre coordinate. If a spot is in the cylindrical part of the cell (outlined), its distance to the centre d_a is the vertical coordinate. If it is in the end cap, Pythagoras theorem is used.

The distribution of locations can be quantified by measuring how far the spots are from the cell centre. An off-centre coordinate can be defined as in Figure 13; in the case where the spot is in the cylindrical part of the cell the value along the coordinate is the vertical distance to the long axis, while if it is in the end caps (defined as the 0.5 µm at each end of the cell) the Cartesian distance to the endpoint of the centre line is used. A histogram of the locations along the coordinates is given for each growth condition and reveals how far from the nucleoid centre the locus is.

If the 2D spot densities are plotted, they describe an elongated shape in short cells, which split into two lobes for longer cells during all growth conditions, and in the longest fast growing cells even into four distinct high-density areas. This corresponds well with the general organization of the *E. coli* chromosome in the cell and provides a basic validation of both the labelling of a specific locus and of the image analysis algorithms. There is a significant heterogeneity between reporter expression levels of the cells, which affects the fluorescent background of each cell and masking the spots in certain cases. This heterogeneity is expected when a protein is expressed from a low-copy number plasmid without any regulatory feedback and does not affect the result of the experiment, as the location of the *malO* sites is not dependent on reporter expression.

RESULTS

For cells growing under the same conditions as in the experiment above, the pattern of the loci distribution changes when the operon is induced by IPTG. As seen in the right column panels of Figure 14, the loci in cells with lengths from 2 to 3 μm are located further out when induced compared to when not. Most notably, when the 3D shape of the cell is taken into account, the loci are more or less excluded from the nucleoid centre. The phenomenon exists but is much less pronounced in longer cells. The *lac* operon is expected to be more induced when growing on glycerol as a carbon source, due to accumulation of cAMP (see page 8). However, this effect is actually much weaker here as evident from Figure 15. There is no strong depletion from the nucleoid centre when induced, only a slight decrease combined with some loci being present further out. Despite of this, there seems like the loci avoid the centre to some extent regardless of induction. This is interesting, since this condition is similar to the one used to investigate the phenomenon earlier[11], where effect was seen clearly.

The chromosome in a rapidly growing cell is very active, with both transcription and replication continuously perturbing it. I theorized that a slower growing cell with less activity on the chromosome would show the effect of a single gene being active more clearly. The same experiment was thus performed in M9 with glucose and amino acids but growing at room temperature (23-24 ^{0}C) and in M9 with only glucose and no amino acids at 37 ^{0}C (also known as minimal medium). The loci were distributed further out from the cell centre when induced in both room temperature (Figure 16) and without amino acids (Figure 17), but the effect is actually weaker than in fast-growing cells. A notable difference between the cases is that the loci to some extent avoids the centre in room temperature cells similar to when in glycerol, while it is much more evenly distributed in minimal media at 37 ^{0}C.

A noticeable difference that is only visible in the 2D heat maps of loci locations, is that the there are more loci present in the inter-nucleoid region between the segregating daughter chromosomes in short cells. This is especially visible when cells are grown in minimal media only and in shorter cells.

Only including the spots in the cylindrical part of the cells gives very similar results to the method used here.

