UPPSALA
UNIVERSITET

UPTEC X 19 034
Examensarbete 30 hp
Juni 2019

Development of a single-molecule tracking assay for the lac repressor in Escherichia coli

Oscar Broström

UPPSALA
UNIVERSITET

Teknisk- naturvetenskaplig fakultet
UTH-enheten

Besöksadress:
Ångströmlaboratoriet
Lägerhyddsvägen 1
Hus 4, Plan 0

Postadress:
Box 536
751 21 Uppsala

Telefon:
018 – 471 30 03

Telefax:
018 – 471 30 00

Hemsida:
http://www.teknat.uu.se/student

Abstract

Development of a single-molecule tracking assay for the lac repressor in Escherichia coli

Oscar Broström

Gene regulation by transcription factors are one of the key processes that are important to sustain all kinds of life. In the prokaryote Escherichia coli this has shown to especially crucial. The operator sequence to which these transcription factors bind to are very small in comparison to the whole genome of E. coli, thus the question becomes how these proteins can find these sequences quickly. One particularly well-studied transcription factor in this regard is the lac repressor. It has been shown that this transcription factors finds its operators faster than the limit of three dimensional diffusion. The leading model for how the repressor does that is facilitated diffusion and this model has gained more experimental evidence, particularly using single-molecule fluorescence microscopy.

This study aimed at measuring the unspecific binding time between the lac repressor and DNA in vivo, but in the end the project evolved to trying to establish a single-molecule tracking assay of the repressor in vivo. In this study a mutant of the repressor was expressed and purified, labelled with a synthetic fluorophore, electroporated into E. coli and tracking was performed under a microscope. One of the three types of experiments were partially analysed with an image analysis software. Unfortunately, analysis was not completed for all experiments which made it difficult to compare the results. In the end the data was compared by eye while also using the results from image analysis. With slight optimism it can be concluded that the assay worked, but it needs more development.

Handledare: Emil Marklund
Ämnesgranskare: Sebastian Deindl
Examinator: Jan Andersson
ISSN: 1401-2138, UPTEC X 19 034

Sammanfattning

Liv är någonting som är väldigt komplext. Det kan bland annat ses vid sjukdomar såsom cancer och Alzheimers. Trots att det har spenderats mycket tid och pengar har forskare ännu inte kunnat bota dessa sjukdomar. Liv kan fundamentalt ses som en väldigt stor mängd av kemiska reaktioner som måste befinna sig i perfekt harmoni, annars skulle det levande förvandlas till något dött. Hur kan den här harmonin uppstå? Det här är i sig en väldigt komplex fråga, men den kan delvis besvaras med reglering av de olika kemiska reaktionerna. Nu blir då frågan vad och hur sker den här regleringen? Paradoxalt nog är det vissa molekyler i några av de kemiska reaktionerna som har ansvar för regleringen. För att i sin tur förstå hur regleringen sker måste man först förstå de fundamentala komponenterna i liv: DNA som är molekylen som lagrar all information som krävs för liv, RNA som bland annat kan agera som en informationsbärare, men under kortare perioder jämfört med DNA och proteiner vilka är molekyler som ger liv många av dess olika funktioner. Förutom komponenterna måste man även förstå de processer som leder till att informationen kan bevaras under längre tid och hur informationen kan översättas till funktionella komponenter. De här processerna kallas för replikation, transkription och translation. Replikation kallas den process då DNA kopieras, transkription kallas den process där DNA läses av och utifrån bildas RNA som innehåller samma information som DNA:t och translation är den process då RNA används som kod för att bilda specifika proteiner. Alla dessa processer kan i någon form regleras, men mycket av den reglering som har studerats hittills har varit reglering av transkription.

Idéen om transkriptionsreglering publicerades i en stuide 1961 av två forskare vid namn Jacques Monod och François Jacob. I sin studie framförde de två teorier hur gener involverade i användandet av laktos som energikälla hos bakterien *Escherichia coli* (vanligtvis kallad *E. coli)* regleras. Den studien användes sedan som bas för vidare studier om just det här systemet och gav Monod och Jacob ett Nobelpris i medicin 1965. Vidare studier utmynnade bland annat i upptäckten och isoleringen av det protein som ansvarade för regleringen av de gener som ansvarar för laktosnedbrytning. Proteinet kallas idag för *lac* repressorn eller LacI och är en transkriptionsfaktor, vilket innebär att det är ett DNA-bindande protein som reglerar transkription, processen då DNA avläses till RNA. Då *lac* repressorn binder till specifika delar av DNA sitter den i vägen för det maskineri som översätter DNA till RNA. Då kan inte de gener som krävs för att utnyttja laktos som energikälla läsas av.

Idag är inte *lac* repressorn den enda kända transkriptionsfaktorn, i människor finns det till exempel 2600 kända transkriptionsfaktorer och de går att hitta i olika variationer i princip i alla former av liv. Att transkriptionsfaktorer är viktiga att förstå kan belysas av det faktum att många är involverade i olika typer av sjukdomar, exempelvis cancer. Förutom sjukdomar är vissa transkriptionsfaktorer involverade i utvecklingen av antibiotikaresistens, ett av de största hoten mot oss människor.

Trots att man har kunnat identifiera många transkriptionsfaktorer och lärt sig mycket annat om dem så är det fortfarande inte helt klargjort hur dessa förhållandevis små proteiner kan lyckas hitta till de specifika delarna av DNA som de har som uppgift att binda. Det här är ett svårt problem i och med att de specifika delarna av DNA är väldigt korta jämfört med längden på allt DNA som finns i en organism. I *E. coli* finns det 4,6 miljoner baspar av DNA. Den specifika ordningen av dessa baspar är det som bär information i DNA. I *lac* repressorns fall letar den efter en kombination av 24 baser. Att repressorn skulle hitta de rätta baserna av rena slumpen är väldigt liten, dessutom skulle det ta väldigt lång tid. Forskare har studerat den här sökprocessen under en väldigt lång tid och en vanlig teknik som används är enmolekylsfluorescensmikroskopi. Tekniken baseras på att använda kraftfulla mikroskop för att se hur enskilda fluorescenta molekyler beter sig i realtid. En anledning till att man vill studera en molekyl i taget är att, precis som vi människor, beter sig molekyler olika. Till exempel, människor som deltar i en löpartävling springer från start till mål, men hur många springer till höger om trädet eller tar kurvan väldigt tätt? Den här typen av information går endast att erhålla om man studerar varje löpare var för sig och samma sak gäller för molekylers rörelse. Från det som går att se i mikroskopet kan en dator användas för att få fram information bland annat om hur molekylerna rör sig, om de binder något och hur fort de rörde sig. En annan anledning är att vissa typer av molekyler finns i väldigt liten uppsättning i celler. Det blir då missvisande att studera dessa molekylers beteende i ett provrör där andelen molekylär kan vara mer än 1000 gånger så mycket som i en cell.

Tidigare har forskare kunnat identifiera att *lac* repressorn letar efter sina specifika baser genom att både flyta runt i sin tredimensionella omgivning och att binda varsomhelst på DNA:t och glida längs DNA:ts yta. Det som händer är att repressorn söker av ytan efter den rätta sekvensen. Den här sökprocessen har studerats både inne i *E. coli* celler men även utanför celler, men än så länge har man inte direkt lyckats uppmäta hur lång tid i genomsnitt *lac* repressorn binder till ospecifik DNA-kod inne i *E. coli*. Den här studien hade som mål att mäta den här tiden, men inom den givna tidsramen lyckades inte den tiden mätas. Istället presenteras ett protokoll för hur man kan göra enmolekyls-experiment med *lac* repressorn i levande *E. coli*. Lärdomarna som erhållits i den här studien kan används till vidare studier med *lac* repressorn, men även för andra transkriptionsfaktorer, vilket i sin tur kan leda till utveckling av nya mediciner eller seger i striden mot antibiotikaresistens.

Table of contents

Abbreviations

1D	one dimension/one dimensional
3D	three dimensions/three dimensional
AIC	Akaike's information criterion
APD	avalanche photodiode
APS	ammonium persulphate
bp	base pairs
CV	column volume
Da	Dalton
DMF	dimethylformamide
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
EMCCD	electron-multiplying charge-coupled device
EMSA	electrophoretic mobility shift assay
FCS	fluorescence correlation spectroscopy
GST	glutathione S-transferase
HMM	hidden Markov model
IPTG	isopropyl β-D-1-thiogalactopyranoside
LacI	*lac* repressor
LAP	linear assignment problem
LB	lysogeny broth
MSD	mean-square displacement
ONPF	orthonitrophenylfucoside
ORF	open reading frame
POE	Per Object Ellipse fit
POI	protein of interest
RDM	Rich Defined Medium
RNA	ribonucleic acid
RNAP	RNA polymerase
rpm	revolutions per minute
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
TCEP	tris(2-carboxyethyl)phosphine
TEMED	tetramethylethylenediamine
TSA	thermal shift assay

1 Introduction

Life from a bottom-up perspective can be seen as a combination of chemical reactions that in harmony tries to avoid reaching thermodynamic equilibrium. One could imagine these chemical reactions as parts, similarly to the parts in car or any other mechanical machine. One of the key features of modern mechanical machines is that they need to be able to respond to external stimuli through some means of regulation. Life is similar in this sense that it requires forms of regulation to remain alive. Regulation occurs generally, but not necessarily during some of the main processes required to sustain life, namely replication, transcription and translation. Regardless, without regulation life would likely not exist, or it would look very different from how we currently see and define life. It is therefore crucial to understand the regulatory processes if we want to understand how life works and what it is.

There are many factors that can regulate life, but one of the key regulatory elements across all domains of life are transcription factors. Transcription factors are deoxyribonucleic acid (DNA) binding proteins that bind to specific DNA sequences leading to either activation or repression of transcription. When transcription factors activate transcription they do so by interacting with parts of the transcriptional complex, causing transcription initiation (Latchman 1997). Repression on the other hand can work in a couple of different ways. Traditionally repression was thought to occur by physically blocking ribonucleic acid (RNA) polymerase (RNAP) from binding to the promoter. However, other mechanisms have been discovered such as inhibition of transitions between different transcriptional complexes and by preventing RNAP from leaving the promoter region (Rojo 2001).

Today, transcriptional gene regulation is a heavily researched area, but up until the mid-20th century the idea that gene expression could be regulated was still not discovered. In 1961 Jacob & Monod were able to observe a regulatory system for gene expression when studying the lactose system in *Escherichia coli*. In their paper Jacob & Monod (1961) proposed two models for how the genes involved in lactose metabolism were regulated (see Figure 1). Today, we know that the repressor proposed in the model is a protein, something that Jacob & Monod did not and that model one coincides better with experimental results (Lewis 2005). Nowadays the repressor is called the *lac* repressor or LacI and the genes it regulates are part of the *lac* operon (Lewis 2005). From their discovery Jacob & Monod received the Nobel Prize in 1965 since it was the first direct observation of gene regulation and to this day LacI still remains one of the most studied transcription factors (Kipper *et al.* 2018).

Due to their importance transcription factors are often involved in diseases such as cancer, diseases impairing human development and hormonal diseases (Latchman 1997), therefore understanding their expression patterns and their mechanism could lead to development of new drugs and therapies. Transcription factors are not only involved in diseases, some are involved in antibiotic resistance (Ramos *et al.* 2005, Deochand & Grove 2017) which is a large threat to global health (Ventola 2015). This means that understanding how transcription

factors work could lead to ways to combat antibiotic resistance. Despite this study focusing on LacI similar methods could be applied to other transcription factors. The regulatory part of the *lac* operon is also commonly used in expression cassettes, both in industry and academia (Rosano & Ceccarelli 2014) which means that understanding the mechanism of LacI could lead to more options for the people who design expression cassettes.

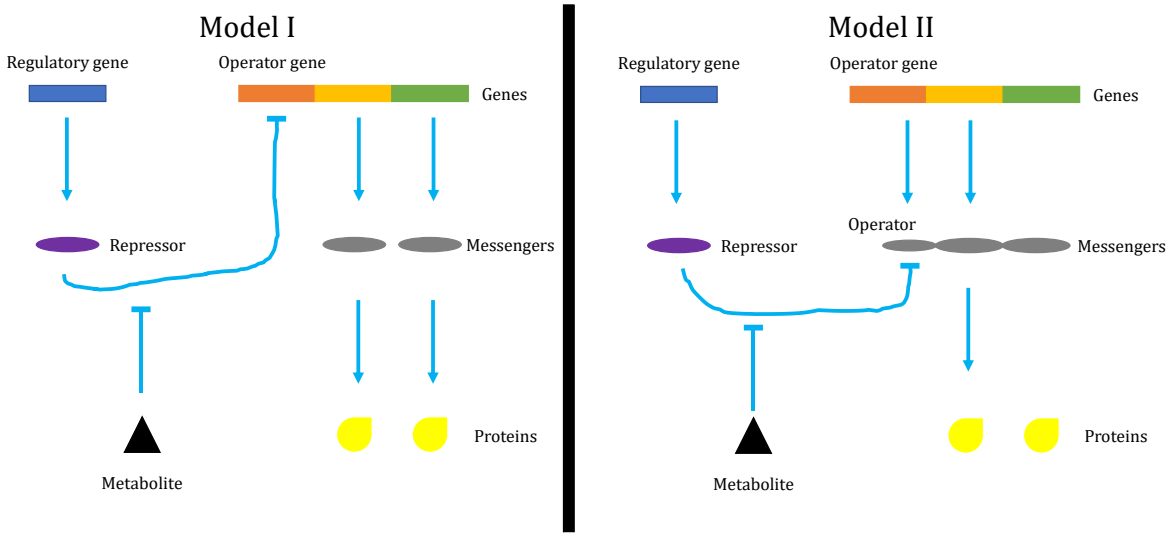

Figure 1. An adaptation of the two models that Jacob & Monod (1961) proposed for regulation of expression of the proteins found within the *lac* operon. In both proposed models a gene coding for a regulatory factor is converted to a repressor (either through transcription or transcription and translation, Jacob and Monod did not know what the repressor was when the two models were conceived). The repressor either acts by binding the operator gene (model I) or the mRNA transcribed from the operator gene (model II). The metabolite can interact with the repressor which destabilises the interaction to the operator, which opens up for RNAP to transcribe the genes within the operon.

One thing that has puzzled scientists for a long time is how transcription factors can find their operons within the vast genome of any organism. The confusion grew larger when Riggs *et al.* (1970) found that for LacI the kinetics for the search process exceeded what is possible with three dimensional (3D) diffusion alone. Because the search process is essential for understanding how transcription factors work scientists started to investigate this conundrum, both through experimental and theoretical means. One crucial paper was published in 1981 by Berg *et al.* where they proposed the theory of facilitated diffusion, a theory which first involves 3D diffusion and one dimensional (1D) sliding along DNA. Although only theoretical the paper inspired further studies and the model is still being investigated through a wide variety of means (Hammar *et al.* 2012, Marklund EG *et al.* 2013, Furini & Domene 2014, Mahmutovic *et al.* 2015, Tempestini *et al.* 2018, Marklund E *et al.* 2018).

Single-molecule fluorescence microscopy has been one of the main techniques used to study how LacI finds its operator sequence. This technique has been applied both *in vitro* and *in vivo*, and *in vitro* LacI has been shown to move ~50 base pairs (bp) while rotating around the DNA helix, which is longer than the helical periodicity of DNA (10.5 bp) (Marklund E *et al.* 2018). *In vivo* LacI has been shown to slide 45 ± 10 bp on DNA, that the repressor is sterically hindered by other proteins bound to DNA and that the repressor tends to miss the naturally occurring operator O_1 several times before binding (Hammar *et al.* 2012).

Although these studies and many others have contributed to knowledge about LacI's search mechanism there are still many unanswered questions that remain, especially in terms of the mechanism *in vivo*. This study aims to answer the question regarding for how long LacI binds unspecific DNA sequences *in vivo*, which is required for the repressor to find its operator sequence. To do this a LacI mutant is labelled with the synthetic fluorophore bifunctional rhodamine B (Corrie *et al.* 1998) *in vitro* before reintroducing the labelled repressor into *E. coli* using electroporation. This facilitates the possibility to track the labelled LacI *in vivo* from which it is possible to deduce the unspecific binding time when performing data analysis while at the same time knowing the orientation of LacI in respect to DNA. The orientation can be determined by the fact that bifunctional rhodamine B is covalently bound to two cysteines. With two covalent bonds the dye cannot rotate around its own axis enabling the use of polarised light to determine the orientation of the dye. The dye's orientation can then be used to infer the orientation of LacI (Marklund E *et al.* 2018).

What this study aims to achieve is to deepen the understanding of how the *lac* repressor finds its operator sequence. The initial goal was to measure the binding time between the repressor and any unspecific DNA sequence *in vivo*. However, within the given timeframe development of a single-molecule tracking assay of electroporated LacI was achieved. In the future, the information brought forth in this study can help to answer what the unspecific binding time for LacI is and can act as a guide when studying other transcription factors.

2 Background

This section aims to provide background on the some of the properties of LacI, the principle behind single-molecule microscopy, how LacI used and can still be studied without single-molecule techniques, provide insight into some of the different options that are available when making a protein of interest fluorescent and a general outline of how the image analysis is done.

