



UPPSALA
UNIVERSITET

*Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Science and Technology 1046*

lac of Time

Transcription Factor Kinetics in Living Cells

PETTER HAMMAR



ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2013

ISSN 1651-6214
ISBN 978-91-554-8674-7
urn:nbn:se:uu:diva-198814

Dissertation presented at Uppsala University to be publicly examined in B42, Biomedicinskt centrum, Husargatan 3, Uppsala, Friday, June 14, 2013 at 09:15 for the degree of Doctor of Philosophy. The examination will be conducted in English.

Abstract

Hammar, P. 2013. *lac* of Time: Transcription Factor Kinetics in Living Cells. Acta Universitatis Upsaliensis. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 1046. 74 pp. Uppsala. ISBN 978-91-554-8674-7.

Gene regulation mediated by transcription factors (TFs) is essential for all organisms. The functionality of TFs can largely be described by the fraction of time they occupy their regulatory binding sites on the chromosome. DNA-binding proteins have been shown to find their targets through facilitated diffusion *in vitro*. In its simplest form this means that the protein combines a random 3D search in the cytoplasm with 1D sliding along DNA. This has been proposed to speed up target location. It is difficult to mimic the *in vivo* conditions for gene regulation in biochemistry experiments; *i.e.* the ionic strength, chromosomal structure, and the presence of other DNA-binding macromolecules.

In this thesis single molecule imaging assays for live cell measurements were developed to study the kinetics of the *Escherichia coli* transcription factor LacI. The low copy number LacI, in fusion with a fluorescent protein (Venus) is detected as a localized near-diffraction limited spot when being DNA-bound for longer than the exposure time. An allosteric inducer is used to control binding and release. Using this method we can measure the time it takes for LacI to bind to different operator sequences. We then extend the assay and show that LacI slides in to and out from the operator site, and that it is obstructed by another DNA-binding protein positioned next to its target. We present a new model where LacI redundantly passes over the operator many times before binding.

By combining experiments with molecular dynamics simulations we can characterize the details of non-specific DNA-binding. In particular, we validate long-standing assumptions that the non-specific association is diffusion-controlled. In addition it is seen that the non-specifically bound protein diffuses along DNA in a helical path.

Using microfluidics we design a chase assay to measure *in vivo* dissociation rates for the LacI-Venus dimer. Based on the comparison of these rates with association rates and equilibrium binding data we suggest that there might be a short time following TF dissociation when transcription initiation is silenced. This implies that the fraction of time the operator is occupied is not enough to describe the regulatory range of the promoter.

Keywords: gene regulation, transcription factor, lac operon, facilitated diffusion, single molecule imaging

Petter Hammar, Uppsala University, Department of Cell and Molecular Biology, Computational and Systems Biology, Box 596, SE-751 24 Uppsala, Sweden.

© Petter Hammar 2013

ISSN 1651-6214

ISBN 978-91-554-8674-7

urn:nbn:se:uu:diva-198814 (<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-198814>)

Ända in i kaklet

List of Papers

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

- I **Hammar, P.**, Leroy, P., Mahmutovic, A., Marklund, E. G., Berg, O. G., Elf, J. (2012) The lac repressor displays facilitated diffusion in living cells. *Science*, 336:1595-98
- II **Hammar, P.**, Walldén, M., Fange, D., Baltekin, Ö., Ullman, G., Persson, F., Leroy, P., Elf, J. (2013) Transcription factor dissociation measurements using single molecule chase in living cells. *Manuscript in preparation*
- III Marklund, E. G., Mahmutovic, A., **Hammar, P.**, van der Spoel, D., Berg, O. G., Elf, J. (2013) LacI follows a helical path while sliding along DNA. *Submitted for publication*

Reprint of paper I was made with permission from the publisher.

Contents

Introduction.....	11
Gene regulation	12
Escherichia coli and numbers	12
Bacterial transcription.....	13
Transcription factors.....	14
The <i>lac</i> operon model	17
Protein-DNA interactions.....	19
Kinetics of operator binding	19
Facilitated diffusion	22
Present study	26
Experimental procedures.....	26
Molecular biology tools.....	26
Single molecule imaging	27
TF binding to a single, specific operator.....	33
Kinetics of binding	34
Sliding on DNA.....	39
Inside and outside the cell.....	39
The length of LacI sliding on DNA.....	41
Introduced roadblocks as physical barriers.....	42
Probability of binding the operator.....	44
Mode of sliding.....	45
Pseudo-specific binding.....	47
Temperature dependent DNA interactions	48
Tracking of free LacI-Venus	50
Dissociation rates	53
Chase assay.....	53
Implication on gene regulation	56
Conclusions.....	59
Future perspective of single molecule imaging.....	60
Svensk sammanfattning	62
Acknowledgements.....	66
List of references.....	68

Abbreviations

TF	Transcription factor
RNAP	RNA polymerase
LacI	Lactose operon repressor
IPTG	Isopropyl β -D-1-thiogalactopyranoside
ONPF	2-Nitrophenyl β -D-fucopyranoside
cAMP	cyclic adenosine monophosphate
CAP	catabolite activator protein
CRP	cAMP receptor protein
TetR	Tetracycline resistance repressor
AraC	TF of the arabinose operon
bp	Base pair
PFS	Perfect focusing system
PSF	Point spread function
SM(T)	Single molecule (tracking)
(r)msd	(root) mean-square displacement
PDMS	Polydimethylsiloxane
MD	Molecular dynamics
HMM	Hidden Markov model

Introduction

The search problem

Some ten years ago, when attending an outdoor student party, I realized that my keys were lost. They ought to be somewhere on a near-endless grass field and I had no choice but to search for them. Although the scenario that followed could well have been described as a drunkard's walk (1, 2), the light from the full moon and the modest previous intake of beer, reduced the random walk to a slightly less hopeless search; thus, the keys were spotted in less steps than what would be expected for an ideal drunkard's walk.

This experience is a personal way of introducing random walks in living cells (2) as a central concept of this thesis. The work presented here focus on the almost metaphysical question: how is it possible for a protein to search, find and bind, one out of millions of false DNA targets, and to do it on a time scale that allows for a particular regulatory element to function properly? The question is not if, since we know these things actually work, but how. The answers involve the possibilities of intracellular processes that direct the proteins to go beyond what can be explained by random diffusive behavior only.

The dynamic chromosomal structure and macromolecular crowding could easily create an impossible needle-in-a-haystack situation for a given protein-DNA interaction. But the system works, and even though many actions in the cell are described by random events, the blueprint DNA, and evolved functions such as replication, transcription and translation put constraints on randomness. A popular scientific comparison to this is the situation in Luke Reinhart's *the Dice Man*: here the characters allow a throw with the dice to decide how they will move on, however, since they choose the alternatives themselves, the outcome is in part predestined.

Outline

I will first give a general background to gene regulation in *E. coli*, and to the biophysical chemistry of DNA-protein interactions. My PhD work is then summarized by describing how single molecule microscopy was used to study the search and binding kinetics of the transcription factor LacI in living cells. In the end the numbers are put in context of actual regulation, to investigate the possibility of the opposite: the lack of time for regulation.

Gene regulation

The proteome of any cell is dynamic, and regulated so that the blend of proteins changes with growth phases, nutrient supplies and stress. At a given time some genes need to be expressed and others turned off. When subject to internal or external changes this expression pattern alter in response. This promotes survival in a competing environment, and is as important for maintaining cell differentiation in higher order organisms, as it is for adaption in single-cell species. This expression control and signal response can be summarized as gene regulation. Due to their *relative* simplicity, prokaryotes are convenient systems when studying fundamental questions that also concerns other forms of life (3).

While few scientific topics have obvious historical starting points, most of them have groundbreaking milestones. For gene regulation such an early mark must be the presentation of a proposed double helix model of DNA sixty years ago (4). Half a decade later Francis Crick went on to publish a paper entitled *On Protein Synthesis* (5) in which he laid the basis for what would be referred to as the central dogma of molecular biology (6). Here, DNA is replicated to copy itself, *and*, its genes are transcribed to make messenger RNA (mRNA) templates out of them. These mRNA molecules are then translated into proteins – the major players in building, regulating and engineering the intracellular environment.

In the same era, transcriptional control as a regulatory step in the information transfer from genes to protein synthesis was conceptually introduced (7); gene regulation and messenger RNA, together with operators, activators and repressors, were the new key words, and the *Escherichia coli lac* operon was the molecular model. A few years later Walter Gilbert and Benno Müller-Hill isolated the *lac* repressor (8) and proved that is was a protein, and then showed that its counterpart, the *lac* operator, is located within DNA (as opposed to being part of mRNA or protein) (9). James Watson, in whose lab Müller-Hill was partially working, was the editor of both papers.

In his book, *The lac operon – A short history of a genetic paradigm*, Müller-Hill starts the last chapter with “The pessimist’s view: ‘These days very few papers are published in *Cell* which deal with the *lac* system. And when a paper is published, it has to present an unconventional point of view.’” (10). However, transcription regulation is in many cases central when studying the function of protein synthesis in any organism and the *lac* operon of *Escherichia coli* (*E. coli*) is still the molecular model.

Escherichia coli and numbers

E. coli is a gram negative bacterium which uses the flagellar motor for locomotion and fimbria for cell-cell communication and host-cell attachment. Its genome of size 4.6×10^6 bp (4400 genes) forms a densely packed nucleoid

with little space between DNA strands (11). It constitutes less than 0.1% of the gut microbial flora (12) and is above all known to the public as a bug found in contaminated food, and to be the cause of stomach disease. This species, and more precisely the human-isolated strain depicted K-12, has for many years served as a model organism in biology (3, 13), where the ease of cultivation and fast growth rate are among the laboratory advantages. With more information available, such as its genome sequence (14), recent findings about nucleoid organization (15), and regulatory functions of non-coding small RNAs (16), it is of growing interest for live cell studies and system level approaches.

For our questions the advantages of using *E. coli* include the ease of gene manipulation and the availability of several well-described regulatory systems (especially the *lac* operon). In addition the thin (<1 μm) cells have low auto-fluorescence, and can be imaged in one focal plane.

Bacterial transcription

Transcription starts when the RNA polymerase (RNAP) finds and binds a designated promoter (17). The mechanism behind the promoter search has been extensively investigated and the different views balance between free diffusion only (18) and facilitated diffusion (19). *E. coli* RNAP is a complex of subunits $\beta'\beta\alpha_2\omega$, the core enzyme, which associates with one of seven different σ subunits to produce the holoenzyme (20). The σ unit constitutes the direct promoter recognition element and the most generic of the seven is σ^{70} (RpoD) which regulates most genes expressed in exponentially growing, non-stressed conditions (21). Binding of RNAP is either non-regulated (constitutive) or directed and enhanced through activator proteins, or reversely inhibited by repressor proteins. The typical promoter covers roughly 35 bp with the conserved consensus sequences -35 (TTGACA) and -10 (TATAAT; Pribnow box) and the +1 start site. Binding of holoenzyme to promoter is known as closed complex formation (20).

Transcription is initiated at the start site when the bound polymerase unwinds the DNA duplex and forms a bubble – the open complex formation. RNAP then produces some truncated ~10 nt RNA sequences (abortive initiation), before the continuation of complete transcription (22). Whether the σ factor is released or moves along with the core during RNA elongation is yet to be settled (17). Elongation of the transcript proceeds in single-nucleotide steps (23), at about 40 nucleotides/s (24). It can slow down due to sequence-specific pausing and possibly also backtracking. When completed, the transcript terminates either through the formation of a destabilizing RNA hairpin loop, or by the action of the helicase Rho (17). Translation is initiated when the anti-Shine-Dalgarno sequence of the ribosomal 16S RNA binds the Shine-Dalgarno sequence of the mRNA (25). RNA polymerases are in short

supply and one mechanism by which they are beneficially shared between competing promoters is through the binding of transcription factors (20).

Transcription factors

Experimental and computational approaches were used to characterize 314 DNA-binding, regulatory proteins in *E. coli*; about forty percent are activators, forty percent are repressors, and the rest have dual functions (26). Seven TFs (CRP, FNR, IHF, Fis, ArcA, NarL, Lrp) have been categorized as global regulators; together they control about 50% of the regulated genome (20). Of the rest, some TFs are regulators of a unique gene or operon and some are regulators of a set of transcriptional units. Bacterial, as opposed to eukaryotic (27), TFs typically act independently by binding to specific sites on the DNA called operators (7, 9).

To illustrate TF-mediated gene regulation, the functionality of a few transcription factors will be described below. The selection is based on the proteins studied and used in this work. Because of its exceptional importance – in this thesis and in general – the *lac* operon has been given its own section.

CAP/CRP

Catabolite activator protein (CAP) or Cyclic AMP receptor protein (CRP) has two names, and regulates more than 100 promoters in *E. coli*. Through the binding of cAMP and a resulting allosteric change, the protein can recognize and bind an activator site. CRP is a homodimer and bends its DNA binding site to an angle of about 80°. By interactions with one or several positions in or near the promoter, it recruits the RNAP and/or stabilizes its binding (28).

The formation of active CRP-cAMP complex and the regulation of the two molecules are complicated; glucose-rich medium reduces intracellular concentrations of both CRP and cAMP. In an intriguing scheme (fig. 1A) the glucose-importer protein (IIA^{GLC}) becomes dephosphorylated as import and phosphorylation of extracellular glucose occurs. In this form IIA^{GLC} binds transport proteins of several other carbon sources and inhibits their function (including lactose-import). This is termed inducer exclusion. In addition, when phosphorylated the protein induces the adenylate cyclase activity and hence the formation of cAMP from ATP, *i.e.* when IIA^{GLC} is dephosphorylated cAMP production is reduced (29). The resulting reduction in CRP-cAMP concentration has the effect that less CRP is produced since its promoter is positively auto-regulated by CRP-cAMP. In contrast, CRP-cAMP acts with negative feedback on adenylate cyclase transcription and activity so that less cAMP is formed when complex concentration is high (30). This influence of glucose on CRP-cAMP levels, and consequentially on promoter activity, is an example of catabolite repression. It can be understood from the other way around: as a response to lowered glucose levels, there is an in-

crease in CRP-cAMP, and in the uptake of other carbon sources. In this way it works as indirect repression and is a complement to the actual repressors. This illustrates how intertwined the positive and negative control systems can be (31, 32). Most CRP binding sites appear to be unsaturated *in vivo*, and thus their differential activator affinity will direct how CRP is distributed among promoters as cAMP levels increase and decrease. Naturally, this regulation implies that going from low to high CRP-cAMP concentration determines which genes will be up-regulated first following a common global signal (33).

L-arabinose metabolism

The metabolic pathway of L-arabinose utilization starts with membrane transport, either through the product of *araE*, or through the proteins expressed in the independent operon *araFGH*. Both these transporter systems respond to extracellular L-arabinose themselves and are the first steps of positive feedback (34). Intracellular processing of the sugar is a step-by-step reaction passing the three enzymes of the operon *araBAD*, which finally sends the product D-Xylose-5 phosphate into the pentose phosphate pathway. The *araBAD* genes are expressed from the promoter P_{BAD} , and the transcription factor of the operon, AraC, has its gene controlled by the divergently transcribing P_C (fig. 1B). In between the two promoters is the binding site for the positive regulator CRP-cAMP which activates either one or both of the two promoters and is counteracted by the presence of glucose (35).

The homodimer AraC was the first protein experimentally shown to take part in DNA looping – the simultaneous binding of a protein to two different DNA sites (36). The TF binds with one monomer each at the half-sites I_1 and the upstream O_2 (within *araC*), for full repression. The interaction of L-arabinose with AraC not only breaks the DNA loop, but also induces the relocation of one monomer binding domain to bind I_2 next to I_1 . This allows for RNAP binding at both promoters and, if available, CRP-cAMP will interact to give full expression (which for the *araBAD* genes is 300 times higher than the basal level) (35).

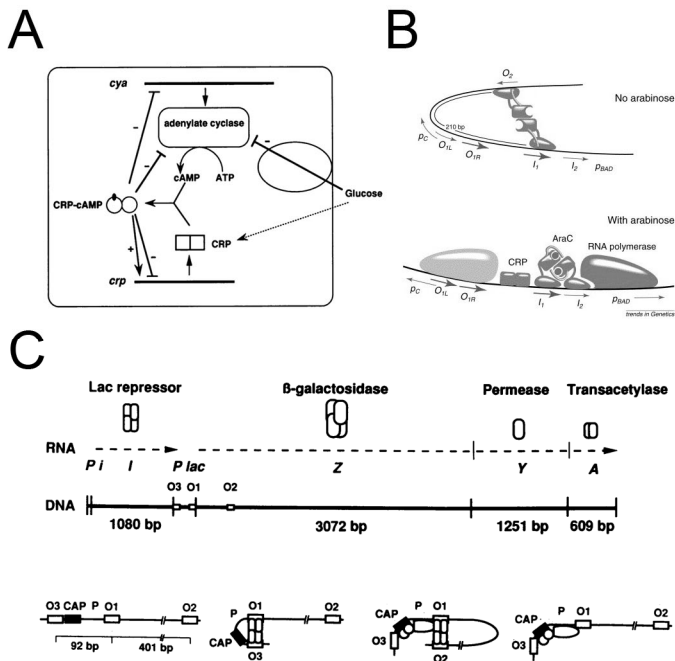


Figure 1. Gene regulation. (A) CRP auto-regulation, (B) *araBAD* repression and activation, (C) the *lac* operon and different modes of DNA looping. Adapted with permission from respective publisher (35, 37, 38).

Tetracycline resistance

The antibiotic tetracycline (Tc) passes through the bacterial membrane, complexes with a divalent cation (typically Mg^{2+}) and inhibits protein synthesis by binding to the A-site of the ribosomal subunit 30S (39). A common resistance mechanism in gram-negative bacteria is to express membrane proteins for specific export of the antibiotic. The protein TetA acts on $TcMg^{2+}$ with an antiporter functionality, where the antibiotic-cation complex is moved out of the cell against the proton gradient. This is needed to restore ribosome function; however overexpression of TetA is toxic since non-specific cation export leads to the collapse of the membrane potential (40).

TetA and its control element, the *tet* repressor (TetR), is expressed from bidirectional genes with overlapping promoters. Two operator sequences, O_1 and O_2 , with only 11 bp spacing edge-to-edge, lie within the promoters. TetR homodimers bind the operators independently, possibly with a weak cooperativity, repressing TetA both at O_1 and the slightly stronger O_2 , and acting with negative feedback on TetR at O_1 only (41). $TcMg^{2+}$ binds TetR, causes an allosteric structural change and induces repressor release, such that both TetA and TetR is expressed (40). In this way the complete *tet* regulatory element (which itself can be transferred between cells via the transposon

TN10) responds to tetracycline uptake by pumping it out, and then reducing the toxicity of the pump by the rebinding of the repressor.