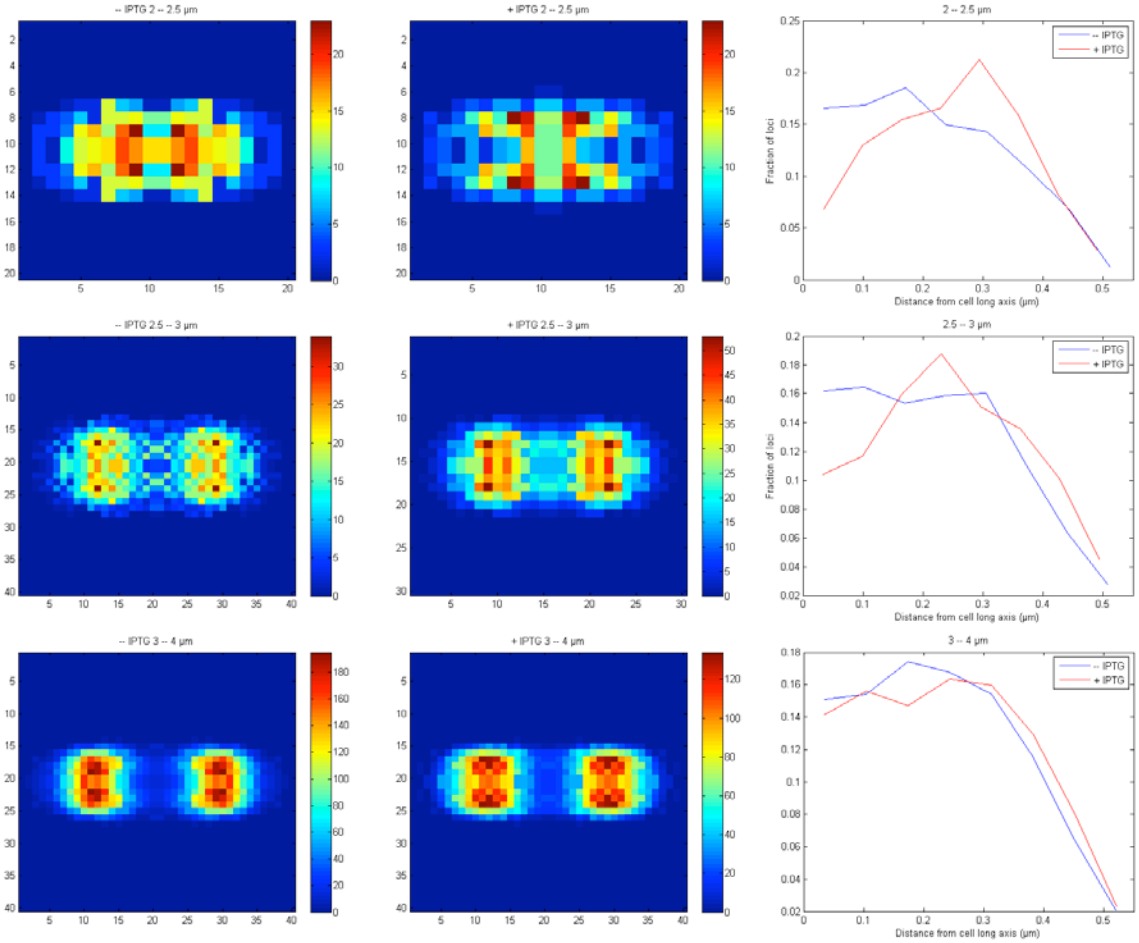

Figure 14: Heat maps of operon location densities in M9 with amino acids and glucose at 37 °C, sorted in different cell length brackets without (left panel) and with IPTG (middle panels). Cells have been scaled to unit length, mirrored horizontally and vertically, and overlaid. The resolution of the heat maps was adjusted for each length bracket to include a reasonable amount of data in each square. Histograms of loci distributions along the coordinate described in fig 13, for each length bracket (right panels). The number of loci locations used for each figure is given in Table 2.