2.1 The *lac* repressor

After the conceptual idea of LacI by Jacob & Monod (1961) it took a couple of years before there was any experimental evidence of its existence. However, when it was finally isolated by Gilbert & Müller-Hill (1966) it was conclusively confirmed that LacI is important for gene regulation. From that point the repressor was further characterised and it was shown that the native state of the repressor is a homotetramer, with each subunit containing 360 amino acids and the total molecular weight of the tetramer being approximately 154 kDa (Lewis 2005). The tetrameric structure can be seen in Figure 2A. Each monomer consists of a few different domains that fulfil particular functions (see Figure 2B for structure). The N-terminal domain (residues 1–45) is the domain that binds DNA and is shortly followed by a hinge region (residues 46–62) (see Figure 2C). The hinge leads into the core of the protein (residues 63–

329). The role of the core is to bind an inducer which could be allolactose (processed form of lactose) or isopropyl β-D-1-thiogalactopyranoside (IPTG). Besides binding inducer, the core also creates a dimer interface that is ~2200 $Å^2$ with another monomer, forming a dimer. There are five key regions in the core that make up the interface. The residues are the following: 70–100, 221–226, 250–260 and 275–290. After the core comes a small hinge (residues 330–338) which leads into the C-terminal domain (residues 339–360). The fold of the C-terminal domains enables interactions of each monomer to form the tetrameric structure (Lewis 2005, Kipper *et al.* 2018).

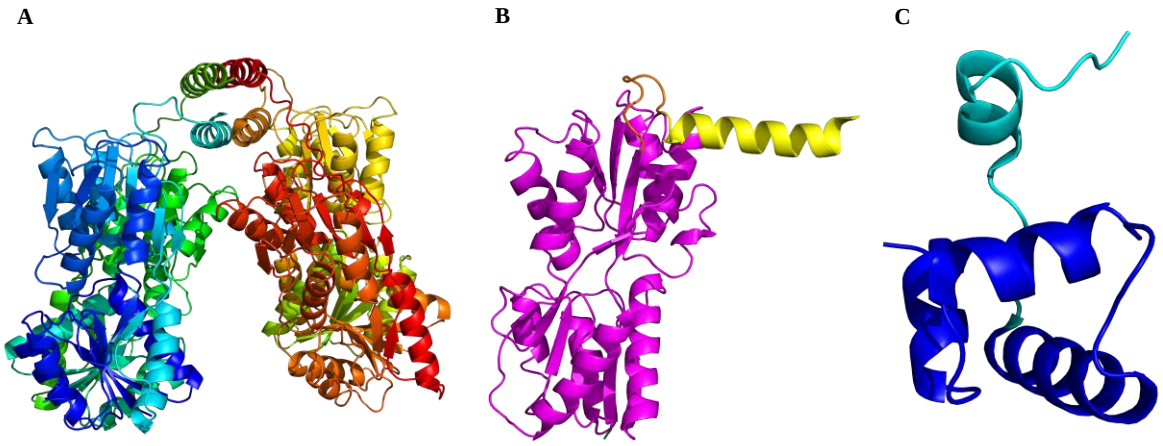

Figure 2. Structure of the *lac* repressor. A) Tetrameric structure of the *lac* repressor as determined by X-ray crystallography with a 2.7 Å resolution (PDB-ID: 1LBI). The N-terminal domain and the hinge between the N-terminal domain and the protein core is missing from the structure (Lewis *et al.* 1996). B) Monomeric form of the *lac* repressor shown in A). In magenta the core domain is shown, in orange the hinge between the core and the C-terminal domain is shown and the C-terminal domain is shown in yellow (PDB-ID: 1LBI) (Lewis *et al.* 1996). C) N-terminal domain (blue) and hinge region (cyan) of a monomeric *lac* repressor bound to an operator sequence and the anti-inducer orthonitrophenylfucoside (ONPF) (not shown in image) with 2.6 Å resolution (PDB-ID: 1EFA) (Bell & Lewis 2000).

In the tetrameric state LacI has the ability to bind two different operators at the same time. These operators do not have to be right next to each other, which is evident from the fact that *E. coli* has multiple operator sites: O_1, O_2 and O_3 with O_1 being the primary operator and O_2 and O_3 auxiliary operators located 401 bp downstream and 92 bp upstream respectively from O_1 (Lewis 2005). When LacI binds two operator sites a looped complex with DNA and LacI is formed, an important factor to ensure full repression as mutation in one or both auxiliary operators yields lower rates of repression (Lewis 2005, Kipper *et al.* 2018). Another consequence of LacI being able to bind two operator sites is that if the C-terminal domain is disrupted causing the tetramer to dissolve into two dimers the dimers still retain the ability to bind an operator site, although with lower repression rate due to the inability to loop DNA. However, if the dimeric interface is disturbed the protein remains monomeric and in the monomeric state the protein does not have very high affinity for DNA (Kipper *et al.* 2018).

The fact that dimeric LacI can bind DNA while the monomeric form cannot somewhat hints at LacI being able to tolerate some mutations (Kipper *et al.* 2018). This becomes further evident from the fact that point mutations in LacI have been heavily studied, which has shown

that the core in particular is resistant to mutations, and that some residues even in the N-terminal (DNA-binding domain) domain can be mutated without reduced activity (Kleina & Miller 1990, Markiewicz *et al.* 1994, Suckow *et al.* 1996, Pace *et al.* 1997). The implications of this is that it is most likely possible to design LacI mutants that are suitable for testing many different hypotheses and that these mutants can be purified easily (Kipper *et al.* 2018).

2.2 Single-molecule fluorescence microscopy

As mentioned in the introduction single-molecule fluorescence microscopy has been an important tool for elucidating the search mechanism of LacI, but why have things been done on the single-molecule level? Could it not be done by just observing the population as a whole? What is attractive about single-molecule techniques is that population-based methods cannot capture any heterogeneity in a population, and for most populations there is always some heterogeneity. For example, heterogeneity can in many instances be crucial in explaining why a system behaves in a certain way. Single-molecule techniques can observe these heterogeneities and thus deepen our understanding of their implications on the population, but also why they happen in the first place (Shashkova & Leake 2017). An example of heterogeneity can be found with bacterial colonies where different cells can have different expression patterns in case of a sudden shift in environment which could eliminate the colony if the colony was homogenous. This strategy is called bet-hedging (Davis & Isberg 2016).

Single-molecule fluorescence microscopy is not one unified technique, rather it can be done in several different ways and the applications are plentiful. The main goal of this section is not to list every single technique, but rather describe the techniques that have been performed and those that were planned for as well. Therefore, the following sections will cover wide-field microscopy, confocal microscopy and fluorescence correlation spectroscopy (FCS).

2.2.1 Wide-field microscopy

Wide-field microscopy is the most commonly used way to do microscopy in biology. This is due to its simplicity in that the light is never manipulated while traveling from the light source to the specimen and then onto the detector (Thorn 2016). In Figure 3 the main components of a wide-field microscope are shown. The first component in the microscope is the light source. In the case of fluorescence microscopy lasers are preferred since they provide high intensity light at specific wavelengths, where the wavelength of choice is dependent on the fluorophore in the specimen. After the light source there is an excitation filter that filters any light that is not of a particular wavelength. This filter is particularly important if the light source is not a laser (during single-molecule experiments the light source is always a laser). The next component is a dichroic mirror that only reflects light of a certain wavelength; therefore, it is very important that the mirror is compatible with the light source. Once the light is reflected off the dichroic mirror it reaches the objective which focuses the light onto the specimen. Because of this the specimen will start fluorescing and the emitted light travels back through the objective, passes through the dichroic mirror and then through an emission filter that

absorbs light of undesired wavelength. The filtered emitted light then goes to the ocular and/or the detector. The detector is connected to a computer to record and store the signal (Thorn 2016).

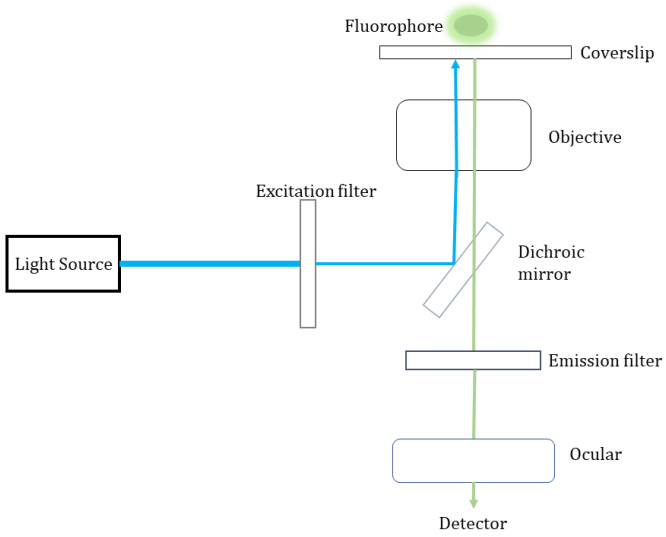

Figure 3. Schematic of a typical wide-field microscope adapted for fluorescence microscopy. The light source (typical a laser) produces light that first passes through an excitation filter to filter out any unwanted wavelengths of light. The light hits a dichroic mirror, which only reflects light of a certain wavelength. The reflected light is directed towards the objective which focuses the light onto the specimen. This causes the specimen to fluoresce and the emitted light passes back through the objective and right through the dichroic mirror as well before being filtered of any undesired wavelengths by an emission filter. Now the light reaches the end of its journey by going into the ocular and/or the detector. The detector records the signal and the data is stored in a computer that is connected to the detector.

Although wide-field microscopes are sufficient for many questions asked by scientists the main problem with the technique is the relatively high abundance of out-of-focus light, which limits the resolution, a crucial point when doing single-molecule microscopy. For that reason other techniques have emerged to try to solve the problem with out-of-focus light, one of them being confocal microscopy (Thorn 2016).

2.2.2 Confocal microscopy

To solve the problem with out-of-focus light confocal microscopy employs two pinholes and, in some cases, scanning mirrors. As can be seen in Figure 4 one pinhole is placed in between the first filter and the dichroic mirror and the other is placed between a lens and the detector. The role of the first pinhole is to filter out all of the out-of-focus light of the excitation beam while the second pinhole takes care of the out-of-focus light from the emitted light beam (Shashkova & Leake 2017). For the pinhole to block all or most of the out-of-focus light the diameter of the pinhole must be really small, so small that only a fraction of the specimen is observed, which equates to only one pixel on the detector. To get the complete image further manipulation must be done. That is why scanning mirrors are used. They ensure that the whole specimen is scanned by the excitation beam leading to the formation of a complete image. This scanning process is often referred to as raster scanning and is quite time consuming since the image is formed pixel by pixel (Thorn 2016). Besides the pinholes and the scanning mirrors a lens is added right before the second pinhole (see Figure 4). The

purpose of this lens is to focus the light towards the second pinhole which means that only the in-focus-light will pass through the pinhole while all the out-of-focus light is absorbed and/or reflected by the material surrounding the hole (Combs 2010).

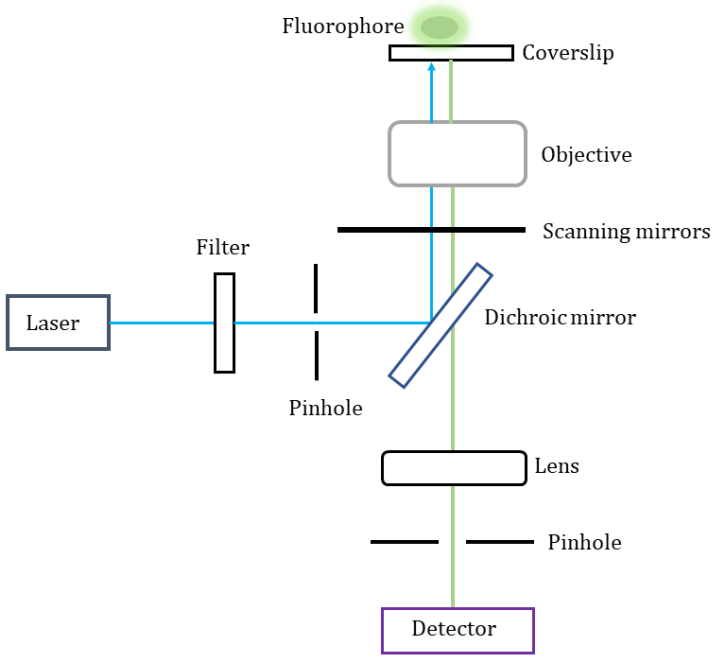

Figure 4. **Schematic of a confocal microscope setup.** A laser is used to generate a beam of excitation light that first passes through a filter to remove any undesired wavelengths. The beam then reaches a pinhole where only the in-focus-light passes through the hole while all the out-of-focus light is filtered away. The in-focus beam is then reflected by a dichroic mirror that only reflects light of a certain wavelength(s). Generally, due to the small size of the pinhole only a fraction of the specimen will be hit by the excitation beam, therefore scanning mirrors are put in place so that the whole sample can be scanned. The objective is there to focus light on the sample which starts fluorescing soon after it is hit by the excitation beam. The emitted beam then travels back through the objective, past the dichroic mirror and into a lens that focuses the light on the second pinhole. Again, the pinhole only lets in-focus-light through and the in-focus light hits the detector which records the signal and the signal is stored in the computer. The scanning mirrors are moved around until a complete image of the sample is formed.

2.2.3 Fluorescence correlation spectroscopy

FCS can be seen as an extension of confocal microscopy, but it utilises statistical analysis of fluctuations in fluorescence intensities in the very small detection volume (sub-femtolitre generally). The intensity fluctuations are then used to calculate an autocorrelation curve based on Equation 1 (Ries & Schwille 2012):

$$G(\tau) = \frac{\langle \delta F(t) \delta F(t+\tau) \rangle}{\langle F(t) \rangle^2} \quad (1)$$

Where $F(t)$ is the fluorescence intensity at the time t, $\langle F(t) \rangle = (1/T) \int_0^T F(t)dt$ is the average fluorescence signal over a given time T, $\delta F(t) = F(t) - \langle F(t) \rangle$ is how much each fluorescence signal at time t differs from the mean and τ is referred to as the time lag (Ries & Schwille 2012). Although this is the fundamental basis of FCS there are many variants of FCS that can lead to different results. In traditional FCS the detection volume is kept stationary

which only yields information about translational diffusion. However, FCS can achieve much higher temporal resolution through the use of avalanche photodiode (APD) detectors (Ries & Schwille 2012). High temporal resolution can be a very important factor when studying certain biological mechanisms. For instance, in a paper by Marklund E *et al.* (2018) a nonconventional type of FCS was utilised to learn about the orientation and rotation of LacI when bound to DNA while at the same time measure the translational diffusion constant. What was different in the approach of Marklund E *et al.* (2018) was that they moved their lasers to track the repressor while it was diffusing along DNA. The tracking setup in conjunction with APDs made it possible for them to measure the rotational diffusion coefficient of LacI without mistaking it for the rotation of DNA around its own axis and acquire the translational diffusion constant.

From the study of Marklund E *et al.* (2018) it becomes evident that the principle of FCS is very powerful. Besides being able to measure translational and rotational diffusion the technique can also be used to determine local concentrations, association and dissociation constants, rate constants and provide information on structural dynamics. Another flexible point of the technique is that it can both be used *in vitro* and *in vivo* (Ries & Schwille 2012).

2.3 Studies of LacI using population-based methods

Before the advent of single-molecule techniques studies were limited only to population-based methods. One population-based method that has been used is the electrophoretic mobility shift assay (EMSA). The assay is based on the core concept of gel electrophoresis. The difference is that once a protein binds to a nucleic acid the mobility of that complex within the gel will decrease. The assay is run under native conditions to ensure binding between the protein of interest (POI) and the nucleic acid. Visualisation is done either through radioactive labelling or through fluorescence. The method is both qualitative and can be used quantitatively as well, which was its primary use when it was developed to study the interaction between LacI and operator DNA (Garner & Revzin 1981, Fried & Crothers 1981). Kinetic information is acquired by titrating the amount of protein as the mobility decreases as the number of bound proteins increases and the necessary quantification of protein/DNA is done in each band is done using autoradiograms or their fluorescence intensity (Fried & Crothers 1981).

Another population-based method is the membrane filter assay. The core part of the assay is a membrane filter made out of nitrocellulose which is negatively charged. Due to its charge it can bind positively charged molecules, for instance proteins, which commonly have a positive net charge. In the case of DNA-binding proteins they bind negatively charged DNA that by itself would not be able to find to the membrane filter. With labelled DNA it is possible to detect protein-DNA complexes on membranes as only protein with and without DNA is able to bind, while unbound DNA passes through the filter. This makes it possible to quantify the amount of protein-DNA complex depending on the labelling strategy (radiolabelling or

fluorescence). Kinetics can be determined if the concentration of DNA is kept constant while the concentration of protein is increased incrementally (Riggs *et al.* 1970).