The *lac* operon model

We tend to emphasize how certain knowledge has been around *forever*. In the first paragraph of many publications on gene regulation, the work of Jacob and Monod is cited that way. Before I give an account of the *lac* operon, I will quote the masters themselves; to give perspective. “It has been known for over 60 years (Duclaux, 1899; Dienert, 1900; Went, 1901) that certain enzymes of micro-organisms are formed only in the presence of their specific substrate.” (7)

Regulation of the *lac* promoter

The *lac* repressor (LacI, sometimes LacR) is expressed just upstream of the actual *lac* operon (fig. 1C) and its monomers form homodimers. Each monomer has a distinct domain in the N-terminus for DNA recognition and binding, which is made possible only after dimerization (42). The core of the protein contains a sugar (inducer) binding pocket that following inducer-binding transmits an allosteric transition through the protein body that changes the DNA-binding affinity (43, 44). Two dimers form a tetramer when all four C-terminal helices oligomerize (43). It has been reported that of the ~20 repressor monomers per chromosome equivalent (8), >99% should be present as tetramers in wild-type cells (45). Many natural and synthetic inducers and anti-inducers have been characterized to bind LacI. Isopropyl β -D-1-thiogalactopyranoside (IPTG) is commonly used to induce the *lac* operon, and 2-Nitrophenyl β -D-fucopyranoside (ONPF) is an efficient anti-inducer (44).

Full activity from the *lac* promoter is seen when LacI is induced by the binding of allolactose, and at the same time a shortage in glucose removes the catabolite repression (46). The promoter region has been subject to extensive and detailed studies, and it is suggested to contain four overlapping promoters, several binding sites for the structural protein H-NS, and two for CRP-cAMP; see *RegulonDB* for an overview (47). The elements likely to be the most important are summarized here. The main operator, O_1 , overlaps with the transcription start site; the two auxiliary (and weaker) operators O_2 and O_3 are located 401 and 92 bp downstream and upstream of O_1 respectively. CRP-cAMP binds adjacent to, just downstream of O_3 (38, 48). As was first seen for the AraC monomer-monomer DNA loop, the two dimer homologs of LacI was shown to enhance repression by co-binding of O_1 and one of the auxiliary operators (49). The role of DNA looping in the *lac* operon has been further described with respect to local concentration effects and binding kinetics (38, 50, 51).

Genes of the *lac* operon

The three genes of the operon are *lacZ*, *lacY* and *lacA*. Encoded by *lacY*, the membrane transport protein lactose permease (LacY) directs the symport, the co-translocation of a galactoside and an H^+ , against the concentration gradient (52). The synthetic inducer IPTG is also actively transported over the membrane. However, when LacY function is impaired, the inducer still crosses the membrane passively and at higher IPTG-concentrations full induction can be reached (53). The main function of LacY is to bind and import lactose.

β -galactosidase (LacZ, from *lacZ*) catalyzes the conversion of lactose substrate molecules (54). The enzyme is multi-functional so that, *in vitro*, about half of the lactose is hydrolyzed and cleaved into galactose and glucose, and half is transformed (through transglycosylation) into allolactose. This isomer acts as inducer for LacI, however is also a substrate for the enzyme in further processing into glucose and galactose (55). Galactose is then converted to glucose by the enzymes of the *gal* operon, and induces expression from the same operon when binding to its repressor GalR (56). This circuit where the expression from the *lac* operon is up-regulated when its proteins sense the same metabolite that it will process gives rise to a cascade effect that is an example of positive feedback.

Literature covering *lacA*, the third gene in the operon, is sparse. Its product is thiogalactoside transacetylase (LacA). The role of this protein, if any, in the *lac* operon is unclear. A few studies suggest that LacY-imported non-metabolizable molecules are acetylated and then diffuse out from the cell (57).

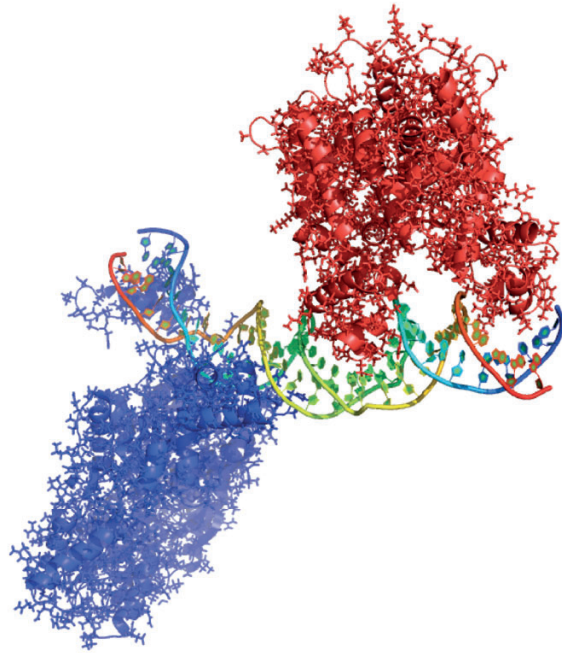


Figure 2. Binding of LacI (blue) and TetR (red) to DNA when their operator sites overlap with one base pair.

Protein-DNA interactions

In the previous section, gene regulation is simply described as the action of the interactions between two or more macromolecules, and this fail to consider the important bond formations between amino acids on the protein binding domain and the nucleotides in the DNA chain. This section will point out that DNA binding proteins are able to recognize certain regulatory elements on the chromosome, and, discriminate between these and all other sequences that compete for their attraction.

Kinetics of operator binding

Almost as legendary as the early papers on DNA structure and gene regulation is the *in vitro* nitrocellulose-filter binding experiments on the *lac* repressor-operator interaction, used in a series of studies and most notably in (58). Not only was this a pioneering method. The results, on association and dissociation rates, salt effects and temperature dependence, lay the ground for numerous theoretical and experimental studies to come. The most highlighted result in this publication was the extremely fast binding at the operator site, which (at $\sim 7 \times 10^9 \text{ M}^{-1}\text{s}^{-1}$) is apparently faster than what would be

possible by 3D diffusion alone; the suggested explanation, which will be further discussed in the section “facilitated diffusion” (p. 22), postulates that operator-binding is preceded by non-operator DNA interactions.

Consequences of non-operator binding

Non-specific interactions between proteins and DNA are characterized by events that slow down the free diffusion of the proteins, and at the same time do not exhibit any regulatory behavior. It was presented for the *lac* repressor tetramer (59) and dimer (60) that on average it is non-specifically DNA-bound 90% of the time in living cells. Binding of inducer IPTG to LacI does not change the repressor-DNA interaction, and presence of DNA does not change the affinity of IPTG for LacI (61). This is in agreement with the observation that complex-formation between DNA and repressor leave their respective structures unchanged (62). The binding domain of the repressor is the same for non-specific and specific interactions (61, 63) and enables a structural switch between non-specific and specific binding configurations (62).

The non-specific configuration is mainly electrostatic, with basic amino acids forming ionic interactions with the phosphates of the DNA backbone (61, 62, 64). A suggested mechanism for the actual non-specific association, is that first the repressor is protonated, and then, if in a “sufficiently dilute” ionic solution, monovalent cations are released from the DNA and replaced by the positively charged repressor binding unit (64). In this way the association between non-specific DNA and repressor is driven by the release of cations (the effect being the same for Na^+ and K^+ (63)). The concentration of ions indeed has huge effects. Increasing concentrations of either monovalent (for the reason described above) or divalent (Mg^{2+} acts as a competitor with the repressor for DNA binding (63)) ions weakens the binding; for example, the effect is 100-fold between 100 and 150 mM Na^+ (59, 64). Binding strengths of repressor for DNA is also significantly decreased, when using a fixed salt concentration and pH, and increasing the temperature from 4 to 38°C. In addition, when pH is varied in the interval 7-8 it is shown that lower pH increases the observed binding constant. This is explained by increased charging of residues in the DNA binding domain and hence electrostatic contributions to the DNA binding (61, 64).

Operator-binding

The ratio between the binding constants for repressor-operator and repressor non-operator respectively is on the order of $\sim 10^8$, however this number depends on pH, salt concentration and temperature (63). It was reported that LacI binds the operator faster at higher temperature (58, 65), and has an “optimal temperature” of $\sim 20^\circ\text{C}$ for dissociation, where the binding time decreases towards higher and lower temperature (65). The association rate decreases with higher pH whilst this effect on dissociation is negligible (58).

It was repeatedly shown that dissociation rates increase with higher salt (58, 65, 66). For association rates it is more interesting. Adding 100 mM cations to a low salt standard buffer decreased the association rate (58); changing from 150 to 250 mM cations also decreased the association rate (65). However, when the interval from 25 to 200 mM was covered it was shown that the association rate increases up to and peaks at ~100 mM and then decreases again (or plateaus, depending on DNA construct) (66, 67). This apparent optimum is understood together with facilitated diffusion, which is described in the next section.

The temperature effect on non-specific and specific DNA binding taken together, showed an increased ratio between specific and non-specific binding at higher temperature. This implies that the non-specific DNA competes for repressor binding more effectively at lower temperatures. This, and an observed temperature dependence on the LacI-IPTG association rate, led to the conclusion that the temperature effects are caused by structural changes in the protein (65).

From this data on binding between *lac* repressor and operator or non-operator DNA, it was concluded that the protein first associates with non-specific DNA through electrostatic interactions, and then undergoes a conformational change that aids the recognition of the operator site. In phase with the correct nucleotide sequence, the protein can form specific hydrogen bonds, and electrostatic and hydrophobic contacts, with the bases of the DNA (43, 62, 66).

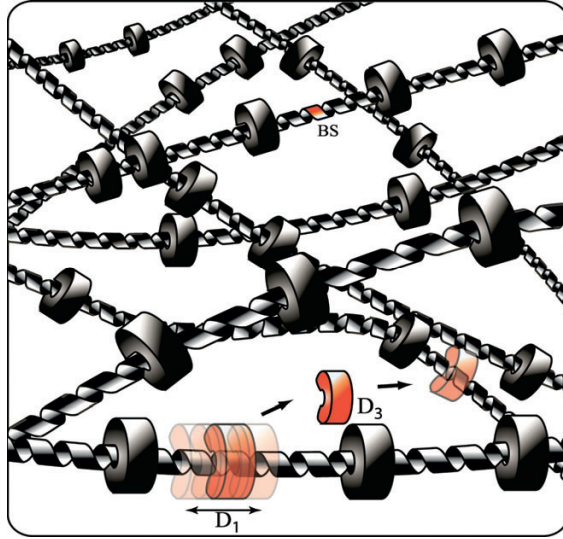


Figure 3. Facilitated diffusion. The red cylinder represents a DNA-binding protein that combines 3D diffusion in the cytoplasm and 1D diffusion along non-specific DNA while searching for its specific binding site (BS). The black cylinders represent other DNA-binding proteins that might affect the search. *Adapted with permission from (51).*

Facilitated diffusion

The first step in bimolecular binding processes is the transportation through diffusion and the second step is interaction and reaction. As opposed to the ATP-dependent movement of for example helicases in the replication fork (68), most reactions are usually preceded by random diffusion. Macromolecules diffuse as an effect of thermal movement in the solvent and the clashes with other molecules in the surrounding (fig. 3). The molecule is said to undertake a random walk, or Brownian diffusion when no additional forces are involved (2). It was also shown that a fluorescent protein, studied in living *E. coli*, exhibit simple Brownian diffusion (69). Here, the distance covered by a molecule as a function of its diffusion coefficient is given by the mean-square displacement (msd) (2), for diffusion in one (1D), two (2D) or three (3D) dimensions:

$$\Delta x^2 = a_{n=1,2,3}Dt \quad (1)$$

with $a_1=2$, $a_2=4$, $a_3=6$ for 1D, 2D and 3D respectively. The diffusion coefficient of the macromolecule is approximated by the Stokes-Einstein relation for spherical particles (70).

The simple bimolecular association reaction rate for two macromolecules (*i.e.* repressor and operator) is:



Riggs and colleagues compared their data on repressor-operator association with the von Smoluchowski equation (eq. 3) for diffusion and reaction between uncharged macromolecules. This gives the theoretical maximal forward rate constant (k_a in eq. 2) due to the diffusion rate limit according to (58):

$$k_a = 4\pi D_{12} r_{12} N_0 / 1000 \quad (3)$$

where D_{12} ($= 5 \times 10^{-7} \text{ cm}^2 \text{ s}^{-1}$) is the sum of the diffusion coefficients, r_{12} ($= 5 \times 10^{-8} \text{ cm}$) is the reaction radius (*i.e.* the distance at which the bimolecular reaction is expected) and N_0 is Avogadro's number. From this they estimated that $k_a \sim 10^8 \text{ M}^{-1} \text{ s}^{-1}$, which is 10-100 times lower than their measured value and thus implies that the *lac* repressor-operator interaction is faster than a diffusion-controlled reaction.

Eq. 3 assumes that every collision leads to a reaction. Indeed, the bottom-line in a diffusion controlled reaction is that the probability of binding within a single macroscopic collision is very high since a first encounter is followed by multiple secondary collisions where the molecules try different configurations aided by thermally driven rotations. However, since only a fraction of the areas of the molecules potentially take part in binding, the probability of reaction might be smaller. Hence the diffusion limited rate is overestimated in eq. 3 (66).

The suggested explanation from the first experiments of Riggs *et al.* (58) was that the electrostatic attractions between positive charges on the repressor and negative phosphate groups within the operator DNA could speed up the reaction. It was further speculated that rather than through a 3D random walk, LacI is directed towards the operator by "relatively long-range electrostatic forces", and searches by rolling or hopping along non-specific DNA. In the end they abandoned this idea, and concluded that electrostatic attraction between operator and repressor would increase the effective reaction radius to the extent that the fast association rate is explained by this effect alone (58).

One-dimensional diffusion

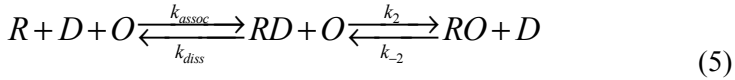
It was then shown that the electrostatic attraction from the operator is not enough to explain the rate enhancement (71). Instead the theory of contribu-

tions from non-specific DNA-interactions around the operator, and thus a target extension along the DNA axis, was developed: the repressor diffusing in 3D (with diffusion constant D_3) binds non-specific sequences by random, and during the time bound ($\tau_{bound,ns}=1/k_{diss}$) is able to diffuse along the DNA chain in a 1D random walk before dissociation and re-association at any uncorrelated non-specific site.

Although the 1D diffusion coefficient $D_1 \ll D_3$ (72, 73), the overall probability to find and bind the operator is increased since the 1D search is more extensive than the 3D search. The target extension is a function of $\tau_{bound,ns}$ and D_1 (71, 72, 74):

$$s = \sqrt{D_1/k_{diss}} \quad (4)$$

The effective capture distance of the operator stretches out from its center in both directions along the DNA and is thus $2s$ (s for *sliding*, *i.e.* 1D diffusion). If the rate of binding the operator is influenced by this, then a decrease in association rate is expected when the length of the DNA fragment is reduced to below the size of the effective target. The random 3D diffusion and reaction with any non-specific site, followed by binding to the operator, is described by a two-step reaction scheme (74):



Binding to non-specific DNA (D) is the intermediate reaction, and transfer to operator (O) the second step such that the overall rate of repressor (R) binding to O is the same as k_a in eq. 2. The positive contribution from sliding, in terms of an increased specific binding rate, is qualitatively best described as a reduction in the number of random trials-and-errors needed until the operator site is bound.

The rmsd covered by a sliding protein during a non-specific search is given by the expression in eq. 1 for 1D diffusion:

$$s_L = \sqrt{2D_1\tau_{bound,ns}} = \sqrt{2}s \quad (6)$$

Hence, the piece of the DNA chain “searched through” increases with the time non-specifically bound. For sliding to occur, the electrostatic attractions need to beat the drive to dissociate, and at the same time the energy barrier should be small enough to allow movements along the DNA. The translocation, as with the non-specific interaction itself, depends on the displacement of cations; when the repressor reaches a new position on the DNA, ions are

displaced, and at the same time ions occupy the position just left (66). The finding described above, that the operator association rate has a salt-dependent optimum is explained by that at too high salt, the effect of non-specific binding and sliding is diminished, and at too low salt, these non-specific interactions trap the proteins on the wrong DNA segments (66, 67).

Sliding events of various lengths has been shown for several different proteins *in vitro*, and this will be discussed later in this thesis. There is however other mechanisms that have been suggested as facilitators in target location.

Intersegment transfer

With two or more DNA binding domains on the protein, it can possibly bind more than one non-specific sequence at a time. Being attached at any location, the protein can interact with another DNA strand that swings by as a result of diffusion. If this “sandwich” complex is unstable and short-lived, and the probability for the protein to stay at any of the two strands following dissociation from one strand is equal, this could speed up the search (74). The proposed mechanism, termed intersegment transfer, is affected by the strength of the non-specific binding, and by the packing and structure of the DNA (45, 74).

Looping

The maximal repression of the *lac* operon is achieved without using the strongest known DNA-repressor interaction (75). Instead, the LacI tetramer (as well as several other *E. coli* TFs) can bind simultaneously to two of three available operator sequences (76). If the initial binding is to any of the auxiliary (helper) operators, then subsequent looping to the main operator is a facilitator. When bound to an auxiliary operator, the local repressor concentration in the promoter region is increased and so is the probability to find and bind the main operator (38, 51).

Present study

In this work *E. coli* and its *lac* operon is used to study the kinetics of protein-DNA interactions in living cells. The motivation can be seen from two sides. First, that the aim is to further develop single-molecule methods for intracellular kinetic measurements previously designed (60, 77), and to use a well-described regulatory system as reference; second, that by modifying an already established set of tools, we can aim to understand how mechanisms that are implied by *in vitro* studies are used in living cells. From my point of view, I started on one side and ended up on the other.

I will first give an account of the single-molecule imaging technology, and microscopy in general. Following measurements on specific binding in relation to physiological parameters, it is then asked whether facilitated diffusion takes place in living cells. Rate constants and proposed mechanisms, including pseudo-specific binding, are compared to previous models and *in vitro* results. Finally, microfluidic devices are used, primarily to make dissociation rate experiments. The papers listed in the beginning and attached in the end are the pillars on which most results are built. Some additional data are denoted as such.

Experimental procedures

Methods are presented in paper I-III and are further described here to summarize how they were used and/or developed in this work.

Molecular biology tools

All *E. coli* strains used branches from a BW25993 background (78) which is the K-12/MG1655 (lacking bacteriophage lambda and the F plasmid) (14) with mutation $\Delta araBAD_{AH33}$ (among others, but for us the most relevant); the genes for L-arabinose metabolism are removed and those involved in uptake are left intact. Many of the strains were constructed starting from JE12, in which the gene encoding for the fluorescent protein Venus^{A206K} is fused to the C terminus of *lacI* (60).

Cloning protocols

Chromosomal deletions, insertions or replacements were done using either the “pKO3” protocol (79) or a lambda Red recombinase-based method (78). I used the first the most. Described in brief, it is itself a good example of gene regulation: a chromosomal region of interest is cloned into a plasmid (pGEMt backbone) and variants of PCR are used to modify the sequence. The insert is sub-cloned into the vector pKO3 which contains selection markers for chloramphenicol resistance, sucrose toxicity, and temperature dependence. The transformed recipient strain is selected on chloramphenicol at 30°C (allowing for amplification of plasmid), followed by a shift to 42°C (plasmid replication is prevented; only cells where the plasmid recombines with the chromosome keep antibiotic resistance). A final shift to sucrose and 30°C selects for cells where the plasmid is excised, and PCR is then used to screen for positive gene replacement.