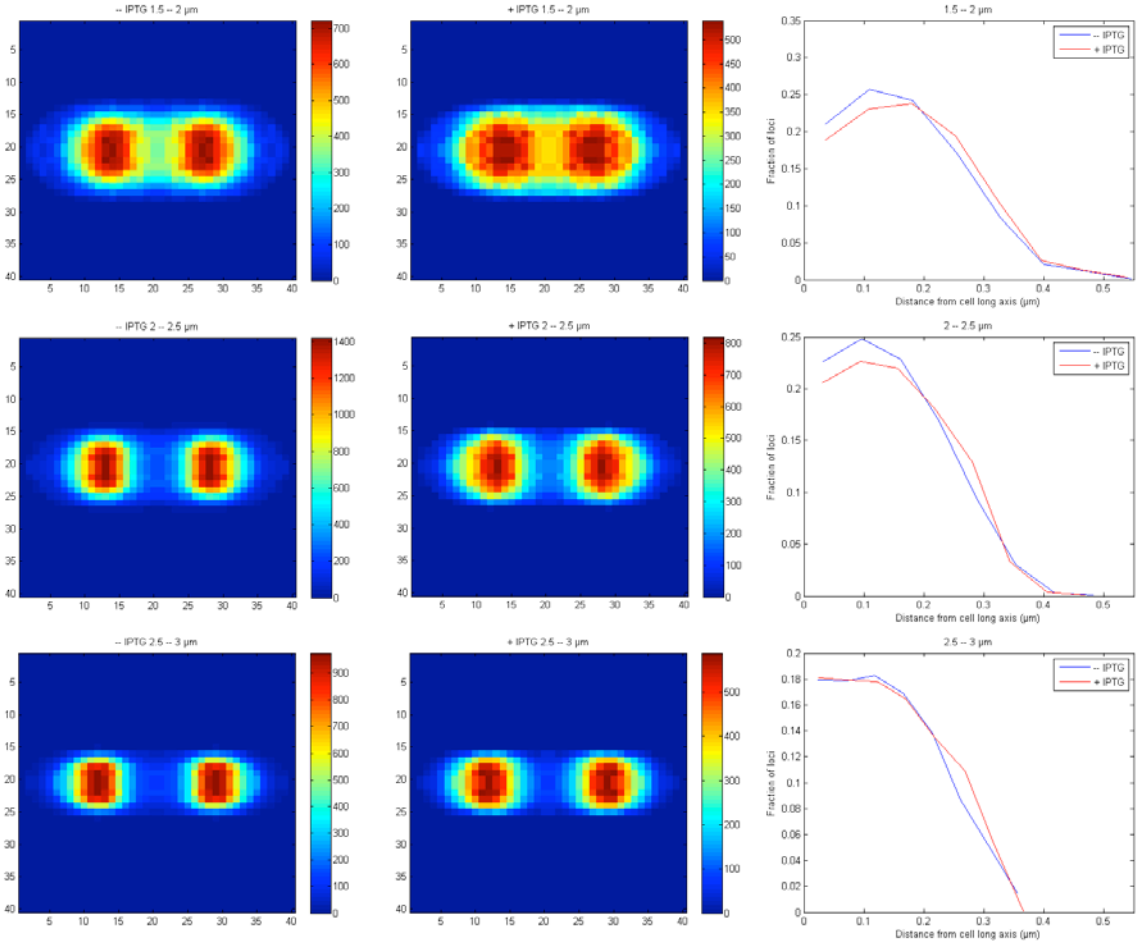

Figure 15: As Figure 14, but in M9 with amino acids and glycerol at 37 °C.

Table 2: The number of loci included in each of the plots in Figures 14-17.

Growth condition	Length bracket (µm)	Number of loci	
		– IPTG	+ IPTG
M9 glucose	2 - 2.5	308	278
amino acids 37 °C	2.5 - 3	1 260	1 040
	3 - 4	6 393	5 205
M9 glycerol	2 - 2.5	1 468	1 666
amino acids 37 °C	2.5 - 3	6 645	6 978
	3 - 4	27 009	26 137
M9 glucose	1.5 - 2	36 537	33 196
amino acids RT	2 - 2.5	48 853	33 236
	2.5 - 3	28 330	19 314
M9 glucose	1.5 - 2	35 785	9 438
37 °C	2 - 2.5	21 817	9 790
	2.5 - 3	8 718	4 782