2.4 Fluorophore labelling of a protein

Perhaps the most important thing for single-molecule fluorescence microscopy is the fluorescent label. There are however many types and approaches when labelling proteins, all of them brings their own advantages and disadvantages. Some of the available approaches are the following: a fusion to a fluorescent protein, polypeptide fusions that can bind fluorescent molecules (two common polypeptides are the HALO and SNAP tags) and covalent binding of a synthetic fluorophore (Sustarsic *et al.* 2014).

The most common of the three approaches is the use of fusion proteins. Although easy to use due to the fact that the gene of the fluorescent protein is placed within the same open reading frame (ORF) as the POI fusion proteins lack some characteristics necessary for detailed mechanistic studies of proteins. Perhaps the most crucial thing is their size, which often are comparable with the size of the POI. The implications of this is that the activity of the POI can either be reduce or completely abolished. Additionally with the added size the diffusion of the fusion will not be comparable to how fast the POI normally diffuses (Kipper *et al.* 2018). Fluorescent fusion proteins are considered to have poor photophysical qualities such low photostability meaning that they photobleach easily and they are not very bright which can make it difficult to distinguish them from a noisy background (Sustarsic *et al.* 2014, Kipper *et al.* 2018).

The use of tags such as HALO and SNAP are also possible and does not necessarily cause the same issues on the photophysical side of things, however the tags are still relatively large, which can be a deterring factor in some applications. Additionally, these tags cannot be added anywhere in the POI as well further limiting the use of these tags (Sustarsic *et al.* 2014).

Lastly, synthetic fluorophores which compared to fusion proteins and tags are much smaller making it less likely that they would interfere with the activity of the POI. Furthermore, the smaller size impacts where in the POI the label can be placed, further increasing the flexibility of synthetic fluorophores. Synthetic fluorophores are also more attractive due to their photophysics. For instance, they are brighter, making it easier to distinguish them from the background and they are also more resistant to photobleaching, which makes it possible to have longer imaging sessions, a necessity to get the complete picture for some mechanisms. However, the problem with synthetic fluorophores is that they require a labelling step *in vitro* regardless of whether the labelled protein will be used *in vitro* or *in vivo*, which can be quite cumbersome (Sustarsic *et al.* 2014, Kipper *et al.* 2018).

Synthetic fluorophores are versatile in regard to where in the POI they can be placed. Therefore, it becomes very important to understand the principles behind labelling. Normally, there are two types of chemistries to form covalent bonds between the dye and the protein;

maleimide chemistry occurs between the dye and a cysteine and succinimide chemistry occurs between the dye and lysines. Regardless of the chemistry it is important to make sure that the target residue is not important for the protein's function. For instance, there might be cases where the target residues are functional, where there are no residues of interest in the protein or where there is an abundance of target residues, which would increase the risk of off-target binding; if so it might be necessary to add or remove certain residues to ensure that labelling with a synthetic fluorophore can occur. However, it is then important to make sure that only non-functional residues are removed and that the addition of new residues do not alter the functionality of the protein (Kipper *et al.* 2018).

The fact that LacI is not sensitive to mutations (Kleina & Miller 1990, Markiewicz *et al.* 1994, Suckow *et al.* 1996, Pace *et al.* 1997) enables a plethora of options for dye placement, making LacI a suitable candidate for fluorophore labelling. Furthermore, LacI does not contain many cysteines compared to lysines (3 cysteines compared to 11 lysines), which makes maleimide chemistry a compelling option as long as the cysteines do not have functional significance and are susceptible for forming covalent bonds with any dye. Kipper *et al.* (2018) showed that this indeed was a viable strategy by screening for mutants that remained functional and structurally intact when two out of the three native cysteines were mutated to alanine. Additionally, the mutation S28C could be done with retained activity. This new cysteine was where they attached their synthetic fluorophore.

2.5 Single-particle tracking

This section will briefly introduce the steps necessary to be able to track individual fluorescent molecules over a period of time *in vivo*. In-depth information can be found in the cited sources. The general process used in this study is based on what Volkov *et al.* (2018) did, however some minor changes were done to better fit this study.

2.5.1 Cell segmentation

The first step necessary for single-particle tracking is to define spatial constraints for the particle. Since this is an *in vivo* study the spatial constraint should be the volume of one cell. For that reason, each cell has to be distinguished from everything this not a cell but also from other cells since the molecules should not be able to move in between cells. To segment cells their borders are identified from phase-contrast images using the algorithm presented by Ranefall *et al.* (2016). The algorithm is called "Per Object Ellipse fit" (POE). Based on the name of the algorithm it tries to fit objects to the shape of an ellipse. The algorithm does so by assigning scores to each object it encounters ranging from 0 to 1 with 1 being a perfect ellipse. Since *E. coli* is rather elliptical the algorithm is well suited to segment these bacterial cells. An example of how the cells look before and after segmentation can be seen in Figure 5.

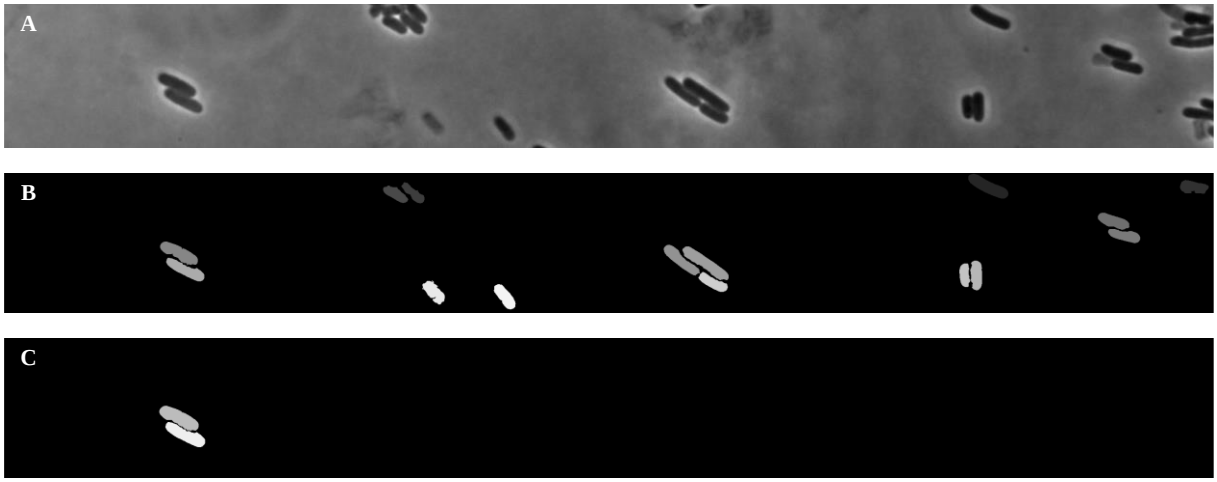

Figure 5. Example of cells before and after cell segmentation. A) Phase contrast image where the image was been oriented correctly according to the bright-field image. No segmentation has been done at this point. B) Segmented cells without applying any manual curation. As can be seen, the algorithm does not detect all cells in A). C) Segmentation after manual curation, such as removing dead cells, removing cells that were not within the laser beam and cells that were partially out-of-frame.

2.5.2 Detection and localisation of spots

The fluorescent spots that appear in the recorded time-lapse have to be detected and localised by the image analysis pipeline. This is done by radial symmetry-based spot detection (Loy & Zelinsky 2003). The algorithm finds points of interest that have high radial symmetry in an image. The spots formed by a single fluorescent molecule are rather circular, often referred to as dots and are therefore rather symmetric. For that reason, this algorithm is suitable when trying to detect fluorescent dots in single-molecule fluorescent microscopy.

With the spots detected they must be localised, meaning that their x- and y-coordinates have to be determined. This becomes more difficult in single-molecule tracking experiments where the molecules are constantly moving. Not only are the particles constantly moving, but the quality of each frame can vary significantly making things even more difficult. For that reason it is necessary to have an algorithm that can deal with this. Lindén *et al.* (2017) developed an algorithm based on a symmetric Gaussian spot model in conjunction with the maximum a posteriori fit. Applying this algorithm results in data that is more accurate.

2.5.3 Building trajectories

Now that the fluorescent molecules can be properly localised the next step is to connect the frames that are recorded during the time-lapse. This process is called trajectory building and in essence it is a game of connect the dots, albeit trickier due to several factors. For instance, the dots might disappear for a few frames due to photoblinking; the event of two particles moving too close to each other is also problematic since the resolution is limited by diffraction, therefore frames where particles seemingly fuse and frames where one “fusion” dot separate into two have to be dealt with. Lastly, the noisy background makes the task more challenging. Luckily, there are algorithms developed to build trajectories, some of them being able to deal with the abovementioned problems. One of these algorithms is the u-track

algorithm (Jaqaman *et al.* 2008). It works by connecting the particles one frame at a time which forms track segments. These segments are then connected by closing the gaps caused by particle disappearance, resolving events of particle fusion and particle separation. The core structure of the algorithm is the linear assignment problem (LAP), which is a combinatorial optimisation problem where there is an equal number of tasks and workers and one worker can only do one task. Each worker has an associated cost for each task, thus the goal is to minimise the cost (Gandibleux *et al.* 2017). Jaqaman *et al.* (2008) used a LAP approach for both the frame connection step and the gap closing, particle fusion and particle separation step.

2.5.4 Extracting diffusion coefficients

From the generated trajectories it is possible to extract diffusions coefficients for each particle. This can be done by calculating the mean-square displacement (MSD) of each particle which is defined as in Equation 2 and from that the diffusion coefficient can be calculated using Equation 3:

$$MSD = \frac{1}{N}\sum_{n=1}^{N}(r_n(t) - r_n(0))^2 \quad (2)$$

$$MSD = 2dD\Delta t \quad (3)$$

Where N are the number of particles/molecules, r contains both the x- and y-coordinates of the particle, t is time in Equation 2, Δt is the time step meaning the time between each frame, d is the dimensions ($d = 2$ in this study) and D is the diffusion constant.

MSD shows how much a particle is displaced between the time step Δt. For each time step a MSD-value can be calculated. The values can then be plotted in a graph with MSD on the y-axis and the time step on the x-axis. Curve fitting is then done and if the curve is linear the particle is following Brownian motion as is illustrated in Equation 3 where the slope of the curve is the diffusion constant D multiplied by $2d$. Judging from Equation 3 the curve should pass through the origin, but in reality, this is usually not the case due to errors in localisation of the particle due to for instance the fact that the particle is moving (presumably) and the levels of noise during data acquisition (Ernst & Köhler 2012).

2.5.5 Hidden Markov modelling

The diffusion constants on their own do not provide any biological context. It is however possible to speculate to what a particular diffusion constant might mean. For instance, if there are two diffusion constants where one is smaller than the other the smaller one could indicate that the molecule is bound to something, while the larger one might be diffusing freely. Pure speculation is not enough, therefore characterisation must be done more robustly. A useful tool for this task is hidden Markov models (HMMs). When an HMM is used to analyse single particle tracking the trajectories are modelled as random transitions between unknown states. The diffusion constants are what define the unknown states (Volkov *et al.* 2018).

HMMs contain a set of observations, a set of states, probabilities for transitions between the states and probabilities for observing each observation. The word hidden in HMMs stems from the fact that the states in the model are unobservable, so are the transitions as well. The Markov part of the name is that the model has to satisfy being a Markov process, which means that transitions between the hidden states are independent from each other, thus the model is memoryless (Ghahramani 2001). In an HMM the number of states must be defined when a model is set up. This can be difficult in a biological context because the number of diffusive states might not be known beforehand (Sgouralis & Pressé 2017). One way to deal with this is Akaike's information criterion (AIC). AIC is used to estimate how much information that is lost when using a particular model and is defined as in Equation 4:

$$AIC = -2l + 2K \tag{4}$$

Where l is the maximised log-likelihood function and K is the number of parameters that will be estimated. The fact that AIC estimates the information loss of a model it becomes important to compare the AIC values between models to be able to tell which one of the models that is the most suitable one (Posada & Buckley 2004).

What the HMM in this study does is that it uses the trajectories and their transitions in the model, takes account of any point-wise localisation error and motion blur and learns the transition patterns of the data. Additionally, the HMM does maximum-likelihood inference of the parameters in the model and AIC as described above is used to find the number of states the model should have (Posada & Buckley 2004, Lindén *et al.* 2017, Volkov *et al.* 2018).

3 Material and methods

The goal here was to first express and purify LacI. LacI was then labelled with bifunctional rhodamine B to be able to use single-molecule fluorescence microscopy. Excess dye is removed through a purification step to remove false positives stemming from free dye diffusion. Since LacI had never been electroporated before suitable conditions for the technique had to be found. This was done through buffer screening. The his-tag on the LacI mutant could potentially affect how LacI diffuses, thus it advised to remove the his-tag, which was done using tobacco etch virus (TEV) protease. To verify the activity of LacI an EMSA was used. Electroporation was carried out to introduce labelled LacI into *E. coli*. To track labelled LacI the cells were put under a microscope.

3.1 Overexpression of LacI

BL21(DE3) cells containing the plasmid pD861.LacIDim2-TEV-His6 (see sequence in Appendix A) was grown overnight at 37 °C, 190/200 revolutions per minute (rpm) (both were used during the study) in lysogeny broth (LB) supplemented with 50 µg/mL kanamycin. The

overnight culture was diluted 1:100 with LB (total volume 1 L) in 5 L flasks, again supplemented with 50 µg/mL kanamycin. The cells were grown until OD_{600} reached 0.6–0.8 at 37 °C, 120 rpm. Before induction, a small fraction of culture was transferred to a separate flask and kept uninduced. To induce expression 20% (w/v) of L-rhamnose to a final concentration of 0.2% (w/v) was added to the culture. The culture was put back into the incubator at 37 °C, 120 rpm for 3 h. 1 mL of both induced and noninduced culture was pelleted by spinning at max speed for 5 min. The pellets were stored at -80 °C. The remaining large cultures were centrifuged for 50 min at 5000 rpm. Most of the supernatant was discarded (left 20–50 mL of media). The remaining media was used to resuspend the pellets. The cells were then centrifuged for 10 min at 4600 g. Supernatants were discarded and the pellets were flash frozen in liquid nitrogen and then stored at -80 °C until further use.

3.2 Purification of LacI

An ÄKTA Pure system (GE Healthcare) placed in a 4 °C cold room was used for purification. Wash buffer (buffer A) containing 20 mM Na_2HPO_4, 500 mM NaCl and 20 mM imidazole at pH 7.4 was prepared. Additionally, elution buffer (buffer B) made of 20 mM Na_2HPO_4, 500 mM NaCl and 500 mM imidazole at pH 7.4 was also prepared. These buffers were sterile filtered with 0.22 µm filters. 30% of EDTA-free (ethylenediaminetetraacetic acid) protease inhibitor (the pill was dissolved in 1 mL Milli-Q water) was added to both buffers on the day of chromatography.

The previously prepared pellets were taken out of the -80 °C freezer and thawed at 4 °C. Once thawed the pellets were resuspended in 7 mL of buffer A. To begin cell lysis 500 µL of 10 mg/mL lysozyme, 20% EDTA-free protease inhibitor and 1 µL of Pierce Universal Nuclease (Thermo Fisher) was added. This was incubated on ice for 40 min. The lysate was then passed through a cell homogeniser twice and then centrifuged for 1 h at 3800 g, 4 °C. The supernatant was filtered through a 0.45 µm filter and stored on ice or at 4 °C until purification.

Purification was done with a 5 mL HisTrap HP (GE Healthcare) column with the flow rate being set to 1 mL/min. The column was first washed with 1 column volume (CV) of buffer B and then equilibrated with buffer A. The supernatant was then loaded onto the column. The column was washed with buffer A until baseline was reached. After washing LacI was eluted with buffer B using a gradient over 90 min. 3 mL fractions were collected. Once done the fractions were kept at 4 °C.

Based on the chromatogram some of the fractions were analysed with denaturing sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). Samples were prepared to a 1:1 ratio with Laemmli buffer (Bio-Rad). The denaturing process began by incubating the samples at 95 °C for 20 min. Then the samples were spun down quickly at max speed and then incubated again at 95 °C for 20 min before being spun down again. The samples were then loaded onto 4–20% Mini-PROTEAN TGX Precast Gels (Bio-Rad). The gels were run at

200 V for 30–35 min. Once complete, the gels were stained with InstantBlue (Expedeon) for 20 min with light agitation. To visualise the gels and take photos of them a ChemiDoc Imaging System (Bio-Rad) was used. The fractions containing LacI were then pooled.

3.3 Labelling LacI with bifunctional rhodamine B

To label LacI with bifunctional rhodamine B (Corrie *et al.* 1998) buffer was first exchanged to labelling buffer (100 mM HEPES, 500 mM NaCl and 10% glycerol at pH 7.4) using Amicon Ultra 15 mL 10 kDa Centrifugal Filters (Millipore). LacI was quantified by measuring its absorbance at 280 nm using a NanoDrop (Thermo Fisher). A 10 mM tris(2-carboxyethyl) phosphine (TCEP) (Sigma-Aldrich) with pH 7.4 stock was prepared in labelling buffer. The stock was added to the solution of LacI so that the concentration of TCEP was 10x greater than that of LacI. The mixture was then degassed at room temperature for 30 min. Meanwhile, N,N´-Bis[2-(iodoacetamido)ethyl]-N,N´-dimethylrhodamine (bifunctional rhodamine B, Santa Cruz Biotechnology, for structure see Figure 6) was thawed at room temperature for 30 min and then dissolved in 60 µL dimethylformamide (DMF). 1.5x or 5x bifunctional rhodamine B was added to the concentration of LacI and the reaction was carried out at room temperature in the dark with light agitation for 1.5 h. 5 mM β-mercaptoethanol was used to quench the labelling reaction.