Cell culture, protein expression and activity

Cells were typically grown in M9 minimal medium supplemented with 0.4% glucose, MgCl₂ and CaCl₂ (referred to as M9-glucose). Amino acids and antibiotics were added as noted in different experiments. Overnight cultures were started from glycerol stocks, diluted next day 1:500 in the same medium, and collected in early- or mid log phase (OD₆₀₀=0.1-0.2). The β-galactosidase (Miller) assay was essentially performed as described: cells are cultured +/- induction, disrupted, and mixed with the substrate 2-Nitrophenyl β-D-galactopyranoside (ONPG) at saturating concentration, which is cleaved by β-galactosidase; one of the products is yellow so when the reaction is stopped its absorbance can be measured as a readout of the enzyme activity which is proportional to the concentration of expressed *lacZ* (80). The repression ratio (efficiency) is calculated as the activity of induced divided by activity of repressed cells. High expression of TetR or LacI were in most experiments obtained using the plasmid pBAD24 (81); the TF is controlled by a modified P_{BAD} promoter and AraC is also expressed from the plasmid. Induction of P_{BAD} is nonuniform between cells at subsaturating arabinose concentrations (less than ~1mM) (82). We induce at 0.2% arabinose (~13mM).

Single molecule imaging

Live cell fluorescence imaging was done using the setup presented below.

The microscope

Data were collected using an inverted Nikon Ti Eclipse microscope for epi-fluorescence imaging. Cells were imaged in two different channels: phase contrast images were acquired for the purpose of cell counting, cell segmen-

tation and cell tracking; fluorescence images were collected to visualize expression, localization and diffusion of LacI-Venus fusion proteins.

For excitation of Venus, we use an Argon⁺ laser (Innova I-304, Coherent) set to 514 nm. In order to remove spatial intensity variations, the laser beam is focused on a 100 μm pinhole (Thorlabs) to create a spatial filter. The filtered beam is collimated with a second lens and the central disk of the resulting airy pattern is cropped with an iris. The beam is 10-fold magnified and the center of it cropped by another iris, so that the Gaussian shape is flattened out and illumination becomes even. The microscope is shown in fig. 4. A lens focuses the image of the iris onto the back aperture plane of the oil immersion objective (100X/ $NA=1.49$ (Nikon)), which sends a collimated beam to the sample. The microscope contains a filter cube with excitation filter (514/10), dichroic mirror (t515.5 rdc) and emission filter (550/50) (Chroma). As will be discussed later, in some experiments an additional 2X lens is placed in front of the aperture of the EMCCD detector (iXon EM+DU-897 from Andor).

The microscope has a hardware autofocus system (perfect focus, PFS) that is stable for days, and an automated xy-stage. Positions can thus be saved and the imaging computer controlled through the software $\mu\text{Manager}$ (83). When microfluidics is used, medium can be switched in the chip using gravitational flow and computer controlled linear actuators that are timed with the time-lapse imaging. The microscope is enclosed in an incubator-hood so that it can be heated and the temperature kept stable.

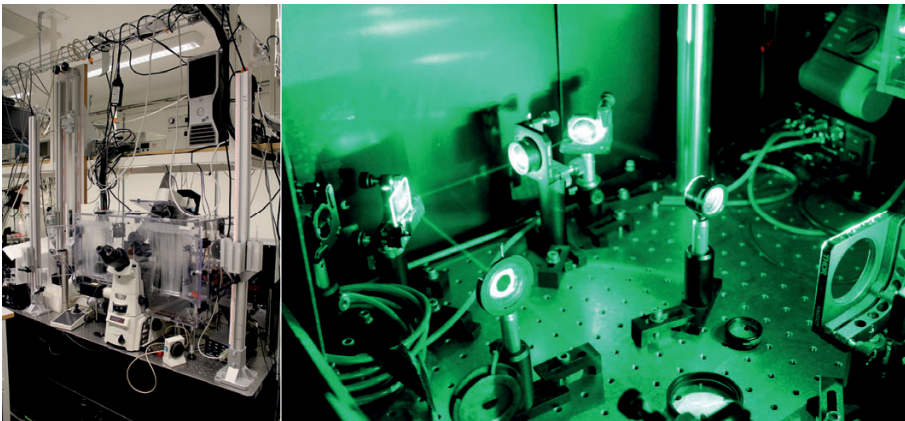


Figure 4. The microscope. (*left*) the microscope is enclosed in an incubator-hood that controls temperature, and is flanked by linear actuators for medium exchange, (*right*) a part of the optical setup where the laser beam passes an iris, mirrors and lenses before entering the microscope at the back (not shown).

Tagging and detection

LacI is expressed as a fusion with a fluorescent protein at the C-terminus which does not affect the N-terminus DNA binding domain but inhibits tetramer formation (60). The gene for Venus is derived from the original *Aequorea victoria* green fluorescent protein. It has a shift in absorption (excitation) maximum to around 515 nm, which changes the fluorophore of the barrel-shaped protein from green to yellow with emission maximum around 528 nm (84). Additional mutations improve folding at 37°C, accelerate oxidation, make it less sensitive to acids and chloride ions and increase the brightness of the fluorophore (85). The maturation time, *i.e.* the average time from protein expression until the fluorophore can be excited, has been estimated to ~7 min *in vivo* (77) and even faster *in vitro* (85). Venus was also subject to the mutation A206K (60) which has been experimentally shown to make the protein monomeric even at overexpression (86).

The principle behind our single molecule imaging is “detection through localization” (60, 77). A molecule that is stationary or slowly diffusing can be temporally separated from fast moving objects. At exposure times longer than ~1 s, light is collected from a near diffraction-limited location if a molecule within this limit is immobile. This spot can then be distinguished from other, non-localized proteins that, at this time scale, diffuse around and spread their emission from the whole cell. In this way bound LacI-Venus can be detected and separated from non-bound. At shorter exposure times also fast moving objects can be detected and this makes it possible to detect and follow fast-diffusing proteins. In general, exposure times should be set according to the mobility of the object. One can think about a road with cars; some are parked, some are driving. A snapshot of the road with an exposure time so short that a driving car does not have time to move within the field of view does not allow for the distinction between the driving and the parked cars. Instead the moving car can be tracked down with many snapshots at fast frame rate. A long exposure will image the moving car as a smear and the parked cars as stationary objects. The distinction can be made. In fig. 5 a short exposure cannot tell if the cat is moving, but can catch the droplets in the running water; a long exposure cannot resolve the water droplets, but shows that the cat is moving. It will later be shown how different exposure times are used, but the rule of thumb is that bound LacI are detected with a long exposure (>1 s) and freely moving LacI are detected with a short exposure (<~3 ms).

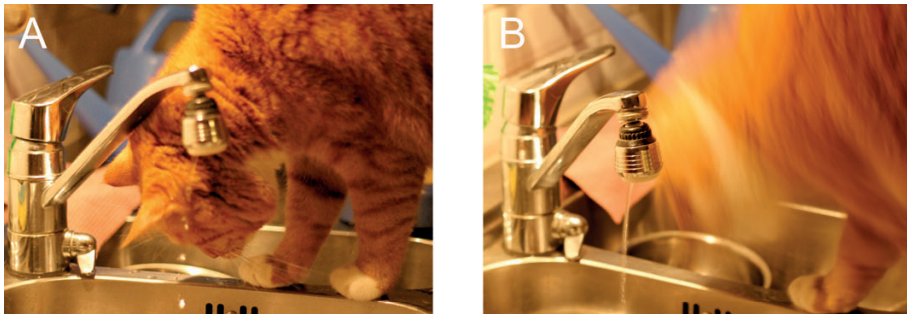


Figure 5. Different exposure times. **(A)** at a short (ms) exposure with the camera the cat appears to be still, and the water droplets freeze. **(B)** at a long exposure (s) it is seen that the cat is moving, and the droplets of the running water cannot be resolved.

It should be remembered that we do not image the bound protein, which have a size of ~ 5 nm, but rather the light from the Venus tag fused to the *lac* repressor. A critical limit in the assays is the total number of repressors. If the protein concentration is 3 per chromosome, the bound to non-bound ratio become 1:2, whereas if the total concentration increase to 10 the ratio become 1:9 and the localized spot is nearly overwhelmed in fluorescent background (the actual concentrations are rough estimates and so is this example). For single molecule detection with 514 nm excitation this is a more challenging issue than the cellular auto-fluorescence.

Preparations and microfluidic devices

Experiments were done either employing commercial culture dishes mounted with cover slips (60) or custom made microfluidic devices (87). Both approaches enable live cell imaging. The culture dishes, when coated with poly-L-Lysine, are convenient for experiments done at room temperature and spanning a time range of the cell cycle or shorter, with the preparations being fast and with a high probability of having all cells in the same focal plane (fig. 6A). The main issues with this approach are, a weakened cell-coverslip attachment at higher temperature, and that growing cells will start to detach and become mobile. The chip design allow for experiments over longer time scales since there is a continuous inflow and exchange of growth medium and space for the cell colonies to grow while still staying in focus (fig. 6B-C). This also simplifies experiments at higher temperatures and enables the shift of growth medium to be done in a controlled way. With the cells filling well defined traps, quantification of data is more straightforward. A drawback is that the polydimethylsiloxane (PDMS, the material of the chips) create a slightly higher background fluorescence compared to the simple culture dishes.

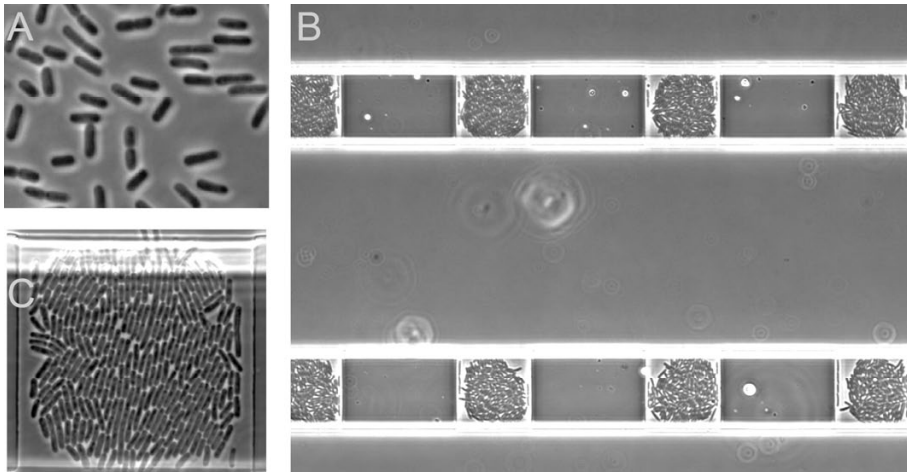


Figure 6. Phase contrast images of living *E. coli*. (A) attached on poly-L-Lysine coated culture dishes at room temperature (magnification 100X), (B) growing at 37°C in a microfluidic chip; the chip contains 3x17 independent traps, each with a size of 40x40 μm (20X); (C) same as in B, although with 100X magnification.

Image analysis

Our imaging data has two parts: one is phase contrast images that are used to quantify the number of cells; the second is fluorescence images that are used to quantify the number of spots (*i.e.* single proteins) per cell. Cells imaged on dishes are well separated and badly ordered (fig. 6A), and were analyzed by manual counting. In the microfluidic chip cells are densely packed, without any spacing in between. An algorithm used for automatic segmentation of cells and tracking of lineages were developed in the lab (87) and applied in some of the experiments in paper II (as indicated). In the kinetic experiments performed in the chip, we assume that the total cell area scales to the number of cells. Fast analysis is therefore achieved by selecting a square area of a cell-trap that is completely filled, calculating the number of pixels and normalizing the number of detected spots with this area.

Images containing fluorescent spots were analyzed using automated spot-detection (although preceded by a large amount of manual evaluation) in MATLAB. The theory behind the spot detection will be briefly covered.

We search for near-diffraction limited spots. According to the Rayleigh criterion (applied to a microscope image), the smallest resolvable distance (d) of two adjacent object points is where the airy disk of each point is separated. λ is the wavelength of the detected light and NA is the numerical aperture of the objective lens (88).

$$d = 0.61\lambda / NA \quad (7)$$

Objects with size smaller than $2d$ appears as circular diffraction disks (described by the point-spread function, PSF) and are thus diffraction limited; they will have a diameter of $2d$ and can be resolved from other particles down to the resolution limit d (88). Instead of a function corresponding to the airy disk, the PSF can be approximated with a Gaussian. When looking for localization of diffraction limited spots, their actual position can be determined with accuracy below the resolution limit by calculating the center coordinates of the PSF. The best estimate of the position is given by the average of the positions of the detected photons. The precision is described with the localization error, which decreases with the standard deviation of the PSF and with an increased number of photons collected from the same object (*i.e.* better statistics) (89).

The EMCCD collects images with each pixel corresponding to 160 nm (100X magnification) or 80 nm (200X). The theoretical size of the PSF for one localized, diffraction-limited protein is $2d \sim 430$ nm, with the emission peak of Venus = 528 nm, and NA of our objective lens = 1.49. Thus, each spot should be found within an area of $\sim 3 \times 3$ and $\sim 6 \times 6$ pixels respectively. The spot can also be described by the full width at half maximum (FWHM) of its intensity profile (fig. 7). Different algorithms for spot detection have been tested with similar result. In principle, all of them work by finding bright spots in an image. In paper I and III, the stable wave detector (SWD) or the Isotropic Undecimated Wavelet transform (IUWT) are used to define spots in an image; the spots are then fitted with a 2D Gaussian function to select for goodness-of-fit, brightness and standard deviation (roughly size). In paper II, a method based on an à trous wavelet 3 plane decomposition (90) and spot detection in the second wavelet plane is used, without applying any Gaussian filter.

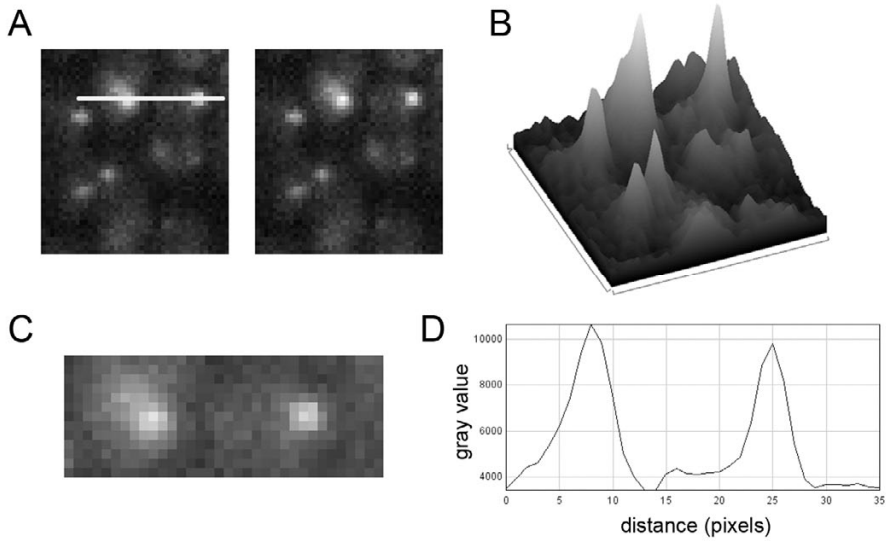


Figure 7. Fluorescence images. **(A)** the two images show a selection of five cells in the fluorescence channel. The two bright spots with a crosscut are magnified in C, **(B)** a surface plot of the cells in A, **(C)** magnified spots as described in A, **(D)** intensity-diagram of the magnified spots in C shows that the full width at half maximum (FWHM) is ~ 4 pixels for each peak; this corresponds to ~ 300 nm.

TF binding to a single, specific operator

The assays for single molecule imaging in living cells developed in the Xie lab have been extended to monitor protein-DNA binding at single operator sites (paper I).

First, the symmetrical artificial operator *lacO_{sym}* (also: *O_{1sym}*, *O_{id}*) was introduced in the end of the *lacI-venus* fusion gene; in a position where it contains the stop codon of the gene, *i.e.* in the position of *lacO₃* in the wt *lac* region. In this way the expression of *lacI-venus* is reduced due to auto-repression from the binding LacI-Venus and thus the total number of repressor molecules is decreased. While the number of bound molecules remains at one dimer per operator site, the background fluorescence is directly dependent on the number of freely diffusing molecules. Thus the auto-repression increases the contrast: *i.e.* while the bound protein will always create a near-diffraction limited spot, it will not be seen unless the variation in background is reduced below the intensity of the single molecule.

We used exposure times of 4 s based on the result on pseudo-specific binding (paper III), which is discussed in the section “Sliding on DNA” on p. 47 and shown in (fig. 13). In this way false-positive detection of non-specifically bound proteins is minimized. In general the exposure time is a

trade-off between reduction of non-specific binders on one hand, and optimal signal from specific binding together with sufficient time-resolution on the other hand.

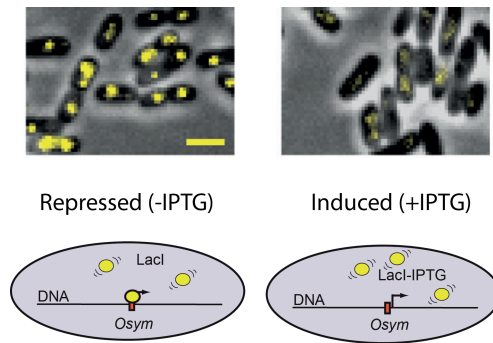


Figure 8. Binding to a single artificial operator site. (*left*) without induction LacI-Venus binds the operator and near-diffraction limited spots are seen over a 4 s exposure, (*right*) when cells are induced the spots disappear.

Kinetics of binding

The association rate measurements were done in the following way. Cells grown in liquid culture were induced with IPTG, collected by centrifugation and immobilized on poly-L-Lysine dishes. At time zero the sample was diluted in 67 volumes of the same medium, without IPTG and supplemented with the anti-inducer ONPF at 1 mM (fig. 8). This means that the final IPTG-concentration is on the order of 4 μ M. This could in principle be enough to reduce pseudo-operator binding, however at the same time 1 mM ONPF competes with IPTG for LacI binding. See figure (fig. 9A-B) for the effective competing concentrations of IPTG and ONPF. The amount of inducer needed is higher for *lacOsym* (fig. 9C) than for *lacO1* (60). To obtain sufficient LacI-*lacOsym* dissociation, we use 300 μ M as starting concentration of IPTG, compared to 100 μ M used in the previous work (60).

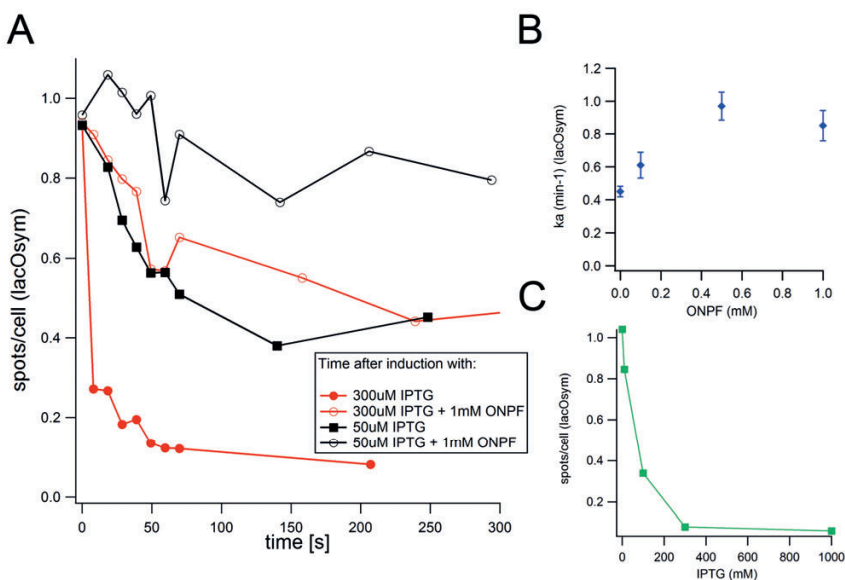


Figure 9. Induction response (additional data). **(A)** IPTG and ONPF are added simultaneously at different concentrations; at 50 μM IPTG and 1 mM ONPF, induction by IPTG is close to blocked, which confirms that the final concentration of 4 μM IPTG in the kinetic experiments should have a minor effect on LacI binding when ONPF is present, **(B)** binding kinetics as a function of ONPF (when starting at 300 μM IPTG); the maximal 2-fold difference with and without ONPF confirms its importance in reducing IPTG wash-out as a rate-limiting step. No additional effect is seen beyond 500 μM , implying that the LacI-ONPF binding is saturated, **(C)** titration of the LacI-*lacOsym* IPTG-response at steady state (>1 min after addition of inducer).