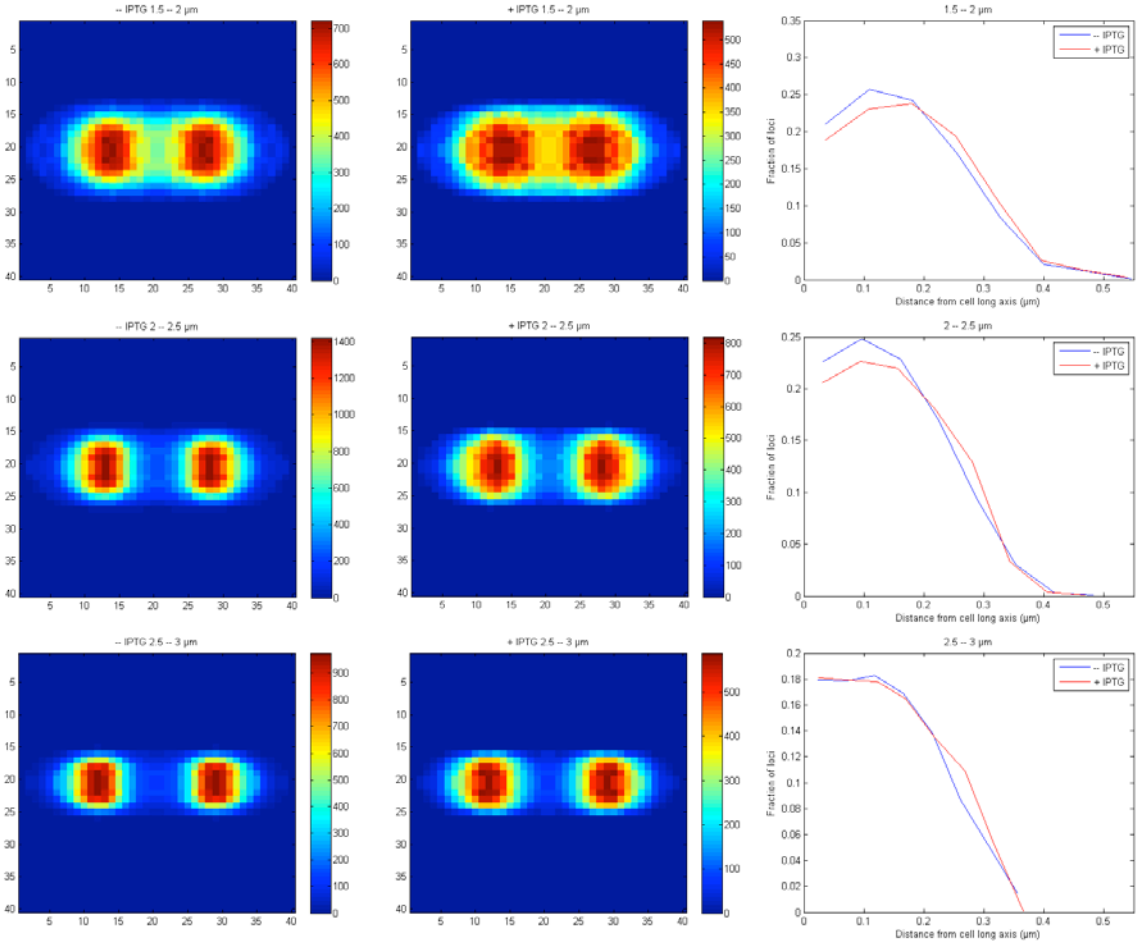

Figure 16: As Figure 14, but in M9 with amino acids and glucose at room temperature.