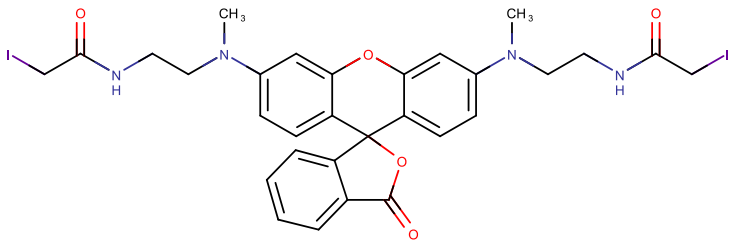

Figure 6. 2D chemical structure of bifunctional rhodamine B. The structure was drawn in MarvinSketch 19.10 (ChemAxon) according to structure 3 in Corrie *et al.* (1998).

Unreacted dye was removed by first mixing the labelled LacI solution with 750 µL of HisPur Cobalt Resin (Thermo Fisher). The mixture was incubated with light mixing for 1 h in the dark at 4 °C. After incubation the resin-protein mixture was transferred to a gravity flow column. Before washing the column was drained of liquid. Washing was done with ~45 CV buffer A (for recipe see Section 3.2). LacI was then eluted with buffer B (for recipe see Section 3.2) in six 1 mL fractions and a larger fraction afterwards. The flow-through, washes and fractions were analysed by denaturing SDS-PAGE. 5x SDS-PAGE sample buffer was mixed with the protein solution so that the final concentration of sample buffer was 1x. The samples were then denatured by incubating them at 95 °C for 20 min and then spun down at max speed. This step was done twice. Mini-PROTEAN 4–20% TGX Precast Gels (Bio-Rad) were used and the gels were run at 200 V for 30–35 min. The gels were analysed with a

ChemiDoc Imaging System (Bio-Rad). Fractions containing labelled protein were pooled and the buffer was exchanged to labelling buffer using Amicon Ultra 15 mL 10 kDa Centrifugal Filters (Millipore).

Impure protein was concentrated as above and loaded on an illustra NAP-10 column (GE Healthcare). The column was drained and then equilibrated with 15 CV labelling buffer before 1 ml of sample was loaded onto the column. The protein was eluted by sequentially adding 200 µL of labelling buffer. Fractions of 200 µL were collected. The fractions were analysed with SDS-PAGE in the same manner as after the Ni-column purification (for results see Figure 16 in Appendix C). Fractions containing protein were pooled. To get rid of possible impurities from the gel filtrated solution it was purified again, this time using Ni-NTA His-Select resin (Sigma-Aldrich). 200 µL of resin was mixed with the pooled fractions and then incubated at 4 °C in the dark with light mixing for 75 min. After incubation the resin-protein mixture was transferred to Mini Bio-Spin Chromatography Columns (Bio-Rad). The columns were then centrifuged for 2 min at 700 g. The columns were then washed with buffer A (for recipe see Section 3.2) and spun for 2 min at 700 g. This step was repeated five times. LacI was then eluted in four steps using buffer B (for recipe see Section 3.2) with the same centrifuge settings. The flow-throughs, the washes and the eluted fractions were analysed by SDS-PAGE as described above (for results see Figure 17 in Appendix C). Eluted fractions that contained labelled LacI were pooled and concentrated using Amicon Ultra 15 mL 10 kDa Centrifugal Filters (Millipore) and all of the labelled LacI was pooled and quantified using A_{280} and A_{547} on a NanoDrop (Thermo Fisher). Amicon Ultra 15 mL 10 kDa Centrifugal Filters (Millipore) were used for buffer exchange to electroporation buffer (100 mM HEPES, 15.625 mM NaCl and 50% glycerol at pH 7.4) and Amicon Ultra 0.5 mL 3 kDa Centrifugal Filters (Millipore) were used for concentrating LacI further. The concentration of labelled protein was then quantified using a NanoDrop (Thermo Fisher).

3.4 Buffer screening

To try to find a suitable buffer for electroporation several buffers were screened. This was done manually and through a thermal shift assay (TSA).

3.4.1 Manual screening

A variety of buffers were mixed and tested to see whether LacI would precipitate or not. In Table 1 a list of the tested buffers can be seen. How the tests were conducted can be seen below the table.

Table 1. Mixed buffers that were used in testing for a buffer that could be suitable for electroporation, help determine which parameters are important or limit the search range within a specific parameter. How the screening was done is described below the table.

Buffer	Buffering agent	Salt/salts	Crowding agent	pH
1	20 mM HEPES	10 mM NaCl 1 mM $MgCl_2$	-	7.4
2	100 mM HEPES	150 mM NaCl	10% Glycerol	7.4
3	100 mM HEPES	150 mM KCl	10% Glycerol	7.4
4	100 mM HEPES	100 mM NaCl	10% Glycerol	7.4
5	100 mM HEPES	50 mM NaCl	10% Glycerol	7.4
6	100 mM HEPES	25 mM NaCl	10% Glycerol	7.4
7	100 mM HEPES	100 mM KCl	10% Glycerol	7.4
8	100 mM HEPES	50 mM KCl	10% Glycerol	7.4
9	100 mM HEPES	25 mM KCl	10% Glycerol	7.4
10	5 mM HEPES	-	10% Glycerol	7.4
11	5 mM HEPES	150 mM NaCl	10% Glycerol	7.4
12	5 mM HEPES	100 mM NaCl	10% Glycerol	7.4
13	5 mM HEPES	50 mM NaCl	10% Glycerol	7.4
14	5 mM HEPES	25 mM NaCl	10% Glycerol	7.4
15	55.67 mM HEPES	250 mM NaCl	50% Glycerol	7.4
16	77.78 mM HEPES	125 mM NaCl	50% Glycerol	7.4
17	88.89 mM HEPES	62.5 mM NaCl	50% Glycerol	7.4
18	94.44 mM HEPES	31.25 mM NaCl	50% Glycerol	7.4
19	97.22 mM HEPES	15.625 mM NaCl	50% Glycerol	7.4
20	98.61 mM	7.8125 mM NaCl	50% Glycerol	7.4

Buffer 1 was tested by trying to change buffer to it in an Amicon Ultra 15 mL 10 kDa Centrifugal Filters (Millipore); unlabelled protein was used when testing. For buffers 2 and 3 they were respectively used in doing a 1:10 dilution of unlabelled protein. The protein was in buffer B (for recipe see Section 3.2) when the dilutions were made. The diluted protein was then concentrated in Amicon Ultra 0.5 mL 3 kDa Centrifugal Filters (Millipore). Continuing, the buffer was then changed using the same filters. Buffers 4–9 were tested by changing into them using Amicon Ultra 0.5 mL 3 kDa Centrifugal Filters (Millipore); unlabelled protein was used. Buffer 10 was tested by diluting unlabelled protein 1:10 and 1:100. Several tubes were prepared of the dilutions and one set was centrifuged for 10 min at 14000 g and 4 °C immediately after mixing while one set was centrifuged after 10 min incubation at 4 °C and one set after 1 h of incubation at 4 °C. The tubes were then incubated for approximately 3 h and then centrifuged for 40 min at 14000 g and 4 °C. Then the buffer was exchanged to buffer 10 using protein that had not been diluted and protein that had been diluted 1:10. Buffer exchange was performed in Amicon Ultra 0.5 mL 3 kDa Centrifugal Filters (Millipore). Buffers 11–14 were tested by changing buffer into them using Amicon Ultra 0.5 mL 3 kDa Centrifugal Filters (Millipore). In the test unlabelled protein was used. When the exchange was complete the samples were transferred to Eppendorf tubes and then centrifuged for 10 min at 14000 g and 4 °C to see whether pellets of protein was formed. The samples were centrifuged again, now for 1 h at 14000 g and 4 °C. For buffers 15–20 labelled protein was used. The buffers were first tested in order starting with 15. Buffer exchange was done using an Amicon Ultra 15 mL 10 kDa Centrifugal Filter (Millipore). The same filter was used when initially testing the buffers. Buffer 19 was further tested by changing from labelling buffer (for recipe see Section 3.3) to it in one step using Amicon Ultra 15 mL 10 kDa Centrifugal Filters (Millipore).

3.4.2 Thermal shift assay

To a 96-well plate 20 µL of the content in a JBScreen Solubility HTS (Jena Bioscience) was transferred using a multichannel pipette. From a 5000x stock solution SYPRO Orange was diluted to 50x in Milli-Q. To the plate the 50x solution was added so that the final concentration of SYPRO Orange was 5x. The assay was run either with LacI or lysozyme. When the assay was run with LacI the final concentration of LacI in the plate was about 13 µM while for lysozyme the final concentration was 1 mg/ml. The total volume in each well was 25 µL. Three wells were used as controls besides the control in the JBScreen Solubility HTS (Jena Biosciences). In one well no protein and dye was added, in another well only protein was added and in one well only dye was added. The plates were then sealed with adhesive plastic to prevent evaporation and then centrifuged for 1 min at 500 rpm. The plate was then placed in a CFX Connect Real-Time PCR Detection System (Bio-Rad). In the machine the plate was first incubated for 30 s at 15 °C. The temperature was then increased with 1 °C increments every 30 s until the temperature reached 95 °C.

3.5 His-tag removal

To remove the his-tag from LacI TEV protease was used since the LacI mutant had a TEV recognition sequence upstream of the his-tag sequence. 1:100 (w/w) of TEV protease (Sigma-Aldrich) was added to labelled LacI. This mixture was dialysed against labelling buffer (for recipe see Section 3.3). Dialysis was performed by hydrating a 10 kDa Slide-A-Lyzer (Thermo Fisher) in labelling buffer for 5 min. The LacI-TEV solution was then added to the dialysis cassette. The sample was left to dialyse for 1 h before the used buffer was discarded and new buffer was added. After another hour buffer was changed again and dialysis was done overnight. Dialysis was carried out at 4 °C with stirring.

Separation of the TEV protease, the his-tag and purified LacI was done using HisPur Cobalt Resin (Thermo Fisher) and Mini Bio-Spin Chromatography Columns (Bio-Rad). 400 µL of resin was mixed with the protein solution and incubated at 4 °C in the dark with light stirring for 2 h. After incubation the mixture was transferred to the columns and centrifuged for 2 min at 700 g. The columns were then washed with 2 CV buffer A (for recipe see Section 3.2) and centrifuged for 2 min at 700 g. Five washing steps were performed. 2 CV buffer B (for recipe see Section 3.2) and 2 min centrifugation at 700 g was used for elution. Three elution steps were performed. Fractions were analysed by SDS-PAGE, both using Laemmli as in Section 3.2 and with SDS sample buffer as in Section 3.3 (for results see Figure 14 and Figure 15 in Appendix C). The gels were imaged using a ChemiDoc Imaging System (Bio-Rad). Fractions containing labelled LacI were pooled and using Amicon Ultra 15 mL 10 kDa Centrifugal Filters (Millipore) the buffer was exchanged to electroporation buffer (for recipe see Section 3.3). LacI was then concentrated with an Amicon Ultra 0.5 mL 3 kDa Centrifugal Filter (Millipore). Concentration was determined using a NanoDrop (Thermo Fisher) at A_{280} and A_{547}.

3.6 Activity validation of LacI

An EMSA was used to verify that LacI was active before introducing it into cells. Initially, 8% polyacrylamide gels were prepared. The recipe that was used can be found in Table 2. With this recipe the TEB concentration is only 0.5x. It should be 1x, thus it is advised to used 1.2 ml of 10x TEB instead.

Table 2. 1x 8% Polyacrylamide gel recipe for EMSA. The TEB buffer contains tris, boric acid and EDTA. APS stands for ammonium persulphate. TEMED stands for tetramethylethylenediamine. The final concentration of TEB in this recipe will be 0.5x. The running buffer contains 1x TEB and it is therefore advised to use 1.2 mL TEB instead.

Component	Volume
30% acrylamide/bis solution, 37.5:1 (Bio-Rad)	3.2 mL
Milli-Q	7.6 mL
10x TEB buffer	0.6 mL
10% APS	400 µL
TEMED	20 µL

Labelled LacI without and with a his-tag and unlabelled LacI was analysed with an EMSA. Functionality was tested using DNA containing the O_1 operator that is labelled with Cy5. The DNA concentration was 10 nM. 0.5–1000 nM of labelled LacI with the his-tag cleaved off was used. The protein was diluted using electroporation buffer (see Section 3.3 for its content). For the unlabelled LacI 0.5–500 nM was used. Again, electroporation buffer was used for dilutions, however two wells containing 500 nM protein were prepared one where electroporation buffer was used to dilute and one where a buffer containing 100 mM HEPES, 250 mM NaCl, 10% glycerol at pH 7.4 was used to dilute the protein. This buffer was used for one well with 500 nM labelled LacI with its his-tag remaining as well. One well with 500 nM labelled LacI with the his-tag diluted in electroporation buffer was also prepared. In all buffers 1 mM of β-mercaptoethanol was added. The gels were run for 1 h at 80 V. The gels were stored overnight in dH_2O before imaging them on a ChemiDoc Imaging System (Bio-Rad).

3.7 Electroporation of labelled LacI

Labelled LacI with a his-tag was used for electroporation. To start with EZ Rich Defined Medium (RDM) (Teknova) was thawed and 1 mm Electroporation cuvettes (Molecular BioProducts) were pre-chilled on ice. ElectroMAX DH5α-E Competent Cells (Thermo Fisher) and electrocompetent DH5α containing a pSMART plasmid with an array of O_1 operators were withdrawn from a -80 °C freezer and thawed on ice. As controls 100 µM Cy3-labelled oligo, electroporated DH5α without adding any protein or DNA and DH5α with LacI added that were not electroporated. For the positive control 1 µL of Cy3-labelled primer was added to 20 µL of DH5α, 0.3 µL of LacI was added to 20 µL of cells. Cells were electroporated using a 1.8 kV pulse on a MicroPulser Electroporator (Bio-Rad). Immediately after electroporation cells were resuspended in 1 mL EZ RDM and left to recover at 37 °C, 190 rpm for 30/45 min. The cells were then harvested by centrifugation for 3 min at

5000/13400 rpm. The supernatant was discarded, and the pellets were resuspended in PBS, 100 mM NaCl, 0.005% X100 Triton, 50 µg/mL proteinase K (Thermo Fisher, pre-heated without proteinase K at 37 °C before) and then incubated at 37 °C, 190 rpm for 20 min. The cells were harvested by centrifugation for 3 min at 5000 and/or 13400 rpm and washed three times with PBS, 100 mM NaCl, 0.005% X100 Triton or EZ RDM. After the last centrifugation step the pellets were resuspended in 50 µL EZ RDM. OD_{600} was measured on a NanoDrop and the cultures were diluted to $OD_{600} = 0.02$ (the pathlength used is in mm) with EZ RDM, 1 µM SYTOX Blue dead-cell stain (SYTOX Blue was omitted before the assay was fully set up). 0.6 µL of diluted culture was then applied to 2% agarose pads made with EZ RDM. The pads sometimes contained 1 µM SYTOX Blue. A coverslip was applied once the cultures had dried on the pad.

3.8 Microscopy

3.8.1 Agarose pads

SeaPlaque GTG Agarose (Lonza) was mixed with EZ RDM (Teknova) to a final agarose concentration of 2%. The mixture was vortexed and then incubated at 65 °C on a heat block until the agarose was completely dissolved. In the meantime, a gene frame (Thermo Fisher) was attached to a microscope slide. To each side of the gene frame additional tape was added to not make the pad too thin. To the melted agarose 1 µM SYTOX Blue was added and 100 µL was pipetted within the gene frame. A plastic cover was put on top of the agarose and to press the agarose down something heavy was put on top of the slide. Once the agarose had dried and the cell samples were ready the weight and plastic cover was removed the pads were cut up with a scalpel. The rest of the agarose was removed, and cells were applied to the pad according to Section 3.7. Once the cells had dried a clean coverslip was gently attached to the strongly adhesive tape of the gene frame and light pressure was applied to ensure that an airtight seal had formed.