Using $\mu\text{Manager}$ the position list was predefined so that following the start of the acquisition all positions could be automatically imaged at a 10 s interval between frames. At each position we acquired fluorescence images with 4 s exposure ($15\text{W}/\text{cm}^2$) at 514 nm followed by a phase contrast image of the same region of interest. As described in the methods section of paper I, the readout is the total number of automatically detected fluorescent spots over the total number of manually counted cells to give the measure of fractional binding. Total duration of the experiment was set to fully cover the whole association time course. The reason for using different positions for different time points is to avoid bleaching of the fluorophore. At the optimized intensity/exposure time used, most Venus proteins appear to be photo-bleached after 2-3 consecutive acquisitions of the same region of interest, and thus the same cells cannot be used to study the increase in protein-binding. Given that all positions used are located within an area of about 1 mm^2 , and simultaneously diluted with a large excess in volume, every position is assumed to be equally treated and comparable. Each image contains 50-500 cells (typi-

cally 200-300) and thus cell-cell variations (for example IPTG/ONPF susceptibility or spot quality) should be averaged out.

From this we end up with association data that are fitted to a simple exponential function to quantify the rate of binding:

$$f(t) = a(1 - be^{-kt}) \quad (8)$$

As shown in paper I, the binding rate to a single *lacOsym* per *E. coli* chromosome in room temperature is $0.74 \pm 0.02 \text{ min}^{-1}$, which is a product of the actual association rate constant and the unknown cellular concentration of *lac* repressor ([TF]). The concentration of the protein could be estimated from old data (8, 60), accounting for that our LacI-Venus is a dimer and its expression auto-repressed, and using visual counting of fluorescent fusion proteins (87). From this we estimate that we have 3-5 *visible* LacI-Venus dimers per cell which gives us the association rate constant of $k_a = 2-3 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$. The early *in vitro* measurements showed a more than thousand-fold higher rate (58). However, a modified expression for the diffusion-controlled non-specific association rate, which accounts for the geometry and compaction of the chromosome, has been derived (74). Also considering the slowdown due to 3D crowding, the upper limit is estimated to $k_{assoc} = 1.5 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ (91). This is how fast the repressor would find and bind an operator site through random and repeated trial-and-error. Our measured rate is just above this and what it means is further discussed in the section “Sliding on DNA” (p. 45).

Difference between operator sequences

The operator *lacOsym* corresponds to the first 10 bp of wt *lacOI* which is then mirrored to give a symmetrical sequence. This artificial operator was shown to have stronger LacI affinity than *lacOI in vitro* (75). Later on, I will discuss how the dissociation rates differ between operator sequences but at this stage the focus is on their effect on association times.

The following experiment was designed exactly as for the case of *lacOsym*, although now with the chromosome containing *lacOI* instead. The measurement showed that LacI binds to a single operator at a rate of $0.60 \pm 0.03 \text{ min}^{-1}$ (fig. 10). This means that it takes about 20-30% longer for the same number of repressors to bind at this operator compared to at *lacOsym*. It is discussed in the section “Sliding on DNA” (p. 44) how this difference could relate to operator recognition (probability of binding). At this point I want to emphasize that the difference is measurable and significant.

An indirect way of comparing association rates to different operator sequences is to study the association to a DNA segment containing several operators. It should here be remembered that the fusion proteins do not form tetramers. Contributions from two-domain associations, such as non-specific

intersegment transfer and specific DNA looping, can therefore be neglected in the analysis.

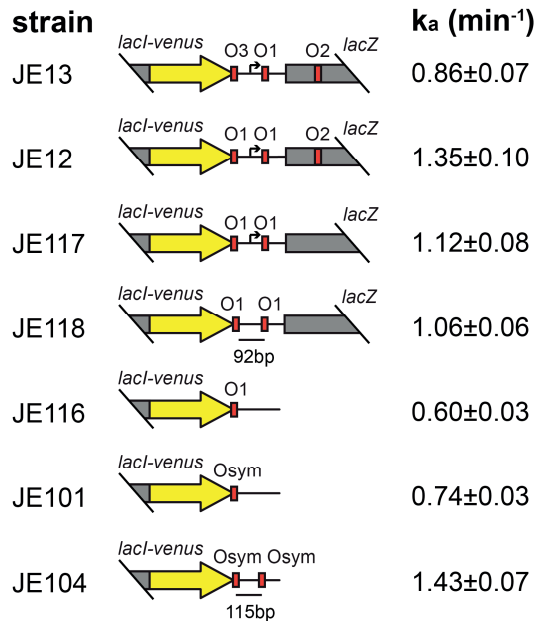


Figure 10. Association rates in the multi-operator scheme. The experiments and analysis are described in paper I.

We measured the association rates in a strain containing the wild type configuration of operators (JE13), in a strain where the weak *lacO3* was replaced by a second *lacO1* (JE12), in a strain where *lacO2* was removed from JE12, and in a strain where the *lac* promoter was deleted. The last strain is especially important as reference since binding to single operator sites as described above were measured in strains where the promoter and *lacZ* gene were deleted and hence no interfering RNAP activity could influence the result. These experiments were conducted in the same way as before, however with two important differences. First, while the change from *lacOsym* to *lacO1* has a small effect on the LacI auto-repression and consequently the repressor concentration, the presence of *lacO3* in this position considerably reduce the same effect. As a consequence the repressor concentration increases in the *lacO3*-strain with two subsequent effects. One is that more repressors should mean faster operator binding. We account for this by measuring the difference in total fluorescence in JE12 and JE13 after subtracting the cellular background and conclude that the latter contains about 1.7 times more LacI-Venus proteins. Another effect is that due to the in-

creased fluorescent background, image analysis (spot detection) becomes more difficult.

Several operators placed within the diffraction limit will produce only one localized spot. The distance between *lacO3* and *lacO2* (where *lacO1* is in between in wt and JE13) is ~500 bp which equals about 170 nm on a stretched piece of DNA. While this maximal distance in principle would give an outstretched fluorescent spot when *lacO3* and *lacO2* are simultaneously occupied, they could not be resolved as two molecules with our analysis tools. This means that our readout is always binding to the operator region, no matter if there are one, two or three operators in this region.

From the association rates measured (fig. 10) we create an additive model that includes the single-operator measurements. This reveals that binding to *lacO1* is two times faster than binding to *lacO2* and ten times faster compared to *lacO3*, which fits nicely with *in vitro* data (48). The numbers should still be taken cautiously given the indirect way we measure. But they provide an idea of the situation. We also see that there is no obvious effect if the promoter and CRP binding sites are intact or not, suggesting that they are bound only a small fraction of the time under our experimental conditions.

The difference in binding time to different sequences can have several implications. One is the probability of binding which is discussed in the section “Sliding on DNA” below; another is binding to the wrong operator. The wild type repressor is a tetramer and has been shown to loop DNA by co-binding of two operators (49). Therefore primary binding to *lacO3* or *lacO2* might be followed by transfer to *lacO1*. Usually this is discussed in terms of dissociation rates, so that the repressor when bound, can dissociate from one operator and have time to re-associate before dissociation from the other operator, and this is modeled as an increase in local concentration (38). However, as was described in the introduction, faster binding to one of several operators would indicate that not only the time bound but also the association rate is increased if looping is much faster than the primary association (51).

Temperature dependence

Later, dissociation measurements are presented. These were done at varying temperatures and using microfluidics. Except for the growth environment, a difference in microfluidics is that the medium is completely changed within a few seconds and that IPTG is likely to be washed out on this time scale. In addition image analysis is done in a fully automated way. Despite these changes, the association rate to *lacO1* differ (increase) no more than 20% compared to when assayed on the poly-L-Lysine dishes.

When association is measured at 37°C (paper II) the rate is ~2.5 times faster than at 25°C. A further increase to 42°C has a less pronounced effect. This effect has also been seen *in vitro* (58, 65). From fluorescence microscopy images, LacI-Venus concentrations seem to be similar at different tem-

peratures. The diffusion coefficient should not change more than a few percent by this temperature shift (2), it is however not obvious how the indirect effect on viscosity propagates. The explanation could instead be that the non-specific DNA-binding of LacI and of all other proteins changes, together with a structural rearrangement of the chromosome. This is further discussed in the next section.

Sliding on DNA

Since the facilitated diffusion model was first presented (74), many scientists have spent their days producing more detailed models, designing intricate computer simulations, and collecting data through the lab hours of their students. In addition biochemical experiments have been conducted to provide more details on the impact from salt effects, DNA structure, and in the case of the *lac* repressor dissecting the property of in principle its every single amino acid. At the end of the day, the early models and data stand intact.

When I now turn the page to be yet another one of these students to look for 1D diffusion; what do I expect to find? More importantly; what do I expect to find, that is new?

The main thing is that 1D diffusion or sliding has not been shown to exist in living cells. It could be that this elegant theory is easily verified outside cells by using appropriate experimental conditions. Since the early experiments on TFs (66), sliding has been demonstrated *in vitro* for restriction enzymes (92), DNA-repair proteins (93) and even the tumor suppressor protein p53 (94). However the conditions that resemble the cellular environment are not easily set up. Therefore the primary question here is: does 1D diffusion or sliding occur in living *E. coli*, and if so, can we measure the average sliding length?

Inside and outside the cell

What are the *in vivo* conditions that the biochemist would have to mimic? The following are likely the most important factors, although not necessarily in order of appearance.

Salt effects

Already Riggs and co-workers observed that increasing the concentration of monovalent ions decreased the rate of binding specific DNA (58). Other groups that have observed sliding *in vitro* have used various salt concentrations and presented different sliding lengths for different proteins, *i.e.* from no sliding (45), to the typical range 50-200 bp (66, 95-97) and up to nearly 1000 bp (98). As was described in the introduction, it is evident that higher salt lowers the protein-affinity for non-specific DNA (*i.e.* more cations in-

creases the shielding of the negatively charged DNA, from the positively charged DNA-binding surface on the protein), so that the molecule stays bound for shorter time and essentially cannot slide as long a distance before dissociation. At the same time, as shown previously (58), the specific association rate is heavily reduced. To make it even more complicated, divalent ions (normally supplemented as $MgCl_2$) competes with the repressor for DNA binding. Obviously it will be indispensable to do experiments using the intracellular salt levels, or even worse, the levels that produce the same conditions as is effective *in vivo*.

The intracellular salt is believed to be roughly represented by a mix of 150-250 mM monovalent (59) and a low concentration of divalent ions. DNA association becomes weak at these concentrations and reliable sliding length measurements are difficult to accomplish. This is why long sliding lengths at low salt are commonly published numbers.

The chromosome

In addition to salt, the DNA itself is an important player. In an *in vitro* experiment where an operator sequence of 20 bp is placed within a 100 bp segment of fragmented DNA, the protein will have essentially 1 out of 80 possibilities to find the right place (excluding 10 bp from each side). If the segment is increased to 10 000 bp, the ratio becomes 1/10 000; in the *E. coli* chromosome this number is 1/5 000 000 if there is one operator per chromosome. Varying the length of the fragmented DNA is a neat way of demonstrating 1D diffusion *in vitro*. However when accounting for that the searching protein might be trapped by non-specific interactions in the wrong region (66), this extreme ratio on the chromosome is of importance. For similar reasons, the coiled and condensed chromosomal nucleoid must be treated differently than linearized DNA fragments commonly found in the test tube (74).

Crowding

The 3D diffusion limit is directly affected by obstacles in the solution. The more macromolecules diffusing around in a solution (or cytoplasm), the more clashes and interactions the searching protein will experience, and its rate of diffusion will be affected in a way that can be modeled by increased viscosity. Macromolecules that occupy the chromosome might also act both as stationary roadblocks for a diffusing protein, or simply cover part of the DNA, not allowing it to be searched. If these roadblocks also are sliding it will lead to sub-diffusion. Indeed, the fraction DNA occupied is estimated to be on the order of 0.3 in the living *E. coli* (51). At the *in vitro* conditions where measurements are done, occupancy numbers close to the *in vivo* situation are unlikely. This chromosomal crowding might reduce the amount of DNA needed to search through. The DNA-bound proteins could even serve as specific roadblocks in terms of regulating the rate of TF binding. In addi-

tion, if they block the path of the sliding protein, then looping or other mechanisms might be required for bypassing.

The length of LacI sliding on DNA

When I started this work the aim was to show that sliding is taking place in living cells. One year later, the question was: is sliding taking place in living cells? The experiment was designed and outlined as in fig. 11. Similar experiments were previously done *in vitro* by (96) and (49), although with other methods for detecting DNA binding.

The assay

Imagine a TF binding to a single operator site (LacI and *lacO_{sym}*) with the association rate constant as measured in room temperature. If a second, identical, operator is placed anywhere on the chromosome the rate of binding to anyone of these sites should be double that of binding a single operator; the time of association should be halved. This is the case if there is no 1D-diffusion, *i.e.* if the TF always binds directly in a one-step reaction. However this is also the outcome if the distance covered by a sliding TF, diffusing along the DNA, is much shorter than the distance between the two operator sites. With 1D-diffusion as part of the search mechanism one can see the operator site as a binding target with a size that increases from the 1 bp precision to about the length covered by the TF before it dissociates away from the DNA. So, if the binding targets are farther apart than the sliding length, then the association time to the first of them should be halved, but if instead they are moved closer, the binding regions start to overlap and the total binding target will not be increased by a factor of two any longer. This relation between sliding length and inter-operator distance is described in paper I.

We measured association rates in two-operator strains with a distance between operators ranging from 25 to 203 bp (fig. 11B-C). From the fitting to the simple most model, where association rates are plotted over operator distance, we have that the rms sliding length is 36 ± 6 bp. In the two-step reaction scheme (eq. 5) this means that the repressor first associates to any available DNA sequence in the chromosome and then scans on average 30-40 bp before dissociating (or binding if the operator center is within the sliding length). Conclusion: 1D sliding seems to exist *in vivo*, but we cannot know from this the details of how it explores these 30-40 bp.

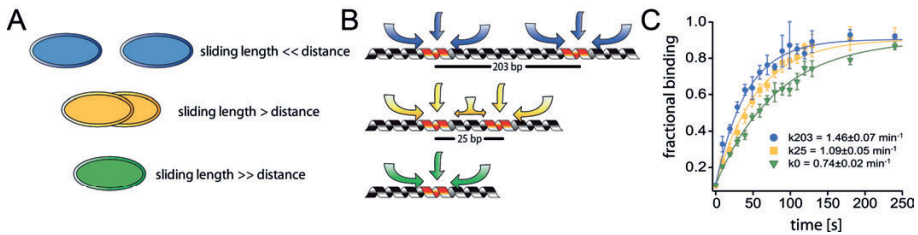


Figure 11. Sliding length assay. **(A)** illustration of the relation between operator distance and sliding length, **(B)** design of the experiment based on the model in A, **(C)** association curves for LacI-Venus and *lacOsym* in constructs described in B.

It is important to note that despite having two operator sites we measure binding to the first of the two sites. The reason is that the two sites are well within the diffraction limit (200 bp on stretched DNA is ~ 60 nm). Note again that there is no LacI tetramer and no looping of DNA in the experiments. Also, the transcriptional unit (*i.e.* CRP-cAMP and RNAP binding) was removed as before, to study as clean a system as possible. It is however seen (fig. 10) that association rates are not reduced in strains with the promoter region intact. This implies that sliding in to and binding at the operator is not affected by transcription at the promoter.

Introduced roadblocks as physical barriers

Although we can measure the average sliding length, it remains to elucidate how this sliding mechanism is actually conducted. We do not know if the TF explores 30-40 bp by making small hops (microscopic dissociation and re-binding), by linear diffusion on the DNA surface, or by sliding along the contour of the helical structure. Additional experiments are needed and we start by asking: can the sliding TF bypass another protein in its way, or is the path obstructed?

LacI can bind close to TetR

With two operator sites for different TFs “overlapping” and sharing on base pair (the last flanking bp in *lacOsym* is the first bp in *tetO₂*) we hypothesized that simultaneous binding could not occur. This turned out to be wrong. As shown in paper I, LacI can bind next to TetR in a situation like this. Overexpression and binding of TetR does not to a significant degree change the number of bound LacI detected, and it does not affect the ability of LacI to repress the *lac* operon. The combined structures of LacI and TetR binding to their operator sites, as in our experiments, display a spacing of ~ 1 nm between the TFs. This is enough to accommodate ~ 3 water molecules (fig. 2). These findings can be related to a paper showing that the central 19 bps of *lacO₁*, and maybe even the central 17 bps (out of 21), are enough for maxi-

mal binding (99). This suggests that the base pairs at the outskirts of the operator are unoccupied although they are evolutionary conserved. The same assumption can be made for the TetR-*tetO*₂ interaction, which seems to be limited to the central 15 (of in total 19 defined) bps of the operator sequence (40). In addition the binding constant between the *tet* repressor and a single operator is reported to $\sim 10^{11} \text{M}^{-1}$ (40). This means that already at the leakage level from the plasmid should the TetR available saturate the operator site. This is complemented in our experiments with a strong induction of the P_{BAD} promoter and as seen in the β -galactosidase experiment (paper I, fig. S14), TetR represses the *lac* promoter very efficiently.

TetR blocks sliding

The effect of the bound TetR on the LacI-*lacO*_{sym} association rate is a factor of 1.75 ± 0.18 (fig.12). What did we expect? If the repressor actually finds the operator through 1D diffusion as shown by the previous experiment, then the roadblock effect should be at most a factor of 2; *i.e.* by blocking from one side. There are two options. The first is that there is a complete block and that we have a measurement error. After all, by accounting for the error bars the result is close to 2. One explanation could be that TetR displays a lower than expected fraction of time bound. Our control experiments, however, argue against this.

The other option is to believe in the measured value, which is significantly below 2. What is then the explanation? Does the repressor occasionally slide past the roadblock? One could suggest that bypassing of roadblocks is a necessity for the search process to work out. If this is the case, then it not only has to manage to get around TetR, it also needs to find the correct binding configuration once the overtaking is completed. The same is the case if the repressor can hop over or around the roadblock. At what distance from the DNA would a hop instead lead to complete dissociation? I will leave this question open for a while and try to answer it in relation to paper III.