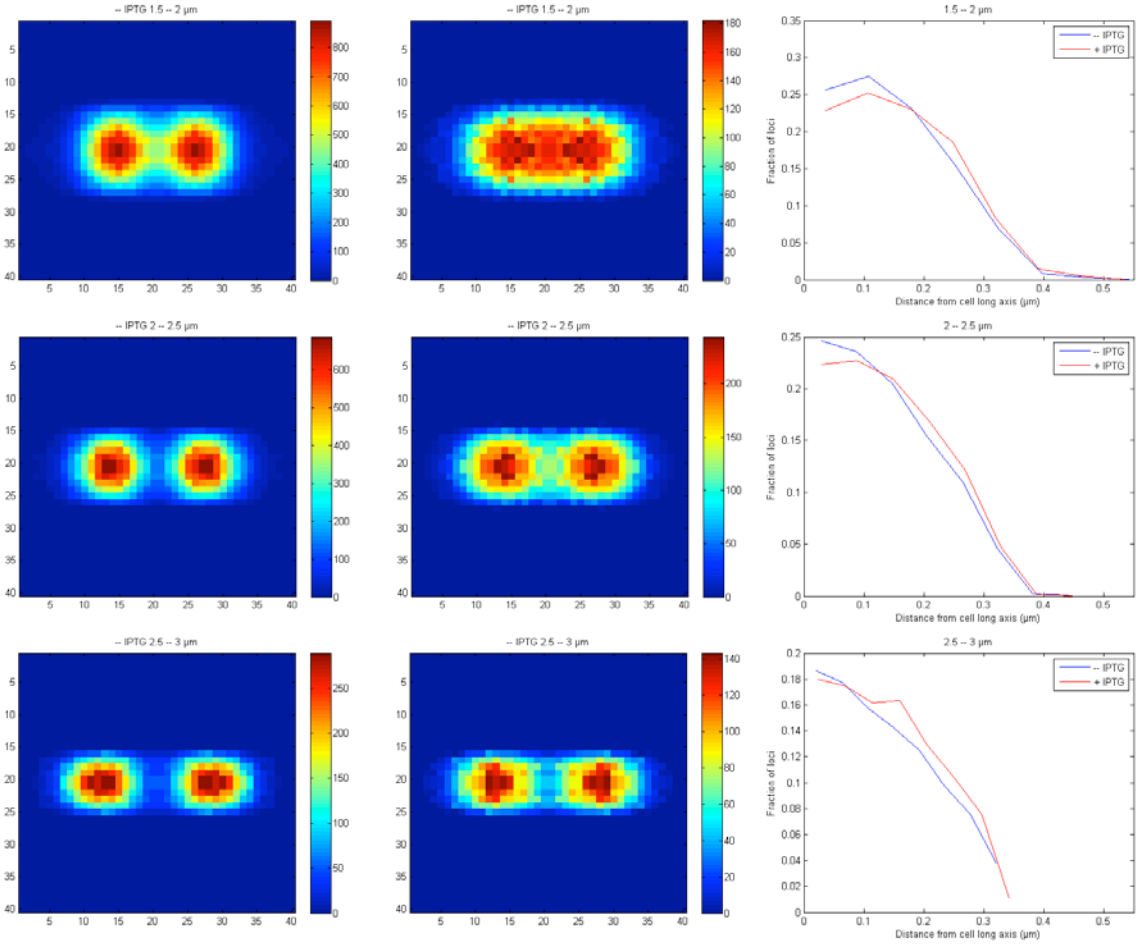

Figure 17: As Figure 14, but in M9 with glucose at 37 °C without supplementary amino acids.

DISCUSSION

The results from the transient induction experiments indicate that the induction response of the lac operon is linear for times between one and four minutes, which correspond to a constant rate of transcription initiation on population-level. This is in accordance with the current model of the regulation of the lac operon. Any other result would have been very difficult to reconcile with established knowledge about the operon. Above four minutes, assuming the transcription initiation rate is constant, the number of expressed proteins per cell will likely reach a plateau when the induction time approaches the bacterial generation time. For short inductions on the other hand, there results agrees with the hypothesis of an induction delay and provides a timeframe of 10-20 s. This is within the interval that would be needed to counter the repressor search time and give an additional repression directly after replication or when the repressor otherwise disassociates. However, such a delay cannot be confirmed definitely due to possible microfluidic issues in the spot counting experiment and lack of repeats. Methodwise is it very promising to see that counting of individual freely diffusing molecules works and that the approach developed here can measure expression at very high quantitative and temporal resolution.

According to all data on the position of the lac locus, it is located more closely to the membrane when actively transcribed. Even more striking is the fact that is seems strongly excluded from

the cell midline in cells growing on glucose with amino acids present. This has been observed earlier in glycerol with amino acids[11], but in that case together with a much larger movement of the whole distribution of loci outwards. In this study the overall change of the distribution is minimal in glycerol. It is possible that the difference is due to where the locus is tagged: upstream – as here – or downstream of the operon. The length of the DNA in between the tagged locations is 3 μm, which is actually about the size of an *E. coli* cell. It would indeed be very intriguing if the two ends of the genes have localizations that follow different patterns. In that case, one could imagine that the upstream location is more important for the transcription rate since this is where RNAP initiates, and the downstream location having some impact on the translation rate since this is where the nascent transcript will be still be attached when translation begins. Possibly, the transcript could also have a higher chance of degradation if the gene somewhere crosses an area where the ratio of nucleases to ribosomes is high where the transcript is unprotected for a longer time.