3.8.2 Optical setup

Microscopy was carried out on an inverted Nikon Ti-E microscope with a CFI Plan Apo Lambda 100X 1.45 NA Objective (Nikon). Phase contrast images were taken using a DMK 38UX304 camera (The Imaging Source). For bright-field and fluorescence images/movies an iXon Ultra 888 (Andor) electron-multiplying charge-coupled device (EMCCD) camera was used. As the source for white light TLED+ (Sutter Instrument) was used. To be able to identify dead cells with SYTOX Blue a 405 nm laser (Cobolt 0.6-MLD) was used. The laser intensity to was set to either 15 or 100 mW. For tracking of bifunctional rhodamine B labelled LacI a 532 nm laser (Cobolt 0.6-DPL) was used. The laser intensity was set to either 150 or 200 mW and the exposure time was set to 10 ms. The lasers were shuttered with an AOTFnC (AA-Opto Electronic). The microscope was controlled with µManager v. 1.4.20 (Edelstein *et al.* 2010). In some experiments image acquisition was made automatic using a custom made µManager plugin. 50 positions were imaged on each agarose pad when acquisition was automatic. The experiments were performed at room temperature.

4 Results

4.1 The dimeric LacI mutant can be properly isolated

By growing the *E. coli* strain BL21(DE3) with the plasmid pD861.LacIDim2-TEV-His6 (for sequence see Appendix A) in LB media supplied with kanamycin it was possible to overexpress the dimeric LacI mutant using L-rhamnose (the amino acid sequence for the mutant and native LacI is found in Appendix B). After overexpression it was possible to purify the protein through affinity chromatography. To verify that it was the LacI mutant that was expressed SDS-PAGE was utilised. The gels can be seen in Figure 7.

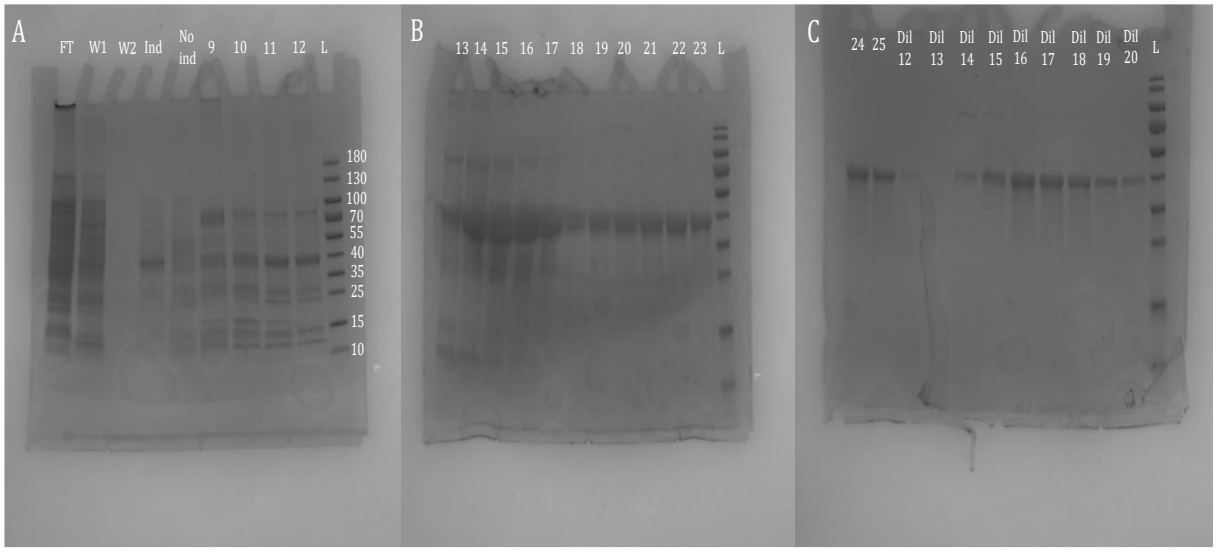

Figure 7. SDS-PAGE gels stained with Laemmli after large-scale protein expression. A) The annotated abbreviations are the following: FT = flow through, W1 = Wash 1, W2 = Wash 2, Ind = Induced cells, No ind = Non-induced cells, L = Ladder (applies to B and C). C) Dil = A 1:10 dilution of the numbered fraction with elution buffer. The ladder that was used for all three gels was PageRuler Prestained Protein Ladder (Thermo Fisher). The molecular weight of monomeric LacI is ~38 kDa and there is a band in between 35 and 40 kDa band on the ladder.

4.2 Bifunctional rhodamine B can be used to label a mutant LacI

The purified LacI was labelled with bifunctional rhodamine B and then purified through affinity purification to remove any unreacted dye. An SDS-PAGE (see Figure 8) was run on the purified sample to verify that the labelling had worked and that the unreacted dye was removed. Major bands can be seen around 40 kDa and between 70 and 100 kDa. The reason for the two bands is that two products are formed during the labelling reaction. The ~40 kDa product stems from the dye binding to Cys36 and Cys43, the two cysteines that were added to the protein for labelling. The band between 70 and 100 kDa contains the by-product of the dye binding a cysteine on one monomer while the other functional group binds a cysteine on another monomer, which essentially crosslinks two monomers. This product is not desired since the diffusion of two cross-linked LacI would be different from desired product due to the size difference.

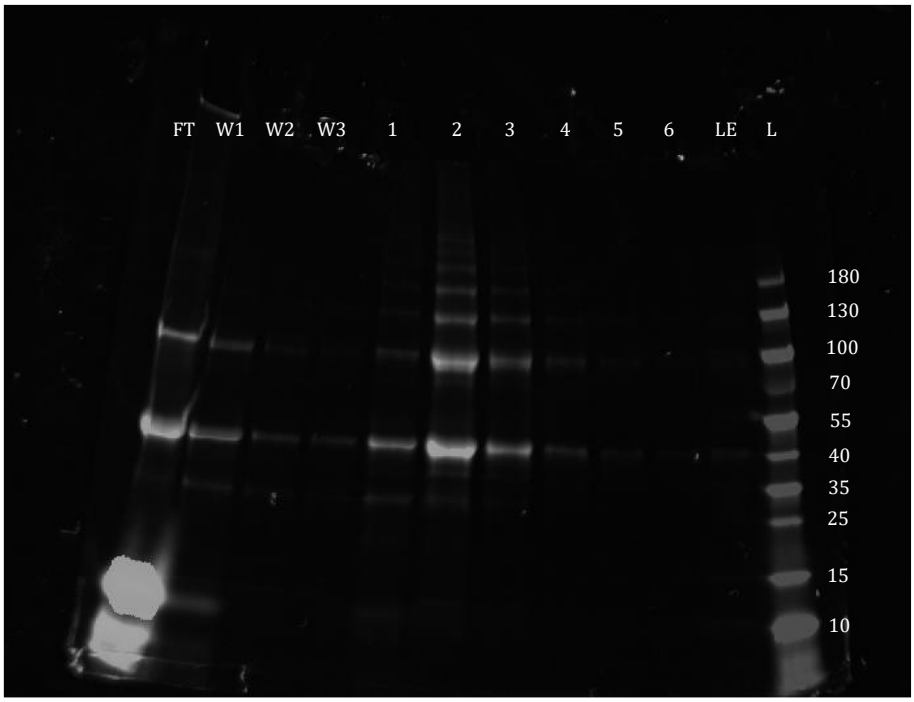

Figure 8. SDS-PAGE after separation of labelled LacI and unreacted dye. When labelling with bifunctional rhodamine B the dye ideally should bind Cys36 and Cys43 which is represented by the band around 40 kDa, however since the dye has two binding sites there is a possibility of the dye binding two monomers which is manifested by the band between 70 and 100 kDa. Abbreviations: FT = flow through, W1 = Wash 1, W2 = Wash 2, W3 = Wash 3, LE = Large elution and L = Ladder (PageRuler Prestained Protein Ladder, Thermo Fisher). This image of the gel was created by merging two images captured on a ChemiDoc Imaging System (Bio-Rad). From FT–LE the bands were visualized using the rhodamine function in the system with an exposure time of 1 s and the ladder was imaged with the inbuilt Cy5 function. The exposure time was 6 s. The two images were merged in ImageJ **(**Schindelin *et al.* 2012**)** and the brightness and contrast of the ladder was adjusted using the auto function. The image was then converted to grayscale in MATLAB R2018b (MathWorks).

4.3 A suitable buffer for electroporation was found through buffer screening

4.3.1 Manual Screening

Electropotation as a technique to introduce proteins or nucleic acids into bacteria requires low salt concentration as the viability of cells is vastly reduced when arcing occurs, which is the consequence of using high salt concentration. A balance has to be struck however since proteins generally require conditions with specific ionic strengths to maintain their function (Sustarsic *et al.* 2014). Finding this balance can either come from previous work, where the POI had been handled in low salt buffers or by finding one or more suitable conditions on your own. Perhaps the simplest way to find suitable conditions is to mix buffers and then tester whether the POI remains soluble in the test conditions. In this study 19 manually mixed buffers were tested with LacI (see Table 1); some buffers were tested with unlabelled LacI and some were tested with labelled LacI. How they were tested can be seen in Section 3.4.1. The results from the testing can be found in Table 3. In many of the conditions LacI precipitated. The amount of precipitate varied, however. Regardless of the amount of precipitate the buffer was deemed unsuitable if precipitate was observed.

Table 3. Screening of manually mixed buffers. Solubility refers to whether LacI precipitated or remained soluble in the buffer. Electroporation compatibility means whether a buffer is compatible or not with the electroporation protocol in Section 3.7. However, electroporation with all buffers was never tested, thus there is no specific breakpoint for whether electroporation would work or not. Instead, tests by Marklund E (unpublished) and data from Sustarsic et al. (2014) act as reference points whether electroporation is compatible with a buffer. Label indicates whether the testing process was done with labelled or unlabelled LacI.

Buffer	Solubility	Electroporation compatible	Label
1	Not soluble	Likely	Unlabelled
2	Soluble	Not likely	Unlabelled
3	Soluble	Not likely	Unlabelled
4	Not soluble	Not likely	Unlabelled
5	Not soluble	Not likely	Unlabelled
6	Not soluble	Likely	Unlabelled
7	Not soluble	Not likely	Unlabelled
8	Not soluble	Not likely	Unlabelled
9	Not soluble	Likely	Unlabelled
10	Not soluble	Likely	Unlabelled
11	Soluble	Not likely	Unlabelled
12	Not soluble	Not likely	Unlabelled
13	Not soluble	Not likely	Unlabelled
14	Not soluble	Likely	Unlabelled
15	Soluble	Not likely	Labelled
16	Soluble	Not likely	Labelled
17	Soluble	Not likely	Labelled
18	Soluble	Maybe	Labelled
19	Soluble	Yes	Labelled
20	Not soluble	Likely	Labelled

4.3.2 Thermal Shift Assay

To try to find optimal conditions with much higher throughput than manually testing buffers a TSA was performed. In the assay 91 different buffers were tested. The buffers used came from a prepared kit from Jena Bioscience. The goal of the assay is to find out the melting temperature of the POI in a specific condition. The melting temperature is determined by measuring the fluorescence of the dye SYPRO Orange which only fluoresces when it interacts with hydrophobic residues. Generally, protein cores are hydrophobic while the surfaces are hydrophilic. When proteins denature from the core will be exposed which should lead to higher fluorescence. This means that ideally the fluorescence should start off low and then increase as the temperature reaches the melting temperature, however if the protein is already denatured or has a relatively hydrophobic surface the assay will not work. The melting temperature is used as an indication of stability in different conditions, since if the conditions are favourable the protein should be more stable and therefore not melt as early compared to unfavourable conditions. By observing the melting temperatures it is possible to pick out conditions that are well suited for the POI (Reinhard *et al.* 2013).

The TSA was run with ~13 µM of unlabelled LacI. Figure 9A shows that the fluorescence is for most, if not all conditions at its peak when the assay is started which means that it is not possible to determine a melting temperature or guess which condition is the most suitable one (Reinhard *et al.* 2013). As a control of the assay 1 mg/ml of lysozyme was tested on the same buffers. Here it was possible to determine a melting temperature, but the intention was not to find the most suitable buffer for lysozyme, therefore no analysis of the data was conducted.

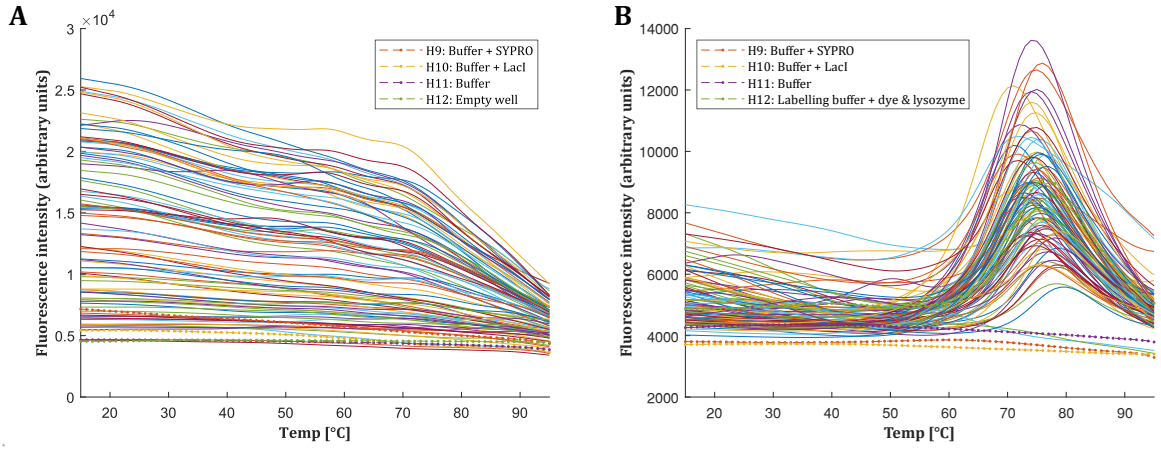

Figure 9. Thermal shift assay of LacI and lysozyme. A) TSA was performed with a dimeric LacI mutant. The controls are highlighted are highlighted by dash-dot lines and in the inset. The fluorescence is high starts off at a high value and generally decreases over time indicating that the assay did not work. B) Lysozyme was run in the assay as well to make sure that there was nothing wrong with the assay itself. Here the fluorescence starts off relatively low and then increases as it should according to figure 1 in the paper by Reinhard *et al.* (2013). Again, the controls for the assay are highlighted with dah-dot lines and in the inset.

4.4 The dimeric LacI mutant could interact with the O_1 operator *in vitro*

An EMSA was performed to verify that LacI was active before it was introduced into *E. coli*. In the assay both labelled protein with and without a his-tag and unlabelled protein was tested (see Figure 10) and the protein was mixed with Cy5-labelled DNA containing the sequence for the O_1 operator. Both LacI with the his-tag cleaved off and the unlabelled protein was titrated, from which it would have been possible to determine the dissociation constant K_D. However, as can be seen in Figure 10 the titration did not work out as intended. An example of how an EMSA should look can be seen in figure 1 of Hellman & Fried (2007). In the gel containing LacI after his-tag removal (see Figure 10A, C & E) neither the protein and/or the DNA ever entered the gel which is especially evident from looking at the bands stemming from rhodamine (see Figure 10C). From these bands it should have been possible based off the concentration of dye and therefore protein to see a shift (compare to the shift observed in Figure 10B, D, F for 500 nM dye for labelled protein). This means that the assay did not work in this case. Figure 10B, D, F show more promising bands however, where shifts can be observed for both labelled and unlabelled protein. This implies that the LacI mutant can bind the O_1 operator *in vitro* and should therefore be able to bind operators *in vivo*.