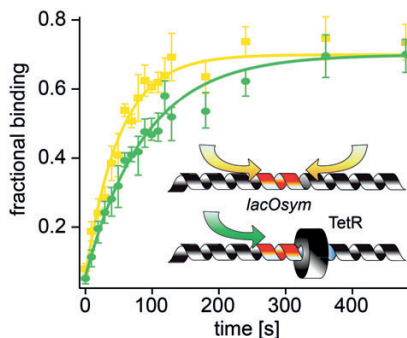


Figure 12. Roadblock experiment. Binding to *lacO*_{sym} is 1.75 times slower when a TetR roadblock can bind adjacent to the LacI-Venus binding site.

Probability of binding the operator

Is it possible that the repressor spends all that time trying to find the operator, and when finally sliding over it, does not bind? Based on the finding that binding to *lacOsym* is faster than binding to *lacOI* (which is also seen for two operators of each kind); this is what we suggest. Since the only difference between the strains is in the operator sequence the probability (or rate) of finding the site must be the same, and only the probability to bind once it has been found can be different.

It means that the previous model (74) in which it is assumed that the repressor always binds, is extended. How the overall association rate (eq. 2) with a finite probability of binding (P_{bind}) is found, is described in paper I (Eq. S1-S5). By assuming that the maximal probability of binding *lacOsym* is 1, and using the measured association rates and sliding length, the maximal probability of binding *lacOI* is found to be 0.12. Since the repressor diffuses back and forth many times before dissociation, the total probability of binding the operator once the operator-containing DNA-segment is reached, and before leaving to any uncorrelated DNA-region, is 0.77. We move on to test this hypothesis by introducing binding sites for roadblocks on both sides of *lacOsym*.

Two-sided roadblocks

In these experiments the auto-repression of LacI was removed in order to position a *tetO₂* operator also upstream of the operator. This increases the fluorescent background as well as the probability of detecting binding to a pseudo-specific site (*i.e.* as described on p. 47). We therefore measured the dynamics going from IPTG to ONPF also in a strain without specific operator, made an approximate fit to the data and subtracted this from the other binding curves. This shows that taking the pseudo-specific binding into account is not only possible but also important. The effect from the roadblocks on LacI binding is obvious. However again the effect is smaller than expected for a 30-40 bp sliding length and the possible explanations for this are the same as for the one-sided roadblock. We then challenge this in the new model containing the probability of binding the operator. The complete analysis is shown in paper I (fig. S13). The data fit the model well and show that the total probability of binding *lacOsym* is in range [0.38, 1] and the rms sliding length is 45 ± 10 bp. As before the corresponding values can be calculated for *lacOI* to be $P_{tot} = [0.29, 0.77]$. When the data for the one-sided roadblock is added and the system is fitted again the result is the same. The reason is that the factor 1.75 fits perfectly with the already found $S_L = 45$ bp and $P_{bind} \sim 0.1$.

Introducing this model shows that the sliding length first presented is a slight underestimate. It also suggests that the one-sided roadblock experiment is valid; the roadblock is there most of the time, sliding from that side

is completely blocked, and the searching repressor does not bypass the roadblock by hopping. The idea that the TF not always binds when sliding over the operator explains why its association rate is higher to *lacO_{sym}* than to *lacO₁* (and why binding to *O₂* and *O₃* is even slower). This should be connected to earlier descriptions of the protein-DNA interactions presented in the introduction. If the repressor has more than one binding conformation (66) it might explain why it sometimes does not bind when passing over its target. The benefit of this could be that compared to with having a single binding conformation it is also less likely to get trapped at other, non-specific and especially pseudo-specific sites.

The connection between roadblocks and the probability of binding is the following. The block from one side gives rise to an association rate reduction of a factor of 2. Sliding events from the open side, which reaches the operator without binding, is hindered from sliding further in the same direction due to the roadblock. Hence the sliding repressor has an increased chance of binding the operator because of the roadblock. In the case of double-sided roadblocks the effect is even more pronounced. For low probability of binding there is no influence on the rate from the roadblocks since the small target size is perfectly compensated by an increased number of return events before leaving the DNA segment. From this reasoning, the observed association rate to *lacO₁*, with a lower probability for the repressor to bind, should display a smaller roadblock effect.

We made a final test. In a strain containing two *lacO₁* operators, with roadblocks in between, the rate reduction for the given S_L and $P_{bind}(lacO_1)$ is predicted to be ~ 1.54 (compare 1.75 for *lacO_{sym}*) per operator site. We measure 1.59 ± 0.10 .

And what about the diffusion limit?

Let us ask the never-ending question: does the *lac* repressor find its operator-site faster than a diffusion-controlled reaction? We show in paper I that the rate of binding is on the order of this limit. However when accounting for that non-specific DNA-interactions slow down the rate (~ 10 times), then the measured sliding length is valuable for explaining the rate we obtain. In summary, LacI binds faster than a random search for a target of a size of one bp could explain.

Mode of sliding

I have already mentioned that our assay cannot resolve the microscopic behavior and the interactions behind the actual movement of the TF over DNA. The experiments using roadblocks suggest that the protein stays close enough to the DNA to be hindered. It is difficult to rule out if the microscopic hops that are suggested as a mode of translocation in a crowded environ-

ment (92) contribute to the facilitated diffusion. If a hop is so small (66) that it is experimentally indistinguishable from a short 1D diffusion step, it might be considered as such a step.

In addition; is 1D diffusion a linear translocation process where the protein diffuses on one surface of the DNA helix? This was demonstrated *in vitro* and also suggested as a way to evade obstacles *in vivo* (100). The opposite idea is that sliding is rotation-coupled, which means that the protein diffuses along the helical structure while keeping its DNA-binding domain in a fix orientation relative the axis. This was investigated analytically, and used to explain why the 1D compared to 3D diffusion coefficients are so small (73). The latter model was recently shown to fit with diffusion coefficients for a range of proteins (101).

A helical path suggested

It is difficult in our two-operator sliding length assay to rule out which is the most likely mode of microscopic diffusion. In contrast, since the results from the roadblock experiments suggest that there is no bypassing of obstacles it might favor a model where LacI is in close contact with DNA. In paper III, molecular dynamics (MD) simulations are used to study this in depth. Starting with an experimentally-based structure configuration of the LacI dimer non-specifically bound to DNA, the repressor is pulled slowly (1 bp in 100 ns and then kept still for another 100 ns) in the axial direction with no angular force applied. As opposed to a linear movement on the surface along the DNA, it clearly and spontaneously follows the helical path as defined by the backbone. It is also seen that the center-of-mass of the protein remains at the same distance from the DNA axis during a one bp translocation.

Based on the observation that the repressor moves in a helical path, umbrella sampling simulations are then employed to calculate the free energies of moving the protein both along the helix and out from (leaving) the DNA. It shows that the roughness (101) for rotation-coupled sliding is ~ 1 k_bT . All together this suggests that LacI stays in close contact with the DNA while sliding. In addition radial pulling shows that the energy cost for dissociation is large. The distance from the DNA at which the repressor undergoes complete macroscopic dissociation seems to be smaller than what is needed to pass over for example a TetR roadblock. Therefore hopping is not likely to help the protein traverse a crowded DNA.

The k_{diss} described in eq. 5 is a macroscopic dissociation rate constant. During the residence time given by k_{diss} the non-specifically bound protein dissociates and rebinds microscopically. This essentially means that most hydrogen bonds are broken such that the protein leaves the DNA, yet stays within the reaction radius and thus has a high probability to rebind to the same base pairs just left. From the simulations the microscopic residence time is estimated to 45 μs and k_{diss} , as a consequence of this, becomes $225s^{-1}$. This *in vitro* estimate gives a macroscopic residence time (~ 4.5 ms) close to

the upper bound measured *in vivo* (60) and smaller than what we measure here. From this the sliding length is calculated to be ~ 70 bp. The corresponding value from *in vitro* experiments is ~ 150 bp (66), which supports the validity of the simulations and interpretations. In addition our *in vivo* measurement of s is 32 bp. The result that the repressor rebinds with high probability following a microscopic dissociation event implies that the degree of diffusion-control for the macroscopic association rate constant is very high (in paper III described as $\alpha \gg 1$). This means that the rate of reaching the surface on the DNA (limited by diffusion) also is the rate of binding non-specifically. The dissociation and rebinding events are so short in both distance and time, that they will not contribute significantly to the diffusion along DNA and hence using eq. 4 to describe sliding holds. With this description the 1D diffusion will consist of a few bp rotation-coupled sliding (as indicated by the MD simulations) at a time, and separated by microscopic dissociation and rebinding events to the same position on the DNA (*i.e.* non-productive hopping).

Pseudo-specific binding

The single-molecule experiment presented in paper III was initiated in the following way. During the cloning procedure of other strains, a strain with LacI-Venus and no operator sites was created as an intermediate step. This increases the concentration of LacI due to the removal of auto-repression. I decided to look at this under the microscope. Surprisingly, at a 1-s exposure, there was a significant amount of fluorescent spots seen in the cells. Clearly less than in strains with operator sites, but they were there. I added IPTG, the spots disappeared. We made two new (very similar) strains and the result was the same. Based on the idea of detection through localization, this implies that what we see are proteins that are bound on the time scale of one second or longer. The induction response indicates that this is DNA-binding. Since the non-specific binding of LacI to DNA is unaffected by IPTG-induction (61), what we see is likely the interactions with operator-like sequences (45) which is referred to as pseudo-operators.

Implications

The impact of pseudo-operators on the specific-operator search kinetics will depend on the unknown binding constants and concentrations of these sites. They will sequester proteins for part of the search time and can of course be considered to be part of the intracellular complexity in general.

In paper III we show the relative amount of spots per cell detected in the same strain, as a function of exposure times (fig. 13). When the exposure times were decreased, the laser effect was increased accordingly, to obtain the same amount of exposure per frame. At 10 ms and below we approach the upper limit of the laser ($\sim 1.2 \text{ kW/cm}^2$), and the EM gain of the camera

was therefore increased to produce a comparable signal. Due to mobility of cells and their chromosomes, spot sizes increases with exposure times, and it is difficult to choose analysis criteria that are equally reliable for all exposure times. Therefore, there might be an underestimate in the absolute differences, especially between the longest and shortest exposure times. Nevertheless the graph illustrates how non-specific and pseudo-site binding decreases over longer exposure times, *i.e.* their interaction times are too short to be observed. Assuming that IPTG do not affect non-specific binding (61), the fraction that is removed with inducer should be a measure of pseudo-operator binding.

Methodological concerns

The presence of these fluorescent spots influences our method; some of the detected binding events in the operator-containing strains will also be at pseudo-sites. This is why we use as long as 4 s exposure times in the kinetic experiments. Also, since the pseudo-site binding is IPTG-affected, it is not as simple as subtracting a flat baseline. This issue is further discussed together with the “two-sided roadblock” experiments above (p. 44).

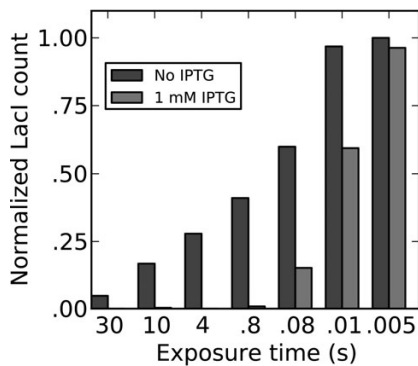


Figure 13. Pseudo-specific and non-specific binding. Detection of LacI-Venus as a function of exposure times and \pm inducer, in a strain lacking *lac* operators. Since the auto-repression is lost when the operator is removed, the LacI-Venus concentration increases and so does the detected non- and pseudo-specific interactions (paper III).

Temperature dependent DNA interactions

We present in paper II that the repressor binds to *lacO1* \sim 2.5 times faster at 37°C compared to at 25°C. If the diffusion coefficients are only marginally affected by the temperature, both directly, and indirectly through viscosity (2, 70) then the explanation is to find somewhere else.

It was shown also *in vitro* that specific binding is faster at higher temperatures (65). Other experiments indicate that non-specific binding is weaker at

higher temperature (64). This implies that k_{diss} increases and that the sliding length decreases accordingly (if the 1D diffusion coefficient is unchanged) (eq. 4). It was suggested that weaker non-specific binding and faster specific binding is explained by that the protein spends less time trapped at non-specific DNA locations (65).

Roadblock effect

In paper II we use roadblocks next to *lacO_{sym}* to show that association and dissociation is affected in the same way (the experiment is described in the section “Dissociation rates”, p. 53). This is also suggested by the equilibrium binding data in paper I; the repression efficiency is unaffected by the presence of a roadblock. Together this implies that the repressor slides into and binds the operator, and slides out and dissociates from the operator in the corresponding way.

However, the experiments show something more. The roadblock effect (fig. 14) is not 1.75 as was measured at 25°C (in paper I) but ~1.3 for these data collected at 37°C. Since the effect from roadblocks is smaller with decreased sliding length (especially when the probability of binding <1), this result indicates that the sliding length decreases with higher temperature.

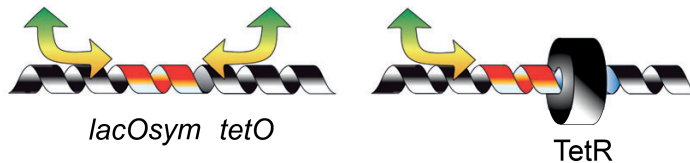


Figure 14. The effect of one-sided roadblocks. (*left*) without roadblocks, LacI can slide in to and out from the operator without obstacles, (*right*) when a roadblock is added sliding is blocked in the same way in both directions.

What does this mean?

Even though the roadblock experiment indicates a shorter sliding length, and thus less competition from non-specific DNA at higher temperature, it is difficult to fit with that the absolute association rate (which is enhanced by sliding) for the operator is faster. A related implication is that if the non-specific binding is weakened, then the fraction of time the protein is non-specifically bound should decrease. This fraction was measured to be 0.9 at room temperature (59, 60). In addition the optimal fraction, in terms of balancing the slow-down and speed-up from non-specific DNA-interactions for obtaining a maximal operator association rate, was estimated to be 0.5 (74, 102). Finally, our suggestion that the probability of binding is <1 might shift this optimality since several visits at the operator site is required for binding. We show in paper I that the fraction that best fits our room temperature-data is 0.7. At the same time this number was also suggested by (103).

There are thus several reasons to investigate this further under our experimental conditions.

Tracking of free LacI-Venus

If the fraction of proteins bound changes with temperature this should be detected in an assay that measures the non-specific DNA-interactions.

Methods for tracking of free proteins in living cells have been implemented in the lab by my colleagues. The diffusion coefficient of a freely moving fluorescent protein was measured by (69). Here, the fluorescent tag (protein) is photoconvertible: when cells are exposed with a violet laser, proteins are converted such that single-molecule excitation can be done with a longer wavelength laser. The protein is highly expressed so that many subsequent trajectories can be collected from the same cell (104, 105).

Neither a mechanical shutter nor the shutter of the detector is suitable for the fast frame rates needed. Tracking is made possible using the principle of stroboscopic illumination (60, 106). An acousto-optical modulator (AOM) shutters the incoming light and is triggered by the EMCCD camera so that imaging is synchronized with excitation of the cells. The camera collects images at a given frame rate and the exposing light is pulsed to hit the sample with the same rate, in the middle of each imaging frame. The AOM defines the exposure time and the camera sets the frame rate. The experiment is used to track thousands of individual molecules in a single cell, where the trajectory length is limited by the bleaching of the fluorophore. Diffusion coefficients are calculated from the MSD of each trajectory as defined by the displacement of spots from frame to frame.

This approach was further developed by introducing a new method for analyzing thousands of short single molecule trajectories (107). Tracking and spot detection is done in a similar way, although now the diffusion coefficients are calculated from single hops between two spots in subsequent frames. This is then used to look for different diffusive states using a statistical model; the Hidden Markov Model (HMM). In this analysis memoryless jumps between hidden diffusive states are assumed, and the algorithm figures out which is the most likely number of hidden states and transition rates, based on the diffusive jumps that can be observed.

The advantage of tracking a highly expressed protein is that a huge amount of data can be sampled from a few cells. The disadvantage is also that the data is sampled from a few cells. We combine the techniques of stroboscopic illumination and HMM state-analysis, with our high-throughput microfluidic imaging setup. In this way many cells can be imaged simultaneously. Low copy, chromosomally expressed proteins can be studied and still a sufficient number of trajectories can be collected. Instead of many trajectories from a few cells, there are few trajectories each from thousands of cells, and at a biologically relevant copy number. This means that we

potentially avoid the saturation of binding sites, which otherwise could mask the effects that we are looking for.

A rectangular slit is set in place of the last iris in the optical setup and this allows us to divide each trap in the microfluidic chip into 4-5 sections which can be imaged independently. This means that up to ~ 200 movies can be taken from a single chip, only exposing the cells once. Each movie is 5-10 s long and all positions can be completed in two hours, resulting in more than 50000 trajectories. Since living cells are studied, the same field of views can be re-used after a few hours recovery from photo-bleaching. We can therefore use the “same” cells to study the effect of different medium and temperature. This experiment is not described in paper I-III and is here presented as an additional, preliminary result. Trajectories were collected with a frame rate of $\sim 250 \text{ s}^{-1}$ and 1 ms exposures in each frame. The data was analyzed by Fredrik Persson as described in (107).

A strain expressing LacI-Venus and lacking a specific operator (JE107) was imaged to study non-specific binding, and a strain expressing a truncated LacI-Venus, lacking the DNA-binding domain (MG1655 Δ lac *atpI::lacI42-Venus*) (108) was used as reference. An initial exposure was done to pre-bleach the cells. With too many fluorophores nothing will be detected since different molecules cannot be separated. This highlights an apparent limitation of this assay. In fig. 15A cells are overlaid with the trajectories from one movie and shows that we collect intracellular diffusive steps while excluding intercellular jumps.

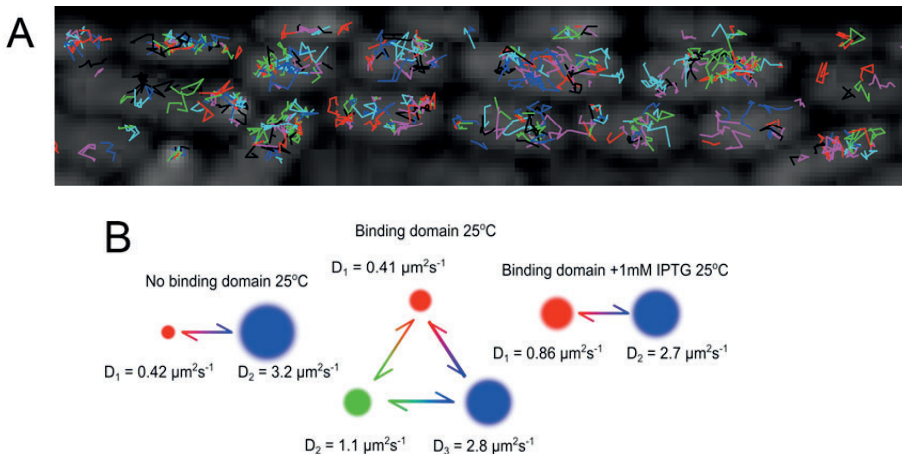


Figure 15. Tracking LacI-Venus, (A) overlay of cells and trajectories from a single movie, (B) comparing diffusive states (*left*) without and (*center*) with the DNA binding domain intact, and (*right*) with IPTG-induction and the binding domain intact. The area of the circles is proportional to the fraction of trajectories in each state.