That the locus is present more often in the region between the segregating chromosomes in cells has to my knowledge not been observed before. This could be interpreted either as a differential location of the loci or as if the whole chromosome organizes differently. There is no reason for the chromosome to change organization upon addition of one specific inducer, so the explanation closest at hand is that the change relates to the *lac* operon specifically. Polymerase and ribosome presence in this region has to my knowledge not been very well studied and it would be very interesting to see if the region has a specific function in gene regulation.

Other more technical explanations for the discrepancies between these results and others can also not be ruled out. In this study all distances are based on the cell contours as identified by automatic segmentation, and calculated as distances from the cell centre. The earlier study[11] uses a fluorescent membrane protein to identify the membrane which of course is very reliable for finding the cell outlines but is does not relate any distances to the cell centre. If the translocation depends on a "push-away-from-centre" mechanism, I expect the method here to give clearer patterns while the opposite might be true if the loci are instead pulled to the membrane.

The mechanism that lies behind this phenomenon is very important for understanding how this relates to the presumed expression delay. If it is purely dependent on co-transcriptional membrane insertion of proteins as claimed[11](also termed *transertion*), then the operon would already need to be transcribed before any translocation can happen. It could serve to keep a higher transcription and translation rate if the gene is already active but not help to initiate transcription. Also, there is no apparent reason why transertion would locate the operon to the inter-nucleoid region. This does however not exclude other mechanisms working concurrently. Even if there is an effect pulling the loci to the membrane, it might also be moved out of the nucleoid upon initial derepression. The movement might be driven by a yet unknown force, or perhaps more likely by random motion upon which the gene is caught by a polymerase in the periphery of the nucleoid. Another option is that polymerases spread through the nucleoid initiates transcription but the transcript is likely to be degraded as long as it does not reach out of the nucleoid to the ribosome-rich regions and can be protected. When it does however, it will be pulled out together with the DNA by translating ribosomes to the cytosol. The transcripts made once in the cytosol could be translated to a much higher extent giving the induction level that we think of as normal in induced cells. These effects can of course be combined; a locus can

be pulled out and then pinned to the membrane by membrane proteins to keep it out in the cytosol.

A way to differentiate between these mechanisms would be to investigate the kinetics of the translocation, i.e. see how long time the change of the distribution takes. If the locus moves on a timescale of seconds is it most likely driven by a derepression- or transcription initiation-related mechanism. If the change of distribution pattern takes minutes, then it can be explained by membrane insertion only. This is possible to investigate using a experimental setup like the one used in this study, and would be very enlightening on this phenomenon.

Acknowledgements

The strain for labeling of the *lac* operon was constructed by Prune Leroy. I would like to thank my supervisors Petter Hammar, Prune Leroy and Johan Elf for great supervision and the opportunity to do this project. I am grateful to the entire Elf lab for assistance, discussions and advice, especially including David Fange and Fredrik Persson who have offered technical advice and assistance even at odd hours. I also want to show appreciation to my scientific reviewer Pia Lindberg and my opponent Marie Ottosson for helpful critical review of this work. Thank you!