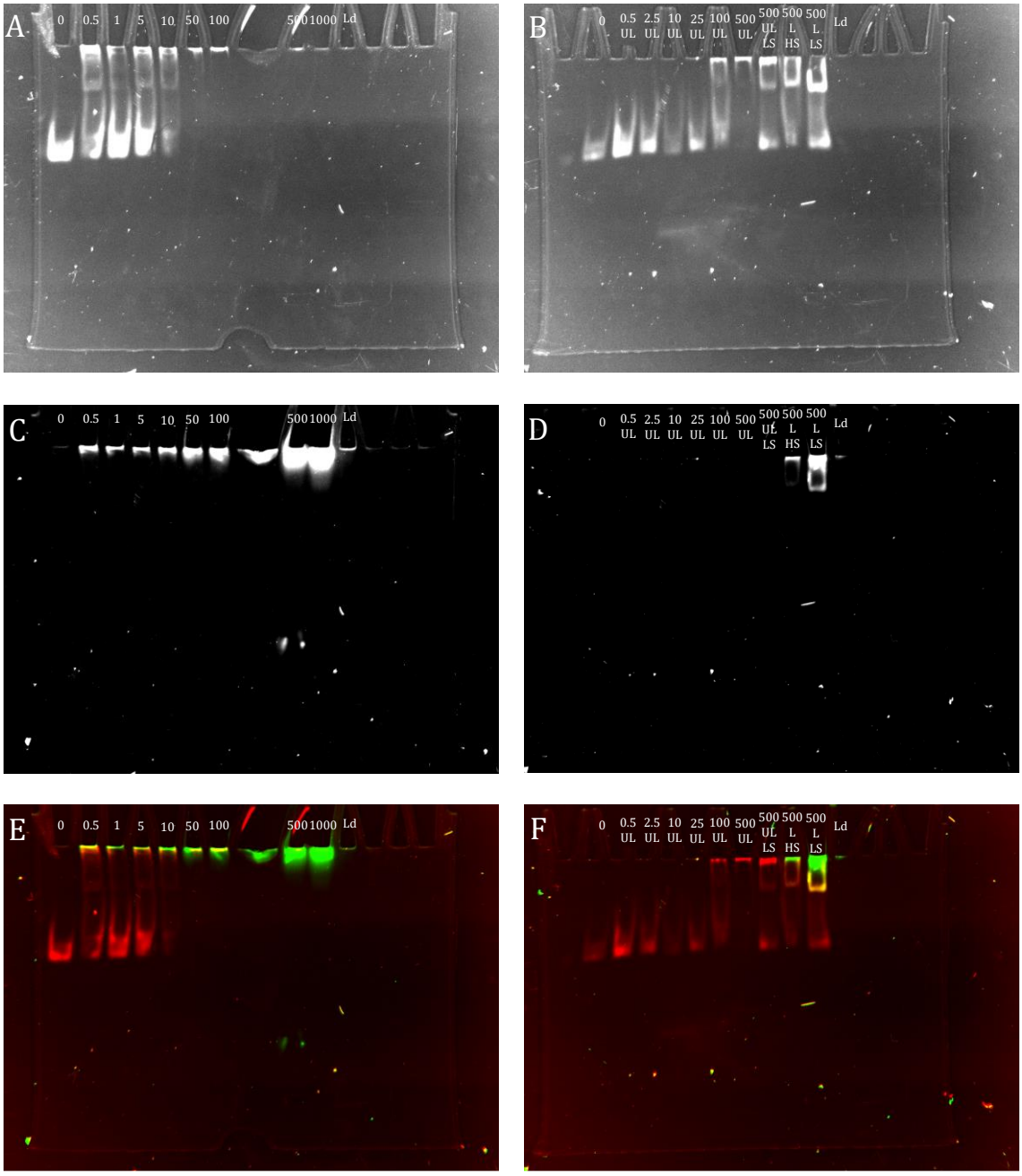

Figure 10. EMSA of unlabelled and labelled LacI with and without a his-tag. A, C & E) Gel with labelled LacI that had its his-tag cleaved off by TEV protease. In this gel a 0–1000 nM titration of dye was performed with a constant DNA concentration of 10 nM. Samples were diluted in electroporation buffer (for recipe see Section 3.3). B, D & F) Gel with unlabelled and labelled LacI that still has the his-tag. Here a 0–500 nM titration of unlabelled protein was performed. At 500 nM of unlabelled protein separate wells with dilutions in low salt buffer (electroporation buffer) and high salt buffer (100 mM HEPES, 250 nM NaCl, 10% glycerol, pH 7.4) were performed. The same was done with the two wells containing labelled LacI. A & B) Gels image with the Cy5 setting on a ChemiDoc Imaging System (Bio-Rad) with 30 s exposure time. C & D) Gels imaged with the rhodamine setting on a ChemiDoc Imaging System (Bio-Rad) with 30 s exposure time. E & F) A merge of the respective gels performed in ImageJ (Schindelin *et al.* 2012) with Cy5 in the red channel and rhodamine in the green channel. Abbreviations: UL = unlabelled, L = labelled, LS = low salt (buffer), HS = high salt (buffer) and Ld = ladder, which was GeneRuler 100 bp DNA Ladder (ThermoFisher). The reason for the ladder not being visible is most likely due to the fact that it diffused out of the gel overnight. The ladder was visible by eye right after the gels were run.

4.5 Tracking of labelled LacI seem to have worked

To perform tracking of LacI labelled with bifunctional rhodamine B two strains of *E. coli* (DH5α and DH5α containing the pSMART plasmid with an array of O_1 operators) were electroporated and dispensed on an agarose pad containing SYTOX Blue. A negative control was performed where labelled LacI was mixed with DH5α cells, but the cells were not electroporated. Image analysis was performed on DH5α cells electroporated with LacI, the image analysis pipeline was not fully completed. Image analysis was not performed for the other two experiments.

4.5.1 Fluorophores could be detected in DH5α

Before performing any image analysis, the movies of all 50 positions acquired were looked at by eye. In Figure 11 bright-field, phase contrast and fluorescence images are showcased for one of the 50 positions taken on an agarose pad where electroporated DH5α cells were suspended. As can be seen in Figure 11C there are fluorophores in all four frames. In the movie containing 1000 frames fluorophores could be spotted in more than four frames for this position.

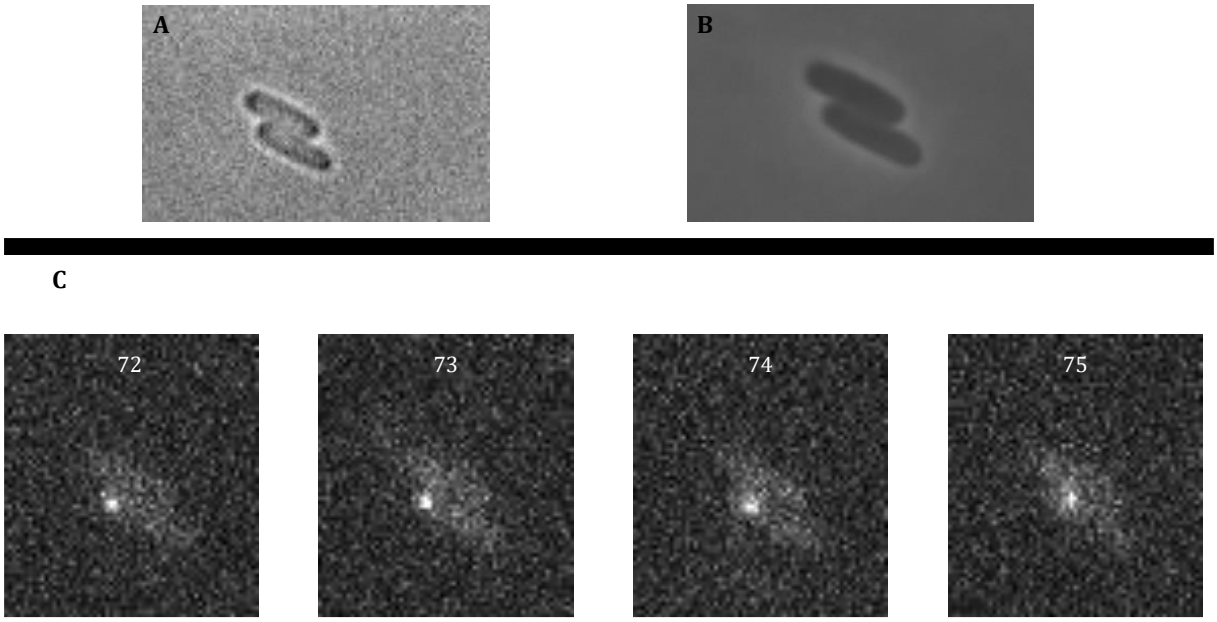

Figure 11. Raw images from DH5α cells electroporated with LacI labelled with bifunctional rhodamine B. A) Bright-field image of cells. B) Phase-contrast image of cells. C) Frames 72–75 from a fluorescence movie containing 1000 frames. Here fluorophores are found in all four frames.

Since fluorophores could be spotted in more than movie from this agarose pad image analysis was performed. Cells were segmented, fluorescent dots were detected and trajectories were built. In Figure 12 a histogram generated by the image analysis software showcases the number of trajectories (N = 81) that the software could detect with the given parameters for image analysis. The histogram displays the number of trajectories of a particular length. Most trajectories were somewhere between 5 and 10 frames long. The software discarded trajectories that were shorter than 5 frames long. The four frames showcased in Figure 11C

were likely a part of the trajectory that was 89 frames long. The image analysis software was never rerun with a different set of parameters to check whether the results were different.

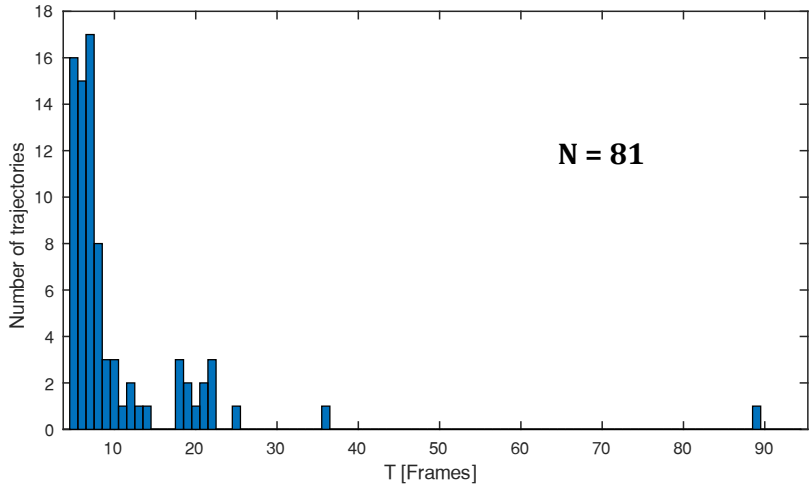

Figure 12. Histogram showcasing the distribution of trajectories of a particular length using initial tracking parameters. The image analysis pipeline was able to detect 81 trajectories from the 50 positions taken on the electroporated LacI agarose pad. Trajectories shorter than 5 frames were discarded by the software.

4.5.2 Fluorophore-like dots could be detected in the negative control

Although no image analysis was performed on the negative control it is still crucial to compare it with the experimental data. The comparison could therefore only be done by eye. Again, 50 positions were used and in Figure 13 bright-field, phase contrast and fluorescence images for one position is shown. The frames shown in Figure 13C are not the only frames that contain fluorophore-like dots. Connecting the dots by eye is more difficult for these four frames compared to the frames displayed in Figure 11C.

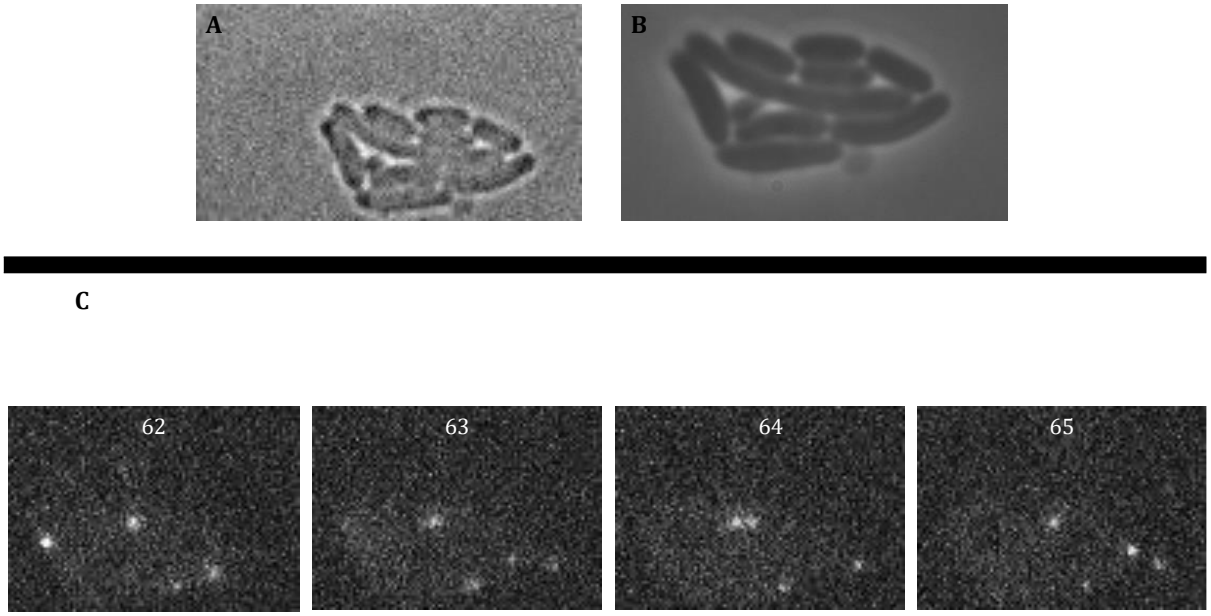

Figure 13. Raw images taken from one position on an agarose pad where DH5α cells were mixed with labelled LacI, but not electroporated. A) Bright-field image of a microcolony of cells. B) Phase contrast image of the same microcolony shown in A). C) Frames 62–65 from a fluorescence movie (1000 frames long). In all of the frames fluorophores seem to be present. Whether the same fluorophores appear in all frames is more difficult to tell compared to Figure 11.

5 Discussion

5.1 The LacI mutant can be purified and labelled with bifunctional rhodamine B

As LacI has been one of the most studied transcription factors it comes to no surprise that it has been successfully purified before (Gilbert & Müller-Hill 1966, Riggs *et al.* 1970, Barkley 1981, Lewis *et al.* 1996, Gatti-Lafranconi *et al.* 2013, Stetz *et al.* 2016, Kipper *et al.* 2018, Marklund E *et al.* 2018). However, in most cases slight variants are purified depending on the type of question researchers want to answer. Purifying a new variant might not be the greatest of scientific achievements, but it is usually worth of notice, especially if the downstream use of the protein is interesting. The mutant purified in this study is a dimeric LacI where nonessential cysteines were removed, two new cysteines (Cys36 and Cys43) were introduced, the C-terminal contains a his-tag and right before the his-tag a TEV recognition sequence (ENLYFQS) can be found, meaning that the his-tag can be cleaved off using TEV protease. The mutant is a dimer is because the his-tag is added. The his-tag is added to make purification easier. The mutant is very similar to one of the mutants used by Marklund E *et al.* (2018), the difference being the TEV recognition sequence, which I added while doing research training at the Elf lab (Autumn 2018). Figure 7 shows successful purification of the mutant by affinity purification. By looking at the pure fractions there seem to be a lot of contamination, but when the fractions were diluted those bands are not as prevalent. That does

not exclude the fact that it might have been better to purify the protein more before continuing in the protocol. One counterargument could be that there are purification steps downstream, however the main protocol only includes affinity-based purification which leaves the risk of continuously copurifying natural *E. coli* protein that have affinity for nickel or cobalt.

The trickier part after purification is to label LacI with bifunctional rhodamine B. There are many reasons for it being a tricky labelling reaction. First of all, the dye has to bind to two cysteines on the same protein. This sounds rather simple, however it might be the case that the dye binds two proteins at the same time. This is evident from looking at Figure 8 where the upper band around somewhere between 70 and 100 kDa must be this side product since the SDS-PAGE is denaturing. The heating process and SDS normally should lead to only monomers, but the covalent bonds that the dye forms will not be broken under these conditions, thus the dye will act as a linker between two monomers making it look like there is a dimer of LacI on the gel. This side product is not desired; however, it should be possible to separate the two products by gel filtration. This is something that should have been tested. It is also possible that only one cysteine binds the dye before the reaction is quenched. This side product might not be as likely, and it is also much harder to detect. Another important aspect are the conditions of the labelling reaction. Thiol-based labelling has narrow pH range (6.5–7.5) where the labelling reaction is more optimal above pH 7.0 (Nanda & Lorsch 2014). The pH range itself might not be problematic as physiological pH is around that range, but it is the fact that the LacI mutant has an estimated isoelectric point of 6.49 (determined by the ProtParam tool on ExPASy, Gasteiger *et al.* 2003). It could be that LacI is not active close to its the isoelectric point, therefore it might be ill-advised to change the pH from 7.4 that has proven to work well enough (Marklund E *et al.* 2018) to another pH. It would however be worth experimenting with pH in the future. Changing the pH might also be suitable since the labelling efficiency was low throughout the many attempts at labelling. It could however be that the low labelling efficiency stems from the fact that the dye has to bind two cysteines which seem to be relatively rare based on this study. Another aspect for the low labelling efficiency could be the fact that the dye was not added sequentially as in Forkey *et al.* (2003) & Beausang *et al.* (2012). The authors never state how efficient their labelling is, which is unfortunate since it would have been helpful to compare their labelling efficiency with mine even though they labelled another protein. Nonetheless it is possible to label LacI with bifunctional rhodamine B (see Figure 8).

5.2 The difficulty of finding electroporation conditions for LacI

From looking at Table 3 and the attempt at doing more large-scale screening in Figure 9 it becomes rather evident that finding suitable buffer conditions for electroporation is not that easy. Although it is possible to make guesses based on general knowledge about proteins and their environment it in the end comes down to trial and error to find a condition where the protein can remain soluble in a condition that has low enough salt content for electroporation to work or causes as little harm as possible to the cells. The difficult thing is that there are

many parameters that can be tweaked: salt concentration, type of salt, buffering agent, concentration of buffering agent, crowding agent and its concentration, pH and temperature (Kramer *et al.* 2012). In my case I kept some parameters fixed (pH and temperature) and it still was difficult to find something that ended up working. Of course, it could have been the case that changing these parameters could have made things easier, but it is not a guarantee.