Results

The result presented in fig. 15B is based on the analysis of ~ 15000 trajectories per condition.

At 25°C and without IPTG-induction (fig. 15B, center) we get 3 distinct states from the HMM analysis. This could represent free diffusion (fast), pseudo-specific binding (slow), and an intermediate state which is an average of molecules interacting differently by non-specific interactions with DNA. When IPTG is added (fig. 15B, right) the slow state disappears, however the intermediate state also becomes slower which implies that some of the slow trajectories end up there. This somehow fits with the data in paper III that pseudo-specific binding and not non-specific binding, is neutralized by the inducer. The fraction of trajectories in the fast state increases when the temperature is shifted to 36°C (from ~ 0.6 to 0.8 , data not shown); the intermediate state disappears and at the same times the diffusion coefficients of the slow and fast states remain unchanged. This implies that [1] the rate of diffusion is mildly affected by the change in temperature (also via viscosity), and [2] the fraction of proteins non-specifically bound is decreased. When the DNA-binding domain is removed from LacI almost all trajectories are found in the fast state at 25°C (fig. 15B, left). The obvious reason is that the DNA-interactions are lost. Little is changed when the temperature is increased (data not shown). This is expected since all (most) proteins are freely moving.

The results should be taken cautiously. Even though the conclusions made are reasonable from a descriptive point of view, the absolute meaning of each state is not clear. Also the transition rates between states, as described in (107), are here a bit ambiguous since they are on the same time scale as the frame rate of the camera and thus difficult to catch. Hence they are not presented.

Fraction of proteins bound

Using Fluorescence correlation spectroscopy (FCS), freely diffusing LacI-Venus without DNA-binding domain was measured to have a $D_3 \sim 3 \mu\text{m}^2 \text{s}^{-1}$ (60). In the same study the fraction of proteins associated to DNA was found to be 0.87 . If in our experiments the average diffusion coefficient of all trajectories (D_{eff}) at 25°C and \pm binding domain is calculated, then D_{eff} (binding domain) / D_{eff} (no binding domain) is an estimate of the fraction of proteins free. From this data it gives 0.67 and implies that 0.33 are non-specifically bound; a dramatically lower number than the previous estimates (59). When the same experiment (not shown in figure) and calculation is made for 36°C the fractions are 0.84 and 0.16 . This confirms the hypothesis that motivated the experiment; that the fraction bound is decreased when temperature is increased. The absolute numbers are however surprising since they significantly diverges from previous measurements and calculations.

The small IPTG-effect (compared to what is described in fig. 13) when the DNA binding domain is intact, and the detection of the slow fraction even when the DNA-binding domain is removed, are two indications that we are not answering the question we are asking. It should be noted though that when the experiment was repeated with a slightly different experimental setup, the result was essentially the same.

I will conclude here by saying that the data partly fits with the expectations. There is however something missing in order to understand the temperature effects observed in different experiments.

Dissociation rates

We now know how the repressor finds and binds the operator. How long does it stay bound? The motivation for asking this question is that the relation between association and dissociation rates, *i.e.* the fraction of time a DNA binding site is occupied, is the basis for understanding how a repressor molecule acts on transcription. The dissociation rate also determines how fast a system adapts to changes in TF concentrations.

Chase assay

The idea is to mimic an *in vitro* chase experiment, much like the early filter binding assays. First LacI-Venus is allowed to bind, and then a competitor in large excess is added so that when LacI-Venus dissociates it is replaced by a non-fluorescent LacI that prevents re-association. In the filter binding experiment, isotope-labeled DNA is outcompeted by the addition of 20-50-fold more unlabeled DNA (58). It is difficult to add proteins *in vivo*. The competitor should therefore be expressed in the cell. Our first idea was to have the *tet* repressor acting as competitor and its binding site was cloned next to *lacOsym* with a one bp overlap (*i.e.* the first and last bp of the respective operator is shared). As was described in the previous section (and paper I), we found that LacI and TetR positioned like this can bind simultaneously, and hence this kind of steric inhibition is not enough. Instead we went on to do an *in vitro* replica by letting LacI compete with LacI (paper II).

Repressor mutants have been identified that exhibit lowered affinity for inducer binding. LacI^{D274N} (109) has a >1000-fold reduction in IPTG-affinity and unchanged operator-binding strength. This single bp substitution was introduced in the *lac* repressor gene (referred to as *lacI_s*) in strains with a Venus fusion and *lacOsym* or *lacOI*, resulting in a chromosomally expressed LacI_s-Venus which does not dissociate in the presence of 1mM IPTG. LacI wild type (wt) is expressed from an arabinose inducible promoter on the plasmid pBAD24. With the addition of IPTG, LacI_s-Venus binds the operator whilst the wt LacI does not. However when fully induced for a long time,

the competitor copy number become so high that either 1mM IPTG does not prevent all LacI from binding; or, the LacI_s-Venus/LacI heterodimers that naturally form (and is dominant when LacI is overexpressed) and bind one IPTG-molecule, does not bind the operator. This is interesting in itself since it is debated whether one or two IPTG is needed for induction of one LacI dimer (110). Our result indicates that one is enough, although this can be a special case for the heterodimer.

We culture and image cells in microfluidic chips as described in the methods section and in paper II. Since too much LacI cannot be used from start, we again do as in the *in vitro* assay (fig. 16A). Although instead of adding competitor at time zero, we induce its expression. At start, with IPTG present, the LacI_s-Venus homodimer binds the operator, whilst neither the LacI tetramer nor the heterodimer do. The medium is changed so that IPTG is removed and 0.2% arabinose is added. Plasmid-expression of LacI is induced. With IPTG removed all repressors can bind and with wt LacI in excess, the initially bound species will be replaced following spontaneous dissociation.

At the start of the experiment there is a short period of increased binding. This is probably due to association of heterodimers (in competition with non-fluorescent LacI) to available operator sites. It can also be binding of LacI_s-Venus homodimers to pseudo-sites (assuming that these weak interactions are inducible). In addition there is a short lag until the induced plasmid produces a significant LacI concentration. To account for both these effects, we chose to consider the actual start of the chase experiment to be 1.5 minute after the switch of medium. Dissociation rates are fitted to this data, also accounting for the possibility of replication-dependent dissociation in the model. Replication is assumed to initiate with the rate of the doubling time (τ_G) (111). We measure the growth rates separately and use these as fixed parameters, such that all spots are supposed to have disappeared at latest τ_G minutes after the start.

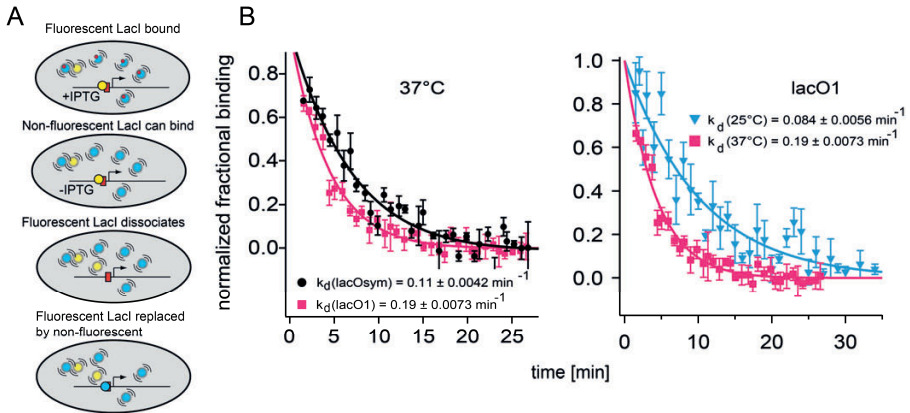


Figure 16. LacI dissociation, **(A)** the chase assay; binding of LacI_s-Venus and wt LacI is controlled by the presence of IPTG, **(B)** dissociation curves for different operators and temperatures.

Results

The doubling times in the microfluidic devices are $\tau_G \sim 26$ min at 37°C and ~ 70 min at 25°C. With dissociation at replication accounted for, the repressor binds to *lacO1* on average for 5 min at 37°C and for >10 min at 25°C (fig. 16B). The corresponding numbers *in vitro* depends on many experimental parameters and varies between intervals 7-20 (high salt) and 40-120 min (low salt) (48, 58, 66) at 25°C. At 37°C the operator binding is weakened less than 2-fold (65). It is reported that the difference between *lacO1* and *lacOsym* is up to 10-fold *in vitro* and 3-5-fold *in vivo* (75). We use this as a control and find that LacI binds to *lacOsym* on average for 10 min at 37°C, *i.e.* the difference is 2-fold.

As presented in paper I, the repressor seems to use 1D diffusion under our experimental conditions. We introduce a binding site for a TetR roadblock next to *lacOsym* and overexpress the roadblock from a chromosomal constitutive promoter. At 37°C, LacI dissociates 1.3 times slower compared to without roadblocks present. If this is due to an effect on sliding, then the association rate should be affected in a similar way. The association rate is measured at comparable conditions. The effect is a factor 1.3 and together these data suggest that the repressor slides in to and out from the operator site.

The dissociation rates are interesting in themselves, considering that all previous data are collected *in vitro*. The main finding here is that the experiments can be done *in vivo*. Next, the numbers are placed in a cellular context.

Implication on gene regulation

LacI-Venus binds to *lacO1* on average for 5 min before spontaneous dissociation or, as we assume, replication. It then takes 15-20 s until it re-associates to the operator (in strains without auto-repression). In between dissociation and association, expression is expected from the P_{lac} promoter. The initiation frequency has been estimated to one per 3.3 s (24), although this is probably lower in glucose rich medium due to a reduction in intracellular CRP levels (30, 112).

The equilibrium binding constant is given by $K=k_d/k_a$. The β -galactosidase assay measures the ratio between maximal (fully induced) and minimal (repressed) P_{lac} expression. The experiment is described in the methods section. If 1mM IPTG is enough for full induction, and expression is linear over the cell cycle as shown (24), then this repression ratio (RR) can be used to calculate the fraction of (cell cycle) time that the promoter is available for transcription (38). The RR relates to K as:

$$\frac{1}{RR} = \frac{1}{1 + [TF]/K} \quad (9)$$

We test this. In the repressed condition the LacI dimer dissociates and rebinds about five times per cell cycle and the ratio between fully induced and repressed transcription becomes ~ 20 . This is calculated from the denominator in equation 9, where our measured association rate is ($k_a[TF]$), and dissociation rate is k_d . The corresponding RR measured with the β -galactosidase assay is ~ 18 . This suggests that the model holds. As a control we repeat the test for *lacO_{sym}*. When repressed, LacI-Venus dissociates and rebinds about three times per cell cycle and the corresponding ratio is ~ 44 . The measured repression ratio is ~ 62 . A possible explanation to this minor discrepancy is a transcription initiation delay, which is related to the interplay between the binding of RNAP and LacI. I will return to this after first discussing the properties of the promoter region.

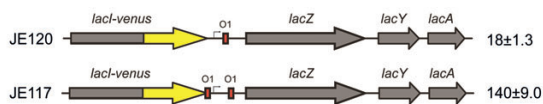


Figure 17. Repression ratios for LacI-Venus. (*left*) strains as described in paper I, (*right*) ratios of induced and repressed β -galactosidase activity. Results are from paper II (one *lacO1*) and additional data (two *lacO1*).

Effect of CRP-cAMP activation

The β -galactosidase experiments in fig. 17 shows that for two independent *lacO1* sites and the dimer binding, the repression ratio increases ~ 8 -fold

compared to when *lacO1* is in the promoter only. This was reported previously (38). It is mainly an effect on the repressed state and is explained as deactivation, *i.e.* the inhibition of CRP-cAMP activation.

With *lacO1* in the *lacO3* position the auto-repression from LacI-Venus is increased, the LacI-Venus concentration is decreased, and its operator occupancy at *lacO1* is decreased from 95% to 90%. With this taken into account and assuming that the dissociation rate is independent and the same at the two *lacO1* sites, then each of them will be occupied 90% of the time. At least one of the operator sites will be occupied >99% of the time. Consequently, either the transcription start site is occupied or, when free, CRP-cAMP is hindered to bind and transcription initiation is weakened.

The fraction of time the operator sites are simultaneously free can be used to calculate a hypothetical repression ratio for this combinatorial effect. The ratio becomes 110; close to the measured 140. This indicates that expression from P_{lac} is strongly repressed by deactivation.

The tetramer and the lac(k) of time

For the tetramer and with *lacO3* intact, CRP-cAMP binding must indeed be hindered by the DNA-loop, and so must RNAP-binding. $RR \sim 740$ for the wild type is 20-25 times higher than for the tetramer binding to a single *lacO1*. RR for the tetramer is about twice the value for LacI-Venus when repressing at a single *lacO1*. We ascribe this to a faster association which for example can be explained by intersegment transfer (74). This should have the same effect in the wild type strain and does not contribute to the 20-25 times difference. If we assume that the tetramer loops DNA, stays bound for the whole cell cycle (which we measure to $\tau_G \sim 26$ min) (76) and only dissociates at replication, then eq. 9 gives that it has to rebind in ~ 2 s. Based on our measurements for LacI-Venus binding to one *lacO1* we can estimate the time it takes for the repressor to bind in the wild type operon. First of all, we show in paper I that association to the first of the operators in a *lacO3-O1-O2* genotype is ~ 1.6 times faster than binding to *lacO1* only. This will contribute to the association rate for the tetramer if looping is fast (51, 113). Second, as described above, the tetramer probably binds twice as fast as the dimer to a single operator. Taken together this implies that the tetramer cannot rebind to the operator-region faster than in ~ 6 s, *i.e.* three times longer than the 2 s needed to explain the measured repression ratio. It does not bind fast enough; there is a lack of time. If the binding time of 6 s together with RR for the tetramer is put into eq. 9, it yields that $1/k_d$, the average time the repressor is bound, is >70 minutes, *i.e.* longer than the generation time.

Delay at transcription initiation

A delay in transcription initiation following dissociation could explain this. In paper II we present the following model:

$$RR_d = \frac{\tau_{on} + \tau_{off}}{\tau_{on}} e^{\tau_d/\tau_{on}} \quad (10)$$

where $\tau_{off}=k_d^{-1}$ is the average time the repressor stays bound, $\tau_{on}=[TF]^{-1}k_a^{-1}$ is the average time it takes to rebind, and τ_d is a possible delay. If $\tau_d=0$ then eq. 10 can be rewritten as eq. 9. A delay following replication where the promoter is silenced for some time would be on the order of 5-6 s to fit our measurements. The delay is probably due to the sequential re-binding of CRP-cAMP and RNAP that also are kicked off the DNA. If CRP-cAMP is in shortage relative LacI (*i.e.* in terms of time to bind the operator), it will sometimes be outcompeted and expression following replication will be silenced. For the average of the cell population this can be seen as a delay in transcription initiation. This does not hold if the activator binds much faster than LacI. Since no delay is needed to explain the *RR* for the tetramer (or the dimer) binding to *lacOI* only, it might indeed be an effect connected to replication.

Operator dependence

A delay following replication is not enough to explain the discrepancy seen for *lacOsym*. It is easy to dismiss this by saying that a small error in the dissociation, association or concentration measurement is enough to obtain the result. However the induced condition in the strain with *lacOsym/LacI-Venus* shows an expression level a factor of five lower than the one containing *lacOI/LacI-Venus*, indicating that the operator sequence somehow weakens the promoter. It is hard to reconcile how this would interfere directly with the CRP-cAMP binding. The significant reduction in the induced level might instead be due to slower open complex formation, if for some sequence-specific reason the bubble is more challenging to form. Another explanation would be that the lower expression level is connected to an initiation delay that also has effect following spontaneous dissociation. Abortive transcription initiation (22) which is effective in, and affected by, the operator sequence, and hence where the *lac* repressor operate might create such a delay due to the step-by-step nature of the initiation mechanism (paper II). A third suggestion is that the bound LacI actually is stabilized by a binding RNAP, in the same way as we see that it dissociates slower in the proximity of a TetR roadblock. This would be possible if RNAP binds next to an already bound LacI and then stays bound for a long time. The reason why this does not affect association of LacI in the promoter region is that the RNAP-promoter interactions are highly transient when LacI is not binding, *i.e.* RNAP rarely sits there alone. In any case a delay of 4-5 s is needed to increase the repression ratio from 44 to 62.

Conclusions

The interior of the cell is a micro-scale, complex version of the *great migration* which constantly circulates the border between Tanzania and Kenya. There millions of wildebeest and zebras search for the green spots, while acting pray for the magnificent cats. In the cell, biomolecules have their own great migration: birth and death, synthesis and degradation; a constant search for a partner to attract.

The problem of operator-recognition is often seen from the perspective of a single DNA-binding protein. In addition the system contains all other proteins, acting at the same time and in similar ways; hundreds of different TFs with copy numbers from a handful up to thousands of each; RNA polymerases, structural proteins, the replication machinery – less abundant but highly dynamic; all of them in a constant search for specific DNA, with temporal attractions to any DNA (62) and when bound acting as physical roadblocks for each other. When asking why repression efficiency simply is not solved with increasing TF copy numbers, one should remember the whole as the limit. In this context facilitated diffusion provides an opportunity to maintain regulation despite a limited TF concentration.

Regulation: search, find, bind and leave

The most important conclusion in this work is that the *lac* repressor displays 1D diffusion *in vivo* and that it slides along a helical path (*search*); that its low probability of binding suggests that the complex between non-specific DNA and repressor have different conformations, and that only one of these can recognize operator DNA (*find*); that the chance of being in the non-recognizing conformation when passing over the operator requires that the repressor typically visits the same site several times before association (*bind*).

From the dissociation rate measurements on the dimer we suggest that the tetramer stays operator-bound for the whole cell cycle (*leave*). The idea of silencing following replication is not new (24). If this model holds, the wild type *E. coli* population might contain less than one LacY per cell when the *lac* operon is repressed. If glucose is depleted and lactose becomes available, the stochastic outcompeting of repressor by the (now up-regulated) activator should result in a delay until all cells can grow. If medium is changed the other way around, all that is needed for repression is a decrease in allolactose and then the LacI tetramer can bind and repress the promoter in a few sec-

onds. I find it fascinating that the essential up-regulation of the *lac* operon can be so slow (114), when the cost-reducing but not life-saving repression of the same is so fast.

Future perspective of single molecule imaging

Gene regulation has its classical papers and so does the single molecule field. These include the first fluorescence imaging (115) of single molecules, and my personal favorite, the movies of rotating F_1 -ATPases (116). For the present and the future, two repeatedly discussed key problems to solve in single molecule fluorescence imaging are image analysis and tagging with fluorophores.

The challenging analysis

Contributions from several groups (including my colleagues) have provided nice tools for cell segmentation, quantification and tracking. While improvements still can be done, this problem seems to approach a solution (at least for *E. coli*) which we also exemplify in paper II. The main challenge will be to transfer knowledge and software between groups, so that people don't do the same thing over and over again.

The spot detection is a mystery. On one hand a fluorescent spot will always be close to diffraction limited, look like the point spread function and approximated with a 2D Gaussian. On the other hand, I have found it difficult use the same analysis tools for all my experiments, which are of course a bit different but always include a few spots in a cellular background. An advantage for spot detection compared to cell tracking is that even though the ideal is to include all true spots and exclude all false positives, some compromises are typically allowed. This is not the case in cell tracking, especially when tracking long cell histories. Despite of this, improvements in spot detection are still of importance.