References

1. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3,** 318–356 (1961).
2. Müller-Hill, B. *The lac Operon: a short history of a genetic paradigm*. (Walter de Gruyter, Berlin ; New York, 1996).
3. Oehler, S., Eismann, E. R., Krämer, H. & Müller-Hill, B. The three operators of the lac operon cooperate in repression. *EMBO J.* **9,** 973–979 (1990).
4. Hammar, P., Leroy, P., Mahmutovic, A., Marklund, E. G., Berg, O. G. & Elf, J. The lac repressor displays facilitated diffusion in living cells. *Science* **336,** 1595–1598 (2012).
5. 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).
6. Gilbert, W. & Müller-Hill, B. Isolation of the lac repressor. *Proc. Natl. Acad. Sci. U. S. A.* **56,** 1891–1898 (1966).
7. Hammar, P., Walldén, M., Fange, D., Persson, F., Baltekin, Ö., Ullman, G., Leroy, P. & Elf, J. Direct measurement of transcription factor dissociation excludes a simple operator occupancy model for gene regulation. *Nat. Genet.* (2014). doi:10.1038/ng.2905
8. Hammar, P. lac of Time: Transcription Factor Kinetics in Living Cells. (2013).
9. Ullman, G., Wallden, M., Marklund, E. G., Mahmutovic, A., Razinkov, I. & Elf, J. High-throughput gene expression analysis at the level of single proteins using a microfluidic turbidostat and automated cell tracking. *Philos. Trans. R. Soc. B Biol. Sci.* **368,** 20120025–20120025 (2012).
10. Li, G.-W. Single-Molecule Spatiotemporal Dynamics in Living Bacteria. (2010).
11. Libby, E. A., Roggiani, M. & Goulian, M. Membrane protein expression triggers chromosomal locus repositioning in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 7445–7450 (2012).
12. Dürrenberger, M., Bjornsti, M. A., Uetz, T., Hobot, J. A. & Kellenberger, E. Intracellular location of the histonelike protein HU in Escherichia coli. *J. Bacteriol.* **170,** 4757–4768 (1988).
13. Bratton, B. P., Mooney, R. A. & Weisshaar, J. C. Spatial distribution and diffusive motion of RNA polymerase in live Escherichia coli. *J. Bacteriol.* **193,** 5138–5146 (2011).
14. Bakshi, S., Siryaporn, A., Goulian, M. & Weisshaar, J. C. Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. *Mol. Microbiol.* **85,** 21–38 (2012).

15. Endesfelder, U., Finan, K., Holden, S. J., Cook, P. R., Kapanidis, A. N. & Heilemann, M. Multiscale Spatial Organization of RNA Polymerase in Escherichia coli. *Biophys. J.* **105,** 172–181 (2013).
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.* **97,** 6640–6645 (2000).
17. Nguyen, A. W. & Daugherty, P. S. Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nat. Biotechnol.* **23,** 355–360 (2005).
18. Shaner, N. C., Lambert, G. G., Chammas, A., Ni, Y., Cranfill, P. J., Baird, M. A., Sell, B. R., Allen, J. R., Day, R. N., Israelsson, M., Davidson, M. W. & Wang, J. A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. *Nat. Methods* **10,** 407–409 (2013).
19. Li, Y. & Altman, S. Polarity Effects in the Lactose Operon of Escherichia coli. *J. Mol. Biol.* **339,** 31–39 (2004).
20. Hansen, L. H., Knudsen, S. & Sörensen, S. J. The Effect of the lacY Gene on the Induction of IPTG Inducible Promoters, Studied in Escherichia coli and Pseudomonas fluorescens. *Curr. Microbiol.* **36,** 341–347 (1998).
21. Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H. R. & Iwata, S. Structure and mechanism of the lactose permease of Escherichia coli. *Science* **301,** 610–615 (2003).
22. 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).
23. Reidl, J., Römisch, K., Ehrmann, M. & Boos, W. MalI, 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.* **171,** 4888–4899 (1989).
24. Nagai, T., Ibata, K., Park, E. S., Kubota, M., Mikoshiba, K. & Miyawaki, A. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* **20,** 87–90 (2002).
25. Miller, J. H. *Experiments in molecular genetics*. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1972).
26. Yu, J. Probing Gene Expression in Live Cells, One Protein Molecule at a Time. *Science* **311,** 1600–1603 (2006).
27. Edelstein, A., Amodaj, N., Hoover, K., Vale, R. & Stuurman, N. in *Curr. Protoc. Mol. Biol.* (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K.) (John Wiley & Sons, Inc., Hoboken, NJ, USA, 2010). at <http://doi.wiley.com/10.1002/0471142727.mb1420s92>
28. Sliusarenko, O., Heinritz, J., Emonet, T. & Jacobs-Wagner, C. High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. *Mol. Microbiol.* **80,** 612–627 (2011).
29. Olivo-Marin, J.-C. Extraction of spots in biological images using multiscale products. *Pattern Recognit.* **35,** 1989–1996 (2002).