Since there are so many parameters to consider it is ideal to be able to test conditions in a high-throughput manner rather than manually preparing conditions. That is why I was advised to do the TSA, but since the assay did not work no information could be gathered from it (see Figure 9A). The first conclusion from the results was that LacI was not compatible with TSA, for instance due to an abundance of hydrophobic patches. This should not be the problem however since Gatti-Lafranconi *et al.* (2013) has managed to perform the assay. What they did was to perform the assay with wildtype LacI together with either inducer (IPTG) or operator DNA. These are not the same conditions nor the same protein (due to the number of mutations the mutant contains) as in this study, but the important thing is that they managed to perform the assay. This makes me believe that the assay should be able to work with the mutant as well. As to why it did not work could have to do with the batch of protein that was used for the assay. The batch that was used was not the same one as was used in the EMSA and electroporation, thus its activity was never tested meaning that the protein could have been denatured before the assay was run.

5.3 EMSA with LacI and O_1 operator suggests that LacI is active

To be able to have confidence in single-molecule tracking data it is of utmost importance to know that the POI is able to perform its function within the cell. There could be a case where the protein cannot bind to a key interaction partner. This would have to be verified before performing single-molecule experiments as it could lead to misinterpretations of the experimental results if the POI is inactive. As an example, assume that a protein that normally binds the membrane is studied but when electroporated into the cell it is not active and thus only diffuses around in the cytoplasm. One would expect for there to be a diffusional state where the diffusion is slow when the protein is membrane bound and that there is a faster state when the protein is searching for the membrane. With impaired function, most, if not all protein would reside in one state. This example is relatively simple and if the function of the protein is well-known it would be relatively easy to conclude that the protein is non-functional. However, if the biological states are not known one would not be able to conclude that the protein is non-functional from observing the data.

EMSA is an assay that can be used to verify a DNA-binding protein's activity before the protein is electroporated into cells. The activity of LacI was verified using an EMSA where LacI was mixed with Cy5-labelled DNA containing the O_1 operator (see Figure 10). The problem with this particular attempt at the assay was the fact that the concentration of TEB in the gel and running buffer was not the same. The gel only had 0.5x TEB while the running buffer was 1x, this subsequently led to higher concentrations of the other components in the

gel since too little TEB was added. This is a very likely reason for the bands being very smeary and that the bands moved slower than expected. It is also evident that in many of the gels both LacI and/or DNA did not enter the gel properly. Whether this was because of the mismatch between running buffer and the gel or poor loading is difficult to determine; perhaps it could be a combination of both. The concentration of LacI is also important. The concentrations in the titrations were solely based on the concentration of bifunctional rhodamine B. This was not accounted for when making the titration of unlabelled protein, thus the concentration given for unlabelled LacI in Figure 10B, D & F is not the actual protein concentration. It is however possible to see that the labelled LacI with a his-tag is active when running at low salt conditions (most evident from Figure 10F). The ladder is also missing from the gels which most likely stems from the gels being stored in dH_2O overnight before imaging the gel. The results from this EMSA are not super clear. Ideally it should be redone without any of the mistakes to better indicate the activity of LacI with O_1 operator DNA.

5.4 The assay seems to work but rigorous testing is needed

After looking at the raw microscope data with and without electroporation (see Figure 11 and Figure 13 for a couple of frames) it seems as if there is some fluorescence on both pads. It is therefore difficult to judge whether what is observed with the electroporated sample is an artefact or that actual fluorophores are present. It is of course possible to argue that what could be seen in the negative control is noise and that the dots seen for the electroporated sample are real fluorophores. Only looking at a couple of frames as in Figure 11 and Figure 13 does not paint a fair picture. Observing the movies is fairer but differentiating noise from signal is not possible to do by eye. The only way to truly compare the two experiments is through image analysis, but it was not performed for the negative control given the timeframe of the project. This is unfortunate since a more honest comparison cannot be made, however the information acquired from the image analysis for the electroporated sample can still be used. What Figure 12 shows is that most of the trajectories are very short and that the longest trajectory found by the image analysis software is an outlier. The frames showcased in Figure 11 are likely a part of those 89 frames that the image analysis pipeline was able to build. The question then becomes whether the short trajectories are all noise while the longer trajectories are fluorophores? Again, this is a difficult question answer without a similar histogram for the negative control. If there was a histogram for the negative control the number of trajectories and their distribution it could be used as a way to possibly classify some, all or none of the trajectories shown in Figure 12 as noise.

One important thing to keep in mind when trying to interpret Figure 12 is that the parameters used to perform image analysis were never changed or optimised. This is imperative, since these parameters govern whether the area of the cells are correct, what constitutes a fluorescent dot in respect to the background and what cells should be discarded based on their SYTOX Blue fluorescence. The parameters are also based on the optical setup, which in this case was tricky to find out since the camera and lasers were moved to another optical setup

after all of the microscopy was performed, but before image analysis was conducted. Some parameters were therefore chosen based on approximations and not on what would have been possible to measure if the optical setup had remained the same. With non-optimised parameters it could be that the algorithms in the software over- or underestimate the number of fluorescent dots and/or the lengths of the trajectories. In the case of overestimation, the most likely scenario is that the algorithms interpret autofluorescence stemming from the cells in the beginning of the movies as fluorescent dots that can be tracked. This autofluorescence is inescapable when bifunctional rhodamine B is used as a dye as it emits light around 547 nm which is in the green part of the spectra where autofluorescence is prevalent in *E. coli* (Mihalcescu *et al.* 2015). This would not necessarily lead to longer trajectories, but more trajectories. If the algorithms instead underestimate trajectory abundance and lengths the thresholds for finding a dot are likely set too high, meaning that many dots would be missed. In the case of the parameters used in this study it is likely that both over- and underestimation occurred, based on comparing the trajectory length of the outlier in Figure 12 with raw data from the cells shown in Figure 11. How I interpreted the raw data the trajectory was longer than 89 frames, however at the same time dots were detected where they were not visible by eye.

One thing to note however is that the data shown in Figure 11 were both outliers Figure 13. Although Figure 12 was not generated with optimised parameters it can at least showcase that the position in Figure 11 was not similar to the other trajectories. Figure 13 was an outlier in the sense that it, at least by eye had more fluorescence than most of the other negative control movies. What can be seen from looking at the movies from these two positions is that the trajectories in Figure 13 do not seem to be as long as the one in Figure 11 and some of the trajectories with intermediate length in Figure 12. It is also difficult to judge by eye whether the dots in Figure 13 stay within the boundary of one cell or whether they move over many cells. If they do move between the cells the dots could be regarded as noise, since a fluorophore is confined to the cell volume if electroporation worked as intended.

What can be concluded then? If one is slightly optimistic it is possible to say that the assay worked since the dots for the negative control in Figure 13 were not present for the same period of time as Figure 11 and that there seemed to be at least some longer-lived trajectories in Figure 12. This claim is not supported by much data; thus, it is necessary to further analyse the data and perform more experiments before any sound conclusion can be drawn.

5.5 Future development and conclusions

Based on the results acquired from the microscopy data it is evident that more work has to be done before any fully supported conclusions can be drawn. What would have to be done is to first try to tweak the tracking parameters as they are vital for determining what is a dot and what is noise in each frame. After parameter tweaking the next step would be to rerun and finish the image analysis on the data for the electroporated sample. Once that is done analysis would be done on the negative control to be able to compare it to experimental system. These

three actions are more short-term developments that should not require too much time. However, if there was more time available all the experiments should be replicated from scratch. Acquiring new protein that has not been stored for a long period of time would not hurt and it would perhaps be possible to increase the labelling efficiency and it would be worth trying to separate the biproducts from the labelling reaction from the desired product. It would also be ideal to remove the his-tag before performing microscopy, which was not done in the experiments shown in Section **Fel! Hittar inte referenskälla.**. During this it would be necessary to verify the activity of LacI again, an ample opportunity to get a nicer EMSA to display other than just verifying the activity. Once all of the steps are completed electroporation and microscopy would be done with the same experimental basis, but more controls would be added in and other types of experiments would be added as well, such as inducing the cells with IPTG while on the pad. Ideally the experiments would be run at 37 °C which would make it possible to perform imaging and then return to the same positions after a few hours to verify that the cells actually grew and divided. The last step would be to analyse the data and based on the results hopefully proceed to the confocal setup to measure the unspecific binding time of LacI to DNA *in vivo*.

Overall, what are the key points that can be taken from this study? The first one is that it is possible to overexpress and purify this LacI mutant. The second point is that LacI can be labelled with bifunctional rhodamine B, although not with very good labelling efficiency. The third point is that there are conditions that can be used to electroporate LacI with. The penultimate point is that with a hint of optimism LacI could be tracked *in vivo*. Lastly, data analysis has to optimised to get better and more data out of the experiments.

6 Acknowledgements

I would like to thank my supervisor Emil Marklund for all the help, advice and support that you have given me during the project. It was very nice to hear words of encouragement even when things did not go that well.

I also want to thank my co-supervisor Johan Elf who acted as a co-supervisor for giving me the opportunity to work in your research group.

I also want to thank my subject reader Sebastian Deindl for the discussions, feedback and advice on trying to find suitable buffer conditions.

I also want to thank Javier Aguirre Rivera for his help with buffer conditions, electroporation and microscopy. It was very helpful to learn how to perform all of these tasks much more efficiently.

I would also like to thank Laura Lehmann for showing me the thermal shift assay and helping me with interpreting the results of the assay.

I also want to thank Carolin Seefeldt for the advice on buffer compositions, protein purification and microscopy.

I also want to thank Ivan Volkov for showing and helping me with the image analysis.

I also want to thank David Fange for showing me how the microscope worked and assisting me when problems arose with the microscope.

I would also like to thank the rest of the people in the Elf and Johansson research groups for making me feel very welcome and for letting me go with you to the *Cell Biology of Prokaryotes* conference. I had the opportunity to learn a lot from that, so thank you.

Lastly, I would like to extend my thanks to the other people, such as friends and family who have supported me during this project.

References

Barkley MD. 1981. Salt dependence of the kinetics of the lac repressor-operator interaction: role of nonoperator deoxyribonucleic acid (DNA) in the association reaction. Biochemistry 20: 3833–3842.

Beausang JF, Sun Y, Quinlan ME, Forkey JN, Goldman YE. 2012. Fluorescent Labeling of Calmodulin with Bifunctional Rhodamine. Cold Spring Harbor Protocols 2012: pdb.prot069351.

Bell CE, Lewis M. 2000. A closer view of the conformation of the Lac repressor bound to operator. Nature Structural Biology 7: 209–214.

Berg OG, Winter RB, Von Hippel PH. 1981. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. Biochemistry 20: 6929–6948.

Combs CA. 2010. Fluorescence Microscopy: A Concise Guide to Current Imaging Methods. Current protocols in neuroscience / editorial board, Jacqueline N Crawley . [et al] 0 2: Unit2.1.

Corrie JET, Craik JS, Munasinghe VRN. 1998. A Homobifunctional Rhodamine for Labeling Proteins with Defined Orientations of a Fluorophore. Bioconjugate Chemistry 9: 160–167.

Davis KM, Isberg RR. 2016. Defining heterogeneity within bacterial populations via single cell approaches. BioEssays 38: 782–790.

Deochand DK, Grove A. 2017. MarR family transcription factors: dynamic variations on a common scaffold. Critical Reviews in Biochemistry and Molecular Biology

Edelstein A, Amodaj N, Hoover K, Vale R, Stuurman N. 2010. Computer Control of Microscopes Using μManager. Current Protocols in Molecular Biology 92: 14.20.1-14.20.17.

Ernst D, Köhler J. 2012. Measuring a diffusion coefficient by single-particle tracking: statistical analysis of experimental mean squared displacement curves. Physical Chemistry Chemical Physics 15: 845–849.

Forkey JN, Quinlan ME, Shaw MA, Corrie JET, Goldman YE. 2003. Three-dimensional structural dynamics of myosin V by single-molecule fluorescence polarization. Nature 422: 399.

Fried M, Crothers DM. 1981. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Research 9: 6505–6525.

Furini S, Domene C. 2014. DNA Recognition Process of the Lactose Repressor Protein Studied via Metadynamics and Umbrella Sampling Simulations. The Journal of Physical Chemistry B 118: 13059–13065.

Gandibleux X, Morita H, Katoh N. 2017. A population-based algorithm for solving linear assignment problems with two objectives. Computers & Operations Research 79: 291–303.

Garner MM, Revzin A. 1981. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. Nucleic Acids Research 9: 3047–3060.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Research 31: 3784–3788.

Gatti-Lafranconi P, Dijkman WP, Devenish SR, Hollfelder F. 2013. A single mutation in the core domain of the lac repressor reduces leakiness. Microbial Cell Factories 12: 67.

Ghahramani Z. 2001. An introduction to hidden markov models and bayesian networks. International Journal of Pattern Recognition and Artificial Intelligence 15: 9–42.

Gilbert W, Müller-Hill B. 1966. ISOLATION OF THE LAC REPRESSOR. Proceedings of the National Academy of Sciences of the United States of America 56: 1891–1898.

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.

Hellman LM, Fried MG. 2007. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. Nature Protocols 2: 1849–1861.

Jacob F, Monod J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. Journal of Molecular Biology 3: 318–356.

Jaqaman K, Loerke D, Mettlen M, Kuwata H, Grinstein S, Schmid SL, Danuser G. 2008. Robust single-particle tracking in live-cell time-lapse sequences. Nature Methods 5: 695–702.

Kipper K, Eremina N, Marklund E, Tubasum S, Mao G, Lehmann LC, Elf J, Deindl S. 2018. Structure-guided approach to site-specific fluorophore labeling of the lac repressor LacI. PLOS ONE 13: e0198416.

Kleina LG, Miller JH. 1990. Genetic studies of the lac repressor: XIII. Extensive amino acid replacements generated by the use of natural and synthetic nonsense suppressors. Journal of Molecular Biology 212: 295–318.

Kramer RM, Shende VR, Motl N, Pace CN, Scholtz JM. 2012. Toward a Molecular Understanding of Protein Solubility: Increased Negative Surface Charge Correlates with Increased Solubility. Biophysical Journal 102: 1907–1915.

Latchman DS. 1997. Transcription factors: An overview. The International Journal of Biochemistry & Cell Biology 29: 1305–1312.

Lewis M. 2005. The lac repressor. Comptes Rendus Biologies 328: 521–548.

Lewis M, Chang G, Horton NC, Kercher MA, Pace HC, Schumacher MA, Brennan RG, Lu P. 1996. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. Science (New York, NY) 271: 1247–1254.

Lindén M, Ćurić V, Amselem E, Elf J. 2017. Pointwise error estimates in localization microscopy. Nature Communications 8: 15115.

Loy G, Zelinsky A. 2003. Fast Radial Symmetry for Detecting Points of Interest. IEEE Trans Pattern Anal Mach Intell 25: 959–973.

Mahmutovic A, Berg OG, Elf J. 2015. What matters for lac repressor search in vivo--sliding, hopping, intersegment transfer, crowding on DNA or recognition? Nucleic Acids Research 43: 3454–3464.

Markiewicz P, Kleina LG, Cruz C, Ehret S, Miller JH. 1994. Genetic Studies of the lac Repressor. XIV. Analysis of 4000 Altered Escherichia coli lac Repressors Reveals Essential and Non-essential Residues, as well as 'Spacers' which do not Require a Specific Sequence. Journal of Molecular Biology 240: 421–433.

Marklund E, Amselem E, Kipper K, Zheng X, Johansson M, Deindl S, Elf J. 2018. Direct observation of rotation-coupled protein diffusion along DNA on the microsecond timescale. bioRxiv 401414.

Marklund EG, Mahmutovic A, Berg OG, Hammar P, van der Spoel D, Fange D, Elf J. 2013. Transcription-factor binding and sliding on DNA studied using micro- and macroscopic models. Proceedings of the National Academy of Sciences of the United States of America 110: 19796–19801.

Mihalcescu I, Gateau MV-M, Chelli B, Pinel C, Ravanat J-L. 2015. Green autofluorescence, a double edged monitoring tool for bacterial growth and activity in micro-plates. Physical Biology 12: 066016.

Nanda JS, Lorsch JR. 2014. Chapter Seven - Labeling of a Protein with Fluorophores Using Maleimide Derivitization. In: Lorsch J (ed.). Methods in Enzymology, pp. 79–86. Academic Press,

Pace HC, Kercher MA, Lu P, Markiewicz P, Miller JH, Chang G, Lewis M. 1997. Lac repressor genetic map in real space. Trends in Biochemical Sciences 22: 334–339.

Posada D, Buckley TR. 2004. Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests. Systematic Biology 53: 793–808.

Ramos JL, Martínez-Bueno M, Molina-Henares AJ, Terán W, Watanabe K, Zhang X, Gallegos MT, Brennan R, Tobes R. 2005. The TetR Family of Transcriptional Repressors. Microbiol Mol Biol Rev 69: 326–356.

Ranefall P, Sadanandan SK, Wählby C. 2016. Fast adaptive local thresholding based on ellipse fit. 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 205–208.

Reinhard L, Mayerhofer H, Geerlof A, Mueller-Dieckmann J, Weiss MS. 2013. Optimization of protein buffer cocktails using Thermofluor. Acta Crystallographica Section F, Structural Biology and Crystallization Communications 69: 209–214.

Ries J, Schwille P. 2012. Fluorescence correlation spectroscopy. BioEssays 34: 361–368.

Riggs AD, Bourgeois S, Cohn M. 1970. The lac represser-operator interaction: III. Kinetic studies. Journal of Molecular Biology 53: 401–417.