The quest for the perfect tag

Fluorescent proteins have drawbacks such as maturation times, photo-bleaching and blinking. At the same time a fusion to a fluorescent protein is the ideal marker: all fluorescence that one sees (excluding cellular background fluorescence) is connected to the protein of interest. If maturation times could be completely diminished so that the protein could be detected directly after translation, and bleaching was hindered so that the same protein could be studied for extended times, these systems would be close to perfect. However the genetically expressed fusion might impair the function of the protein of interest and therefore a small molecule that can be used to genetically tag the protein is probably *the* solution.

Next step

The method for tracking of individual proteins in living cells using microfluidic, allows for the combination of a fast switch of medium and the collection of many trajectories at high time resolution. The development will continue and we will try to complement the analysis with segmentation of cells. This is important in order to remove boundary effects when the trajectories are affected by the cell membrane. It might also be possible to appoint different diffusive states to different locations in the cell (*i.e.* the chromosome contra cytosol). If we can use this method to quantify non-specific binding then it should also be applicable to other low-copy proteins.

The assays for measuring association and dissociation should (could) be generalized to study other proteins and interactions. There is for example the LacI tetramer, for which it first is required to find a non-impairing fluorescent tag as discussed above. Since in paper II we indirectly estimate its binding constant, direct measurements are of course of value. The proposed kinetics is also a nice challenge given that the estimated binding time is 6 s in the wild type scenario or even down to 2 s if the traditional model holds.

Turning back from the future, I draw the attention to Jacques Monod. He was part of opening the area of gene regulation, and his words will close this book: “Anything found to be true of *E. coli* must also be true of elephants” (117). It is not clear if the elephants of Serengeti agree, but it is an appealing thought.



Svensk sammanfattning

”Vi har alla krupit fram ut Gogols kappa”, sade den ene av mina favoritrys-sar (Dostojevskij) om den andre.

Den här doktorsavhandlingen har formatet av tre delarbeten, antingen publicerade eller i färd med att färdigställas. Boken börjar med en separat text som kallas för kappa och syftar till att knyta ihop dessa tre till en sammanhängande berättelse. Låt oss se vad som kryper fram om vi öppnar den.

I celler utgår tillverkningen av ett visst protein från en specifik gen på kromosomen. Produktionen regleras i hög grad av andra proteiner, så kallade transkriptionsfaktorer, som binder kemiskt på DNA (arvsmassan), nära den aktuella genen och stimulerar eller blockerar tillverkningen. De här regler-systemen kan beskrivas som en termostat, vilken hela tiden känner av omgivningen och talar om för exempelvis ett element att öka eller minska värme-produktionen. Det möjliggör för cellen att stänga av den resurskrävande produktionen av vissa proteiner om de inte behövs vid en viss tidpunkt. Till- lika kan tillverkningen snabbt trappas upp då förändringar i cellens närhet kräver nya funktioner. För att regleringen ska ha en kort svarstid är det nöd- värdigt att bundna transkriptionsfaktorer på en given signal kan lossna från DNA, men också att de snabbt kan hitta tillbaka; i bakterien *Escherichia coli* (förkortas *E. coli*) handlar det om att lokalisera en unik DNA-sekvens bland närapå fem miljoner felaktiga platser på kromosomen.

I min avhandling undersöker jag mekanismerna som styr hur transkript- ionsfaktorerna hittar till sina specifika platser på kromosomen och mäter hur snabbt det sker. Jag undersöker också hur länge molekylerna sitter bundna när de väl har kommit fram.

I molekylärbiologins unga historia har många mätningar gjorts med vad vi brukar kalla biokemiska metoder i provrör. I fallet med att studera hur proteiner interagerar med DNA så innebär det att de två molekylerna renas fram från levande celler, och sedan mäter man hur fort de kan binda varandra eller hur starkt de sitter bundna. På så vis går det att få detaljerade beskrivningar av hur saker fungerar. Genom matematisk modellering går det att omsätta de uppmätta siffrorna till hur det borde vara i levande celler.

Det är generellt svårt att mäta saker i levande organismer, dels för att de flesta metoder tar kål på dem, men också för att det finns så mycket okänt som påverkar i cellen, och som man inte behöver ta hänsyn till i provröret.

Därför är en central del i mina studier att jag har utvecklat metoder för att mäta de här sakerna i just levande *E. coli*-celler.

Metoderna bygger nästan uteslutande på att med hjälp av fluorescensmikroskopi titta på enskilda proteiner i cellerna. Det protein, eller den transkriptionsfaktor som har varit min modell, är ett vanligt förekommande studieobjekt i mikrobiologiska lab. Det heter LacI eller laktos-repressorn, vilket betyder att den binder på DNA och förhindrar att andra proteiner som är inblandade i nedbrytningen av just laktos tillverkas. Poängen är att om cellerna inte får laktos, så behövs inte heller de här proteinerna, och därför förhindras dess tillverkning. Om däremot laktos tillsätts så känner LacI av detta och lossnar. Därefter kan de proteiner som bryter ner laktos börja tillverkas och cellen, som i första hand äter glukos, kan börja leva på detta socker istället.

Storleken på en bakterie-cell är en miljondel av storleken på en människokropp. Proteinerna i cellerna är ytterligare tusen gånger mindre. Se figur 2 i kappan som visar LacI och ett annat protein som sitter bundna till DNA, i verkligheten är de ungefär fem miljarddelar av en meter stora. Det vi tittar på i mikroskopet är dock inte proteinerna i sig, utan ljuset ifrån dem när de märkts in med en fluorescent markör. Det innebär att när vi lyser på celler med grönt ljus, så utstrålar markören ett annat ljus, vilket vi kan filtrera ut och fånga upp med en mycket känslig kamera. Ljuset från en enda molekyl sprids runt en punkt vars minsta storlek begränsas av diffraktionsgränsen. Genom att använda olika exponeringstider går det att fånga upp proteinernas olika tillstånd. Om de rör sig fritt i cellerna så kommer de att under en sekund hinna täcka stora delar av cellens yta och därmed blir signalen utsmedad. Om de däremot är bundna hela tiden och ljuset samlas ihop från ett enda ställe, så kommer vi att detektera en skarp ljus prick.

Inne i kappan (figur 5) finns katten Sigge avbildad. Under en kort exponeringstid på 1/100 s så går det inte att avgöra om han rör sig eller inte, däremot går det att urskilja vattendropparna i kranen. Om han istället exponeras under ett par sekunder så blir han utsuddad samtidigt som kranvattnet uppfattas som en jämn stråle. Anledningen är att katten rör sig och avbildas på flera olika platser under bildtagningen, samtidigt som vattendropparna faller så snabbt att de dras ut. Med hjälp av den här principen kan vi skilja på proteiner som är DNA-bundna och sådana om rör sig fritt. Detektionen ger tillbaka en så kallad fluorescent prick som vi sedan analyserar med hjälp av egenutvecklad programvara. Sammanfattningsvis så är detta grunden i hur vi mäter bindningstider.

I den enklaste hypotetiska modellen för hur snabbt en transkriptionsfaktor hittar rätt plats, söker den runt genom att testa sig fram. Den utför en slumpmässig vandring där den flyttar runt genom diffusion, dvs. dess förflyttningar beror på termisk rörelse och på att den knuffas av omgivande

molekyler. Populärvetenskapligt brukar detta beskrivas som en fyllots vandring. Man kan tänka sig att fyllet, istället för att gå raka vägen hem, för varje steg hen tar, har lika stor sannolikhet att ta ett steg åt fel håll som åt rätt håll. Hen kommer att komma hem förr eller senare, men det kommer att ta lång tid.

Från de tidiga biokemi-experimenten i provrör är det känt att den här enklaste modellen inte är tillräcklig för att beskriva hur LacI hittar till sitt bindningsställe; den hittar fram snabbare än så. Här har man jämfört resultaten med beräkningar på vad som är fysikaliskt möjligt, för att föreslå alternativa mekanismer. En länge levande teori är den om faciliterad diffusion. Den säger att förutom den slumpmässiga vandringen i alla riktningar, så kommer molekylerna när de träffar på DNA att även röra sig fram och tillbaka på detta, också det i en slumpmässig vandring. Det gör att varje träff med DNA leder till att transkriptionsfaktorn kommer att leta igenom även dess närliggande platser. Därmed krävs färre slumpmässiga test innan den kommer rätt och på så vis går det att förklara att den binder så snabbt som den gör.

Man kan tänka sig en boll som studsar runt med oförminskad kraft i ett stort rum. Någonstans finns ett litet hål i golvet eller väggen. Om bollen studsar för evigt kommer den tillslut att träffa det lilla hålet, men chansen är liten. Om något monteras framför hålet, som kan leda bollen rätt, eller naturligtvis om hålets storlek ökas, så kommer chansen öka och den hittar fram snabbare. Ett annat visuellt exempel är den parabol som ökar effektiviteten hos antenner.

Experiment i levande celler är mer komplicerade än i provrör, eftersom cellerna innehåller så många fler och olika komponenter än de få som blandas i experiment utan celler. Vi har mätt upp att en LacI-molekyl hittar fram och binder in till rätt ställe på en dryg minut i 37°C, men det finns omkring fem likadana proteiner och därför tar det omkring 15 sekunder innan ett av dem hittar rätt. Vi har undersökt faciliterad diffusion för LacI i *E. coli* och visat att det existerar. Transkriptionsfaktorn träffar slumpmässigt på DNA och scannar därefter av ca 40 av dess positioner (baspar) innan den lossnar och söker av ett nytt ställe. Om den träffar rätt binder den. Vi kan också se att om ett annat protein binder starkt till DNA, intill det ställe där LacI ska sitta, så kan inte LacI passera; dvs. dess sökning längs DNA blockeras. Med hjälp av dator-simuleringar kallade molekylärdynamik kan vi också dra slutsatsen att sökningen längs med DNA sker genom att LacI roterar och följer DNA-strängens struktur. Alternativet, vilket vi motbevisar, skulle vara att den rör sig enbart på en sida av DNA.

Slutligen har vi med liknande experimentella metoder som beskrivits ovan mätt hur länge LacI sitter bunden och blockerar proteintillverkning. Det är minst fem minuter i taget. Det protein vi har studerat är skilt från det som finns naturligt i cellen. Den varianten kan istället binda samtidigt till två närliggande ställen på DNA och vår hypotes är att den sitter bunden hela

cellens livscykel, dvs. den tid det tar för cellen att dela sig. Tiden varierar beroende på vilken näring cellerna får och hur fort de därför växer. I det här fallet tar det ungefär 25 minuter för en cell att bli två. För att bägge dottercellerna ska överleva så måste cellen kopiera sitt DNA (kromosomen) i två innan delningen är färdig, och idén är att det är när detta sker som LacI trillar av. Det här är spännande från regleringspunkt därför att om LacI-molekylen enbart lossnar en gång per celldelning (eller replikation som kopiering av DNA heter), så sitter den strängt taget bunden så länge det är möjligt. Därmed är tillverkningen av de proteiner den reglerar nedtryckt till minsta möjliga nivå. Man brukar dock säga att för många proteiner finns det en läckagenivå, ett minsta antal i cellerna. Syftet är att det behövs en beredskapsplan. I det här fallet syftar det på att om glukos i omgivningen plötsligt tar slut, men laktos finns tillgängligt, så kan cellerna snabbt ställa om och direkt ha möjligheten att bryta ned laktos. Vi föreslår dock att även efter replikationen finns det så lite tid för tillverkning innan LacI binder tillbaka och blockerar, att det här läckaget måste vara mycket lågt. Resursförbrukningen är verkligen minimerad.

Sammanfattningsvis så visar resultaten att problem som tidigare bara kunde studeras med biokemiska metoder för framrenade komponenter nu också kan angripas i levande bakterier. Vi presenterar nya data för hur en av cellens mest fundamentala funktioner är reglerad. Men vi understryker också nyttan av att kombinera ny mätteknik med avancerade datasimuleringar och matematisk modellering för att bryta ny mark inom molekylärbiologin.

De små och snabba förlopp vi studerar kan tyckas ogreppbara. För att inte blir tokig (eller för den delen religiös) när man försöker förstå cellernas reglering, måste man, något motsägelsefullt, tro på det vi inte kan se; på vad som går snabbare än vad vi kan uppfatta. För att spetsa till det ytterligare kan jag nämna att kromosomen i *E. coli* har över 4000 gener, och den tillverkar ungefär lika många olika typer av proteiner. Många av dem är precis som LacI ständigt på jakt efter en annan molekyl att binda till. I en vuxen människa lever hundra tusen miljarder (14 nollor) bakterier, av många olika arter och de flesta ofarliga, ja rentav viktiga för oss. Vi bär på en sammanlagd bakterievikt av 1-2 kg. Men bakterierna är små och trots den ringa vikten är de till antalet tio gånger fler än människans "egna" celler. Tänk alltså att det finns hundra tusen miljarder bakterier, som var och en bär på tusentals proteiner med olika funktioner. Föreställ er nu scenen i filmen *Men In Black*, när de tittar ut genom dörren på en förvaringsbox på tågstationen, och visar att vår civilisation bara är ett skruvt i det stora alltet. Det här är samma sak, fast tvärtom.

Acknowledgements

First I want to thank my unsuccessful experiments, for the lessons given; second my successful experiments, a necessity for writing this text.

Johan, master Elf: for the freedom in the lab, for the investments on all levels, for the travels; for your nightly and never-ending support. Thanks for questioning me and listening to me.

Otto Berg, for the patience and superb tuition, and for the amazing papers of 1981 – a good year it seems.

Professor Clint Eastwood, for being yourself and providing yours and the Feynman-perspective; well, some kind of perspective at least.

The Uppsala hobbits

Prune, for the dinners and kindness, the drinking riots, the *car*, the supervision and funny times in the office; Brian, my first optics guru, for all the things you taught me, for your peculiar way of debating, for the pub and football nights and the invitation to NY; Paul, for looking after me in the beginning; ADHD-Mats, for putting myself in perspective, and for putting me on that raft, for great friendship and excellent collaborations; David, for always being around, in- and outside the lab; and for never giving up on me and my questions; Ryuyo, for your goodness and wind from the orient; Arash, for the things I never could have done, and the time you saved me; Anel, for rewarding discussions, for blobPalette, and for staying within ten orders of magnitude; Vasili, double-Dr, for the fun and for the different kind of feedback; Stoyan, Speedo Man, for the eastern touch and endless football talks; Nina, for the teaching and teamwork; Andreas for your interesting way of understanding biology; Cia, for always keeping up the spirit and keeping track of my lab bench; Gustaf, Fourier-mannen, for transforming my sanity; Erik, for the dark music and bright mind, and productive collaborations; Fredrik, for the lessons, for the guerilla blast (and other blasts), and for much-needed support during the last months; Ozzy, for the re-birth of the beer-club and your everlasting energy; Irmeli, for pushing me towards this day; Kalle, for keeping track of me during a chaotic period.

ICM. Janos Hajdu, for a surprising mentorship; the biophysics crew for the entertainment in the lunch room, and Harri, Magnus and Niklas for the same in the lab; Harriet, for the raised fist and for keeping up the fight; past and

present students, postdocs (and others), for the funny interactions, beer clubs and football; Nadja and Zvoni, for the first months of red wine and nightly talks in Flogsta; Ankarklev, for the long and great discussions, and for understanding that there are other ways than the normal ones.

Family. My parents, thanks for everything, for always being there, you're the best! Gustav, the artist and moviestar, for being so much more than my brother; the rest of my family, in Sweden and Poland; Susana, who has been around throughout my life, for great support when needed, and for sequencing my DNA.

Sandviken. Viktor, my second brother, for a life-long unity, and for the honor in 2009; Kenny the incredible who never stops surprising me, for all the fun, and for the honor in 2010; Jocke the dude, I never carried your bag in Sydney, but some day we will open that book store; Affe and Ronnie for all the escapades and fun we have had; Mange; for never getting rid of the Storvik-spirit (and for long time, neither that amazing car); Macke, for being such a good person in a life that changed but didn't end.

Diana, the most important person in 9th grade and a wonderful teacher; three high school teachers: Inger, for the struggles and obstacles, for helping me write that book, and for the final appreciation; Anders, for the wake-up-call, telling me "you are not serious enough"; Erik, the biologist and scientist, for the inspiration that initiated this journey.

Linköping. Ola, for the nights and days in Ryd (and sometimes in school) and for the proof-readings; Erik, for the interesting explorations of the humanity; Anna, for balancing my unorganized approaches; Martin, for the adventures with the caravan and the ÖB-tent, and for bringing me in to UR-Fas; Kristina, for great times, with and without claws and tails; Maria, for early inspiring me to do research; Greta and Polona, for opening my eyes during a time when the world changed; all the fabulous dinosaurs.

Jan-Ingvar, for the opportunities, for encouraging me to do this, for the football and politics; Pernilla, for the supervision (also after beer-hour), fun and inspiration, and for filling in the details missed out by the Boss.

And... the friendly people of Iran, the humble people of Tibet, and the travelers of 2005, who were around when I went out looking for Natasha; Zlatan, for the entertainment; Sigge, for the joy and company during long nights.

Finally, Jenny, fantastic sweet troll, for all the fun and for being at my side; I hope the adventures never end – not even the sky is the limit!

List of references

1. L. Mlodinow, *The Drunkard's walk : how randomness rules our lives*. (Pantheon Books, New York, ed. 1st Vintage Books, 2009), pp. xi, 252 p.
2. H. C. Berg, *Random walks in biology*. (Princeton University Press, Princeton, N.J., ed. Expanded, 1993), pp. 152 p.
3. M. Schaechter, G. View From Here, Escherichia coli and Salmonella 2000: the view from here. *Microbiology and molecular biology reviews : MMBR* **65**, 119 (Mar, 2001).
4. J. D. Watson, F. H. Crick, Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737 (Apr 25, 1953).
5. F. H. Crick, On protein synthesis. *Symposia of the Society for Experimental Biology* **12**, 138 (1958).
6. F. Crick, Central dogma of molecular biology. *Nature* **227**, 561 (Aug 8, 1970).
7. F. Jacob, J. Monod, Genetic regulatory mechanisms in the synthesis of proteins. *Journal of molecular biology* **3**, 318 (Jun, 1961).
8. W. Gilbert, B. Muller-Hill, Isolation of the lac repressor. *Proceedings of the National Academy of Sciences of the United States of America* **56**, 1891 (Dec, 1966).
9. W. Gilbert, B. Muller-Hill, The lac operator is DNA. *Proceedings of the National Academy of Sciences of the United States of America* **58**, 2415 (Dec, 1967).
10. B. Müller-Hill, *The lac Operon : a short history of a genetic paradigm*. (Walter de Gruyter, Berlin ; New York, 1996), pp. ix, 207 p.
11. D. S. Goodsell, Escherichia coli. *Biochemistry and molecular biology education : a bimonthly publication of the International Union of Biochemistry and Molecular Biology* **37**, 325 (Nov, 2009).
12. P. B. Eckburg *et al.*, Diversity of the human intestinal microbial flora. *Science* **308**, 1635 (Jun 10, 2005).
13. J. Lederberg, E. L. Tatum, Gene recombination in Escherichia coli. *Nature* **158**, 558 (Oct 19, 1946).
14. F. R. Blattner *et al.*, The complete genome sequence of Escherichia coli K-12. *Science* **277**, 1453 (Sep 5, 1997).
15. R. T. Dame, The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Molecular microbiology* **56**, 858 (May, 2005).
16. S. Brantl, Bacterial chromosome-encoded small regulatory RNAs. *Future microbiology* **4**, 85 (Feb, 2009).
17. F. Wang, E. C. Greene, Single-molecule studies of transcription: from one RNA polymerase at a time to the gene expression profile of a cell. *Journal of molecular biology* **412**, 814 (Oct 7, 2011).
18. F. Wang *et al.*, The promoter-search mechanism of Escherichia coli RNA polymerase is dominated by three-dimensional diffusion. *Nature structural & molecular biology*, (Dec 23, 2012).

19. C. Bustamante, M. Guthold, X. Zhu, G. Yang, Facilitated target location on DNA by individual Escherichia coli RNA polymerase molecules observed with the scanning force microscope operating in liquid. *The Journal of biological chemistry* **274**, 16665 (Jun 11, 1999).
20. D. F. Browning, S. J. Busby, The regulation of bacterial transcription initiation. *Nature reviews. Microbiology* **2**, 57 (Jan, 2004).
21. A. Ishihama, Functional modulation of Escherichia coli RNA polymerase. *Annual review of microbiology* **54**, 499 (2000).
22. S. R. Goldman, R. H. Ebright, B. E. Nickels, Direct detection of abortive RNA transcripts in vivo. *Science* **324**, 927 (May 15, 2009).
23. E. A. Abbondanzieri, W. J. Greenleaf, J. W. Shaevitz, R. Landick, S. M. Block, Direct observation of base-pair stepping by RNA polymerase. *Nature* **438**, 460 (Nov 24, 2005).
24. D. Kennell, H. Riezman, Transcription and translation initiation frequencies of the Escherichia coli lac operon. *Journal of molecular biology* **114**, 1 (Jul, 1977).
25. T. M. Schmeing, V. Ramakrishnan, What recent ribosome structures have revealed about the mechanism of translation. *Nature* **461**, 1234 (Oct 29, 2009).
26. E. Perez-Rueda, J. Collado-Vides, The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12. *Nucleic acids research* **28**, 1838 (Apr 15, 2000).
27. G. Orphanides, T. Lagrange, D. Reinberg, The general transcription factors of RNA polymerase II. *Genes & development* **10**, 2657 (Nov 1, 1996).
28. S. Busby, R. H. Ebright, Transcription activation by catabolite activator protein (CAP). *Journal of molecular biology* **293**, 199 (Oct 22, 1999).
29. P. W. Postma, J. W. Lengeler, G. R. Jacobson, Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiological reviews* **57**, 543 (Sep, 1993).
30. H. Ishizuka, A. Hanamura, T. Kunimura, H. Aiba, A lowered concentration of cAMP receptor protein caused by glucose is an important determinant for catabolite repression in Escherichia coli. *Molecular microbiology* **10**, 341 (Oct, 1993).
31. T. Kuhlman, Z. Zhang, M. H. Saier, Jr., T. Hwa, Combinatorial transcriptional control of the lactose operon of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 6043 (Apr 3, 2007).
32. V. Sasson, I. Shachrai, A. Bren, E. Dekel, U. Alon, Mode of regulation and the insulation of bacterial gene expression. *Molecular cell* **46**, 399 (May 25, 2012).
33. O. G. Berg, P. H. von Hippel, Selection of DNA binding sites by regulatory proteins. II. The binding specificity of cyclic AMP receptor protein to recognition sites. *Journal of molecular biology* **200**, 709 (Apr 20, 1988).
34. C. M. Johnson, R. F. Schleif, In vivo induction kinetics of the arabinose promoters in Escherichia coli. *Journal of bacteriology* **177**, 3438 (Jun, 1995).
35. R. Schleif, Regulation of the L-arabinose operon of Escherichia coli. *Trends in genetics : TIG* **16**, 559 (Dec, 2000).
36. T. M. Dunn, S. Hahn, S. Ogden, R. F. Schleif, An operator at -280 base pairs that is required for repression of araBAD operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 5017 (Aug, 1984).
37. H. Ishizuka, A. Hanamura, T. Inada, H. Aiba, Mechanism of the down-regulation of cAMP receptor protein by glucose in Escherichia coli: role of autoregulation of the crp gene. *The EMBO journal* **13**, 3077 (Jul 1, 1994).

38. S. Oehler, M. Amouyal, P. Kolkhof, B. von Wilcken-Bergmann, B. Muller-Hill, Quality and position of the three lac operators of E. coli define efficiency of repression. *The EMBO journal* **13**, 3348 (Jul 15, 1994).
39. S. R. Connell, D. M. Tracz, K. H. Nierhaus, D. E. Taylor, Ribosomal protection proteins and their mechanism of tetracycline resistance. *Antimicrobial agents and chemotherapy* **47**, 3675 (Dec, 2003).
40. W. Saenger, P. Orth, C. Kisker, W. Hillen, W. Hinrichs, The Tetracycline Repressor-A Paradigm for a Biological Switch. *Angew Chem Int Ed Engl* **39**, 2042 (Jun 16, 2000).
41. I. Meier, L. V. Wray, W. Hillen, Differential regulation of the Tn10-encoded tetracycline resistance genes tetA and tetR by the tandem tet operators O1 and O2. *The EMBO journal* **7**, 567 (Feb, 1988).
42. J. Chen, S. Alberti, K. S. Matthews, Wild-type operator binding and altered cooperativity for inducer binding of lac repressor dimer mutant R3. *The Journal of biological chemistry* **269**, 12482 (Apr 29, 1994).
43. M. Lewis *et al.*, Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* **271**, 1247 (Mar 1, 1996).
44. M. D. Barkley, A. D. Riggs, A. Jobe, S. Bourgeois, Interaction of effecting ligands with lac repressor and repressor-operator complex. *Biochemistry* **14**, 1700 (Apr 22, 1975).
45. R. Fickert, B. Muller-Hill, How Lac repressor finds lac operator in vitro. *Journal of molecular biology* **226**, 59 (Jul 5, 1992).
46. J. D. Hopkins, A new class of promoter mutations in the lactose operon of Escherichia coli. *Journal of molecular biology* **87**, 715 (Aug 25, 1974).
47. H. Salgado *et al.*, RegulonDB (version 3.0): transcriptional regulation and operon organization in Escherichia coli K-12. *Nucleic acids research* **28**, 65 (Jan 1, 2000).
48. M. Pfahl, V. Gulde, S. Bourgeois, "Second" and "third operator" of the lac operon: an investigation of their role in the regulatory mechanism. *Journal of molecular biology* **127**, 339 (Jan 25, 1979).
49. W. T. Hsieh, P. A. Whitson, K. S. Matthews, R. D. Wells, Influence of sequence and distance between two operators on interaction with the lac repressor. *The Journal of biological chemistry* **262**, 14583 (Oct 25, 1987).
50. J. M. Vilar, S. Leibler, DNA looping and physical constraints on transcription regulation. *Journal of molecular biology* **331**, 981 (Aug 29, 2003).
51. G. W. Li, O. G. Berg, J. Elf, Effects of macromolecular crowding and DNA looping on gene regulation kinetics. *Nat Phys* **5**, 294 (Apr, 2009).
52. H. R. Kaback, E. Bibi, P. D. Roepe, Beta-galactoside transport in E. coli: a functional dissection of lac permease. *Trends in biochemical sciences* **15**, 309 (Aug, 1990).
53. P. R. Jensen, H. V. Westerhoff, O. Michelsen, The use of lac-type promoters in control analysis. *European journal of biochemistry / FEBS* **211**, 181 (Jan 15, 1993).
54. D. H. Juers, B. W. Matthews, R. E. Huber, LacZ beta-galactosidase: structure and function of an enzyme of historical and molecular biological importance. *Protein science : a publication of the Protein Society* **21**, 1792 (Dec, 2012).
55. R. E. Huber, G. Kurz, K. Wallenfels, A quantitation of the factors which affect the hydrolase and transgalactosylase activities of beta-galactosidase (E. coli) on lactose. *Biochemistry* **15**, 1994 (May 4, 1976).
56. A. Majumdar, S. Adhya, Demonstration of two operator elements in gal: in vitro repressor binding studies. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 6100 (Oct, 1984).

57. A. Marbach, K. Bettenbrock, lac operon induction in Escherichia coli: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA. *Journal of biotechnology* **157**, 82 (Jan, 2012).
58. A. D. Riggs, S. Bourgeois, M. Cohn, The lac repressor-operator interaction. 3. Kinetic studies. *Journal of molecular biology* **53**, 401 (Nov 14, 1970).
59. Y. Kao-Huang *et al.*, Nonspecific DNA binding of genome-regulating proteins as a biological control mechanism: measurement of DNA-bound Escherichia coli lac repressor in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 4228 (Oct, 1977).
60. J. Elf, G. W. Li, X. S. Xie, Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191 (May 25, 2007).
61. A. Revzin, P. H. von Hippel, Direct measurement of association constants for the binding of Escherichia coli lac repressor to non-operator DNA. *Biochemistry* **16**, 4769 (Nov 1, 1977).
62. C. G. Kalodimos *et al.*, Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science* **305**, 386 (Jul 16, 2004).
63. M. T. Record, Jr., P. L. deHaseth, T. M. Lohman, Interpretation of monovalent and divalent cation effects on the lac repressor-operator interaction. *Biochemistry* **16**, 4791 (Nov 1, 1977).
64. P. L. deHaseth, T. M. Lohman, M. T. Record, Jr., Nonspecific interaction of lac repressor with DNA: an association reaction driven by counterion release. *Biochemistry* **16**, 4783 (Nov 1, 1977).
65. P. A. Whitton, J. S. Olson, K. S. Matthews, Thermodynamic analysis of the lactose repressor-operator DNA interaction. *Biochemistry* **25**, 3852 (Jul 1, 1986).
66. R. B. Winter, O. G. Berg, P. H. von Hippel, Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The Escherichia coli lac repressor--operator interaction: kinetic measurements and conclusions. *Biochemistry* **20**, 6961 (Nov 24, 1981).
67. M. D. Barkley, Salt dependence of the kinetics of the lac repressor-operator interaction: role of nonoperator deoxyribonucleic acid in the association reaction. *Biochemistry* **20**, 3833 (Jun 23, 1981).
68. R. McMacken, K. Ueda, A. Kornberg, Migration of Escherichia coli dnaB protein on the template DNA strand as a mechanism in initiating DNA replication. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 4190 (Oct, 1977).
69. B. P. English *et al.*, Single-molecule investigations of the stringent response machinery in living bacterial cells. *Proceedings of the National Academy of Sciences of the United States of America* **108**, E365 (Aug 2, 2011).
70. P. H. von Hippel, O. G. Berg, Facilitated target location in biological systems. *The Journal of biological chemistry* **264**, 675 (Jan 15, 1989).
71. P. H. Richter, M. Eigen, Diffusion-controlled reaction rates in spheroidal geometry. Application to repressor--operator association and membrane bound enzymes. *Biophysical chemistry* **2**, 255 (Oct, 1974).
72. O. G. Berg, C. Blomberg, Association kinetics with coupled diffusional flows. Special application to the lac repressor--operator system. *Biophysical chemistry* **4**, 367 (Jul, 1976).
73. J. M. Schurr, The one-dimensional diffusion coefficient of proteins absorbed on DNA. Hydrodynamic considerations. *Biophysical chemistry* **9**, 413 (May, 1979).

74. O. G. Berg, R. B. Winter, P. H. von Hippel, Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **20**, 6929 (Nov 24, 1981).
75. J. R. Sadler, H. Sasmor, J. L. Betz, A perfectly symmetric lac operator binds the lac repressor very tightly. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 6785 (Nov, 1983).
76. S. Oehler, E. R. Eismann, H. Kramer, B. Muller-Hill, The three operators of the lac operon cooperate in repression. *The EMBO journal* **9**, 973 (Apr, 1990).
77. J. Yu, J. Xiao, X. Ren, K. Lao, X. S. Xie, Probing gene expression in live cells, one protein molecule at a time. *Science* **311**, 1600 (Mar 17, 2006).
78. K. A. Datsenko, B. L. Wanner, One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 6640 (Jun 6, 2000).
79. A. J. Link, D. Phillips, G. M. Church, Methods for generating precise deletions and insertions in the genome of wild-type Escherichia coli: application to open reading frame characterization. *Journal of bacteriology* **179**, 6228 (Oct, 1997).
80. J. H. Miller, *A short course in bacterial genetics : a laboratory manual and handbook for Escherichia coli and related bacteria*. (Cold Spring Harbor Laboratory Press, Plainview, N.Y., 1992).
81. L. M. Guzman, D. Belin, M. J. Carson, J. Beckwith, Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of bacteriology* **177**, 4121 (Jul, 1995).
82. D. A. Siegele, J. C. Hu, Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 8168 (Jul 22, 1997).
83. A. Edelstein, N. Amodaj, K. Hoover, R. Vale, N. Stuurman, Computer control of microscopes using microManager. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.] Chapter 14*, Unit14 20 (Oct, 2010).
84. M. Ormo *et al.*, Crystal structure of the Aequorea victoria green fluorescent protein. *Science* **273**, 1392 (Sep 6, 1996).
85. T. Nagai *et al.*, A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature biotechnology* **20**, 87 (Jan, 2002).
86. D. Landgraf, B. Okumus, P. Chien, T. A. Baker, J. Paulsson, Segregation of molecules at cell division reveals native protein localization. *Nature methods* **9**, 480 (May, 2012).
87. G. Ullman *et al.*, High-throughput gene expression analysis at the level of single proteins using a microfluidic turbidostat and automated cell tracking. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **368**, 20120025 (Feb 5, 2013).
88. D. B. Murphy, *Fundamentals of light microscopy and electronic imaging*. (Wiley-Liss, New York, 2001), pp. xii, 368 p.
89. R. E. Thompson, D. R. Larson, W. W. Webb, Precise nanometer localization analysis for individual fluorescent probes. *Biophysical journal* **82**, 2775 (May, 2002).
90. J. C. Olivo-Marin, Extraction of spots in biological images using multiscale products. *Pattern Recogn* **35**, 1989 (Sep, 2002).
91. G. W. Li, J. Elf, Single molecule approaches to transcription factor kinetics in living cells. *FEBS letters* **583**, 3979 (Dec 17, 2009).
92. I. Bonnet *et al.*, Sliding and jumping of single EcoRV restriction enzymes on non-cognate DNA. *Nucleic acids research* **36**, 4118 (Jul, 2008).

93. P. C. Blainey, A. M. van Oijen, A. Banerjee, G. L. Verdine, X. S. Xie, A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 5752 (Apr 11, 2006).
94. A. Tafvizi *et al.*, Tumor suppressor p53 slides on DNA with low friction and high stability. *Biophysical journal* **95**, L01 (Jul, 2008).
95. C. Kleinschmidt, K. Tovar, W. Hillen, D. Porschke, Dynamics of repressor-operator recognition: the Tn10-encoded tetracycline resistance control. *Biochemistry* **27**, 1094 (Feb 23, 1988).
96. T. Ruusala, D. M. Crothers, Sliding and intermolecular transfer of the lac repressor: kinetic perturbation of a reaction intermediate by a distant DNA sequence. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 4903 (Jun 1, 1992).
97. D. M. Gowers, G. G. Wilson, S. E. Halford, Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15883 (Nov 1, 2005).
98. A. Jeltsch, C. Wenz, F. Stahl, A. Pingoud, Linear diffusion of the restriction endonuclease EcoRV on DNA is essential for the in vivo function of the enzyme. *The EMBO journal* **15**, 5104 (Sep 16, 1996).
99. C. P. Bahl, R. Wu, J. Stawinsky, S. A. Narang, Minimal length of the lactose operator sequence for the specific recognition by the lactose repressor. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 966 (Mar, 1977).
100. M. Kampmann, Obstacle bypass in protein motion along DNA by two-dimensional rather than one-dimensional sliding. *The Journal of biological chemistry* **279**, 38715 (Sep 10, 2004).
101. P. C. Blainey *et al.*, Nonspecifically bound proteins spin while diffusing along DNA. *Nature structural & molecular biology* **16**, 1224 (Dec, 2009).
102. M. Slutsky, L. A. Mirny, Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. *Biophysical journal* **87**, 4021 (Dec, 2004).
103. M. Bauer, R. Metzler, Generalized facilitated diffusion model for DNA-binding proteins with search and recognition states. *Biophysical journal* **102**, 2321 (May 16, 2012).
104. L. Niu, J. Yu, Investigating intracellular dynamics of FtsZ cytoskeleton with photoactivation single-molecule tracking. *Biophysical journal* **95**, 2009 (Aug, 2008).
105. S. Manley *et al.*, High-density mapping of single-molecule trajectories with photoactivated localization microscopy. *Nature methods* **5**, 155 (Feb, 2008).
106. X. S. Xie, J. Yu, W. Y. Yang, Living cells as test tubes. *Science* **312**, 228 (Apr 14, 2006).
107. F. Persson, M. Linden, C. Unoson, J. Elf, Extracting intracellular diffusive states and transition rates from single-molecule tracking data. *Nature methods* **10**, 265 (Mar, 2013).
108. T. E. Kuhlman, E. C. Cox, Gene location and DNA density determine transcription factor distributions in Escherichia coli. *Molecular systems biology* **8**, 610 (Sep 11, 2012).
109. W. I. Chang, P. Barrera, K. S. Matthews, Identification and characterization of aspartate residues that play key roles in the allosteric regulation of a transcription factor: aspartate 274 is essential for inducer binding in lac repressor. *Biochemistry* **33**, 3607 (Mar 29, 1994).

110. L. Swint-Kruse, H. Zhan, K. S. Matthews, Integrated insights from simulation, experiment, and mutational analysis yield new details of LacI function. *Biochemistry* **44**, 11201 (Aug 23, 2005).
111. S. Cooper, C. E. Helmstetter, Chromosome replication and the division cycle of *Escherichia coli* B/r. *Journal of molecular biology* **31**, 519 (Feb 14, 1968).
112. T. P. Malan, A. Kolb, H. Buc, W. R. McClure, Mechanism of CRP-cAMP activation of lac operon transcription initiation activation of the P1 promoter. *Journal of molecular biology* **180**, 881 (Dec 25, 1984).
113. L. Finzi, J. Gelles, Measurement of lactose repressor-mediated loop formation and breakdown in single DNA molecules. *Science* **267**, 378 (Jan 20, 1995).
114. A. L. Koch, Lag in adaptation to lactose as a probe to the timing of permease incorporation into the cell membrane. *Journal of bacteriology* **124**, 435 (Oct, 1975).
115. T. Hirschfeld, Optical microscopic observation of single small molecules. *Applied optics* **15**, 2965 (Dec 1, 1976).
116. H. Noji, R. Yasuda, M. Yoshida, K. Kinosita, Jr., Direct observation of the rotation of F1-ATPase. *Nature* **386**, 299 (Mar 20, 1997).
117. H. C. Friedmann, From "butyribacterium" to "E. coli": an essay on unity in biochemistry. *Perspectives in biology and medicine* **47**, 47 (Winter, 2004).

Acta Universitatis Upsaliensis

*Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Science and Technology 1046*

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology.

Distribution: publications.uu.se
urn:nbn:se:uu:diva-198814



ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2013