Rojo F. 2001. Mechanisms of transcriptional repression. Current Opinion in Microbiology 4: 145–151.

Rosano GL, Ceccarelli EA. 2014. Recombinant protein expression in Escherichia coli: advances and challenges. Frontiers in Microbiology, doi 10.3389/fmicb.2014.00172.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. 2012. Fiji: an open-source platform for biological-image analysis. Nature Methods 9: 676–682.

Sgouralis I, Pressé S. 2017. An Introduction to Infinite HMMs for Single-Molecule Data Analysis. Biophysical Journal 112: 2021–2029.

Shashkova S, Leake MC. 2017. Single-molecule fluorescence microscopy review: shedding new light on old problems. Bioscience Reports, doi 10.1042/BSR20170031.

Stetz MA, Carter MV, Wand AJ. 2016. Optimized expression and purification of biophysical quantities of the Lac repressor and Lac repressor regulatory domain. Protein expression and purification 123: 75–82.

Suckow J, Markiewicz P, Kleina LG, Miller J, Kisters-Woike B, Müller-Hill B. 1996. Genetic Studies of the Lac Repressor XV: 4000 Single Amino Acid Substitutions and Analysis of the Resulting Phenotypes on the Basis of the Protein Structure. Journal of Molecular Biology 261: 509–523.

Sustarsic M, Plochowietz A, Aigrain L, Yuzenkova Y, Zenkin N, Kapanidis A. 2014. Optimized delivery of fluorescently labeled proteins in live bacteria using electroporation. Histochemistry and Cell Biology 142: 113–124.

Tempestini A, Monico C, Gardini L, Vanzi F, Pavone FS, Capitanio M. 2018. Sliding of a single lac repressor protein along DNA is tuned by DNA sequence and molecular switching. Nucleic Acids Research 46: 5001–5011.

Thorn K. 2016. A quick guide to light microscopy in cell biology. Molecular Biology of the Cell 27: 219–222.

Ventola CL. 2015. The Antibiotic Resistance Crisis. Pharmacy and Therapeutics 40: 277–283.

Volkov IL, Lindén M, Rivera JA, Ieong K-W, Metelev M, Elf J, Johansson M. 2018. tRNA tracking for direct measurements of protein synthesis kinetics in live cells. Nature Chemical Biology 14: 618.

Appendix A: Plasmid sequence

The DNA sequence for pD861.LacIDim2-TEV-His6 is found here:

agctcctgaaaatctcgataactcaaaaaatacgcccggtagtgatcttatttcattatggtgaaagttggaacctcttacgtgccgatcaa
gaagacggtcaaaagcctccggtcggaggcttttgactttctgctatggaggtcaggtatgatttaaatggtcagtattgagcgatatcta
gagaattcgtccaccacaattcagcaaattgtgaacatcatcacgttcatctttccctggttgccaatggcccatttttcctgtcagtaacgag
aaggtcgcgaattcaggcgctttttagactggtcgtaatgaaattctttttaagaaggagatatacatatgaaaccagtaacgttatacgat
gtcgcagagtatgccggtgtctcttatcagaccgtttcccgcgtggtgaaccaggccagccacgtttctgcgaaaacgcggtgtaaagt
ggaagcggcgatgtgtgagctgaattacattcccaaccgcgtggcacaacaactggcgggcaaacagtcgttgctgattggcgttgcc
acctccagtctggccctgcacgcgccgtcgcaaattgtcgcggcgattaaatctcgcgccgatcaactgggtgccagcgtggtggtgt
cgatggtagaacgaagcggcgtcgaagccgcgaaagcggcggtgcacaatcttctcgcgcaacgcgtcagtgggctgatcattaact
atccgctggatgaccaggatgccattgctgtggaagctgccgcgactaatgttccggcgttatttcttgatgtctctgaccagacacccat
caacagtattatttttctcccatgaagacggtacgcgactgggcgtggagcatctggtcgcattgggtcaccagcaaatcgcgctgttagc
gggcccattaagttctgtctcggcgcgtctgcgtctggctggctggcataaatatctcactcgcaatcaaattcagccgatagcggaacg
ggaaggcgactggagtgccatgtccggttttcaacaaaccatgcaaatgctgaatgagggcatcgttcccactgcgatgctggttgcca
acgatcagatggcgctgggcgcaatgcgcgccattaccgagtccgggctgcgcgttggtgcggatatctcggtagtgggatacgacg
ataccgaagacagctcatgttatatcccgccgttaaccaccatcaaacaggattttcgcctgctggggcaaaccagcgtggaccgcttg
ctgcaactctctcagggccaggcggtgaagggcaatcagctgttgcccgtctcactggtgaaaagaaaaaccaccctggcgcccaat
acgcaaaccggcagcggcgaaaacctgtattttcagagcggcagcggccatcaccatcaccatcacggttagagcggccgccaccg
ctgagcaataactagcataacccccttggggcctctaaacgggtcttgaggggttttttgctgaaaggaggaactatatccgggtaacgaa
ttcaagcttgatatcattcaggacgagcctcagactccagcgtaactggactgcaatcaactcactggctcaccttcacgggtgggccttt
cttcggtagaagtcttctcatgaccaaaatcccttaacgtgagttacgcgcgcgtcgttccactgagcgtcagaccccgtagaaaagatc
aaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggat
caagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttctagtgtagccgtagttagccc
accacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtctt
accgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttgga
gcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacagg
tatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgg
gtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcctttt
acggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgag
ctgataccgctcgcgcagaaaggcccacccgaaggtgagccaggtgattacatttgggccctcattagaaaaactcatcgagcatcaa
atgaaattgcaatttattcatatcaggattatcaataccatatttttgaaaaagccgtttctgtaatgaaggagaaaactcaccgaggcagtt
ccataggatggcaagatcctggtatcggtctgcgattccgactcgtccaacatcaatacaacctattaatttcccctcgtcaaaaataaggt
tatcaagtgagaaatcaccatgagtgacgactgaatccggtgagaatggcaaaagtttatgcatttctttccagacttgttcaacaggcca
gccattacgctcgtcatcaaaatcactcgcatcaaccaaaccgttattcattcgtgattgcgcctgagcgaggcgaaatacgcgatcgct
gttaaaaggacaattacaaacaggaatcgagtgcaaccggcgcaggaacactgccagcgcatcaacaatattttcacctgaatcagga
tattcttctaatacctggaacgctgtttttccggggatcgcagtggtgagtaaccatgcatcatcaggagtacggataaaatgcttgatggt
cggaagtggcataaattccgtcagccagtttagtctgaccatctcatctgtaacatcattggcaacgctacctttgccatgtttcagaaaca
actctggcgcatcgggcttcccatacaagcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatc
agcatccatgttggaatttaatcgcggcctcgacgtttcccgttgaatatggctcat

Appendix B: Amino acid sequences

The amino acids sequences for native LacI and the overexpressed mutant are listed here.

Native amino acid sequence of LacI:

MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNYIPNRVAQQ
LAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVVSMVERSGVEACKAAVH
NLLAQRVSGLIINYPLDDQDAIAVEAACTNVPALFLDVSDQTPINSIIFSHEDGTRLGV
EHLVALGHQQIALLAGPLSSVSARLRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQ
TMQMLNEGIVPTAMLVANDQMALGAMRAITESGLRVGADISVVGYDDTEDSSCYIP
PLTTIKQDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLAPNTQTASPRAL
ADSLMQLARQVSRLESGQ

Amino acid sequence of overexpressed mutant:

MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTRCKVEAAM**C**ELNYIPNRVAQQ
LAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVVSMVERSGVEAAKAAVH
NLLAQRVSGLIINYPLDDQDAIAVEAAATNVPALFLDVSDQTPINSIIFSHEDGTRLGV
EHLVALGHQQIALLAGPLSSVSARLRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQ
TMQMLNEGIVPTAMLVANDQMALGAMRAITESGLRVGADISVVGYDDTEDSSCYIP
PLTTIKQDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLAPNTQTGSGEN
LYFQSGSGHHHHHH

Appendix C: SDS-PAGE gels

Here are images of SDS-PAGE gels where the product did not end up being used for electroporation or the gels were from purification of the flow through shown in Figure 8. Figure 14 and Figure 15 both show the same gels which contain labelled LacI that was treated with TEV protease (Sigma-Aldrich) to remove the his-tag. In Figure 14 Laemlli (Bio-Rad) and InstantBlue (Expedeon) was used to stain the gel. In Figure 15 the fluorescence from the labelled protein is used to detect it and the ladder is detected using the Cy5 setting on a ChemiDoc Imaging System (Bio-Rad).

Both Figure 16 and Figure 17 show gels acquired from purification of the flow through that still contains labelled LacI, which is shown in Figure 8. Figure 16 shows the initial gel filtration that did not manage to separate the free dye from labelled protein very well. The fractions containing protein were however pooled and purified again, this time using affinity chromatography. The repurification is shown in Figure 17 where the free dye could be well separated from labelled LacI.

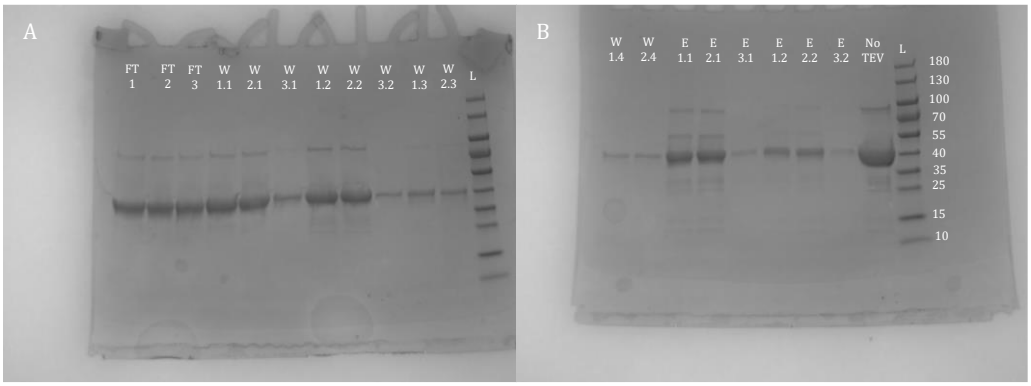

Figure 14. InstantBlue stained SDS-PAGE gel after treatment with TEV protease. A) This gel shows the different flow throughs and some of the washes when performing affinity purification with Mini Bio-Spin Chromatography Colums (Bio-Rad) and HisPur Cobalt Resin (Thermo Fisher). Three of these columns were used and the bands that can be seen in Figure 8 are found on this gel as well indicating that there was cleavage of the his-tag. B) Two washes, some of the elutions and a control is shown. In the elutions there are bands for labelled LacI meaning that cleavage was to 100% efficient. There are also very feint bands around 55 kDa in some of the elutions. These bands are the TEV protease purchased from Sigma-Aldrich, which has a molecular weight of 52 kDa due to it containing a glutathione S-transferase (GST) tag and a his-tag. Naming convention in both gels: Flow through = FT, wash = W, elution = E, ladder = L (PageRuler Prestained Protein Ladder, Thermo Fisher) and No TEV = labelled LacI without the addition of TEV protease. The first number in for every well indicates the three columns that were used while the second number indicates which wash/elution step that is shown.

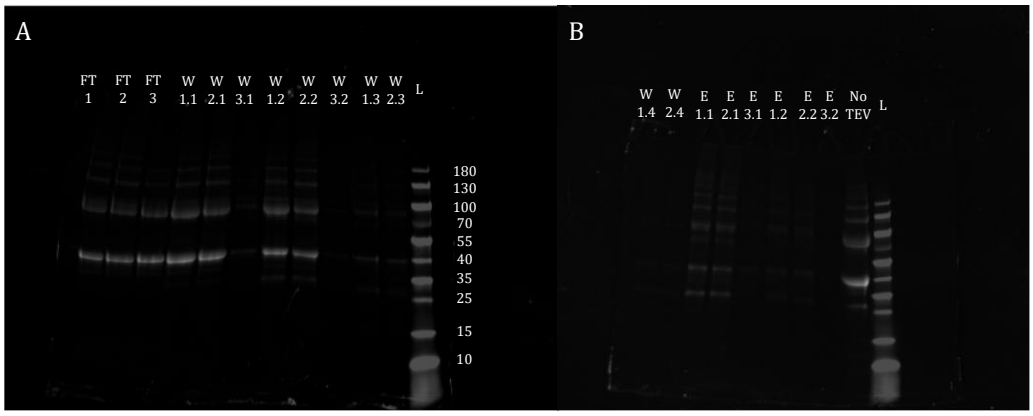

Figure 15. Rhodamine SDS-PAGE after treatment with TEV Protease. The gels shown here are the same gels as in Figure 14 but now the fluorescence of bifunctional rhodamine B is used to visualise the bands. The exposure time was set to 1 s for rhodamine on a ChemiDoc Imaging System (Bio-Rad). A) Flow throughs and washes indicating that much of the labelled protein could be cleaved off. B) Washes, eluate and a no cleavage control. Here it is possible to see some labelled protein that was not cleaved by the protease, although most of the labelled protein seemed to have been cleaved. Naming convention in both gels: Flow through = FT, wash = W, elution = E, ladder = L (PageRuler Prestained Protein Ladder, Thermo Fisher) and No TEV = labelled LacI without the addition of TEV protease. The first number in for every well indicates the three columns that were used while the second number indicates which wash/elution step that is shown. The images of the gels were made by merging the images where the ChemiDoc Imaging System (Bio-Rad) was set to the rhodamine setting and Cy5 setting for the ladder. The raw images for the ladder acquired from this were still difficult to see, therefore the brightness and contrast was adjusted using ImageJ (Schindelin *et al.* 2012); merging was also done in ImageJ. The image was then converted from colourised to grayscale using MATLAB R2018B (MathWorks).

Figure 16. SDS-PAGE after size exclusion chromatography with labelled LacI. The fractions collected here were passed through an illustra NAP-10 (GE Healthcare) column. As can be seen in fraction 3 and 4 most of the free dye is removed, but there is still some in fraction 4. For the other fractions containing labelled LacI there is substantially more dye though, which is indicated by the lower bands. The image was created by merging the image with rhodamine exposure (exposure time 1 s) and Cy5 exposure (exposure time 6 s) for the ladder in ImageJ (Schindelin *et al.* 2012). The brightness and contrast of the ladder was also adjusted. The merged image was then converted into grayscale in MATLAB R2018b (MathWorks). L indicates the ladder, which was PageRuler Prestained Protein Ladder (Thermo Fisher).

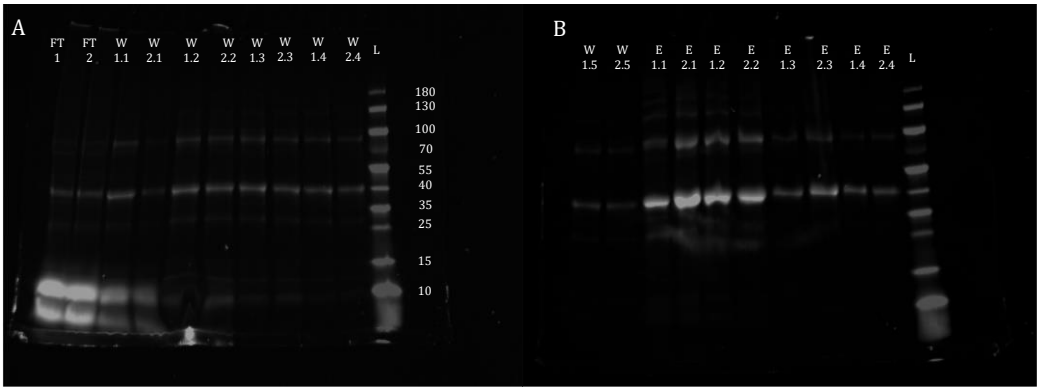

Figure 17. SDS-PAGE after affinity purification of pooled gel filtration fractions. Since the gel purification was not too successful the fractions containing labelled LacI were pooled (fraction 3-8, see Figure 16) and instead purified with Mini Bio-Spin Chromatography Colums (Bio-Rad) and Ni-NTA His-Select resin (Sigma-Aldrich). A) Flow throughs and washing steps. The very bright bands at the bottom of the gel indicates free bifunctional rhodamine B and although there is some labelled protein in all of the fractions it is very little compared to the eluate fractions in B). B) Two wash fractions and eluate fractions. All eluate fractions lack any low molecular weight bands while the bands at the corresponding LacI molecular weights are present. The images for both gels were created by merging the rhodamine and Cy5 image (used to be able to see the ladder) in ImageJ (Schindelin *et al.* 2012). The exposure time for rhodamine was 1 s and for Cy5 it was 6 s, furthermore the brightness and contrast were adjusted for the ladder to make it easier to see it. The merged imaged was then converted to grayscale in MATLAB R2018b (MathWorks). Gel annotations: FT = flow through, W = wash, L = ladder (PageRuler Prestained Protein Ladder, Thermo Fisher) and E = eluate. The first number indicates the column (either 1 or 2) and the second number indicates the wash/elution step.