University of Groningen

Initiation rites in molecular biology

Schavemaker, Paulus Eelke

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2018

Citation for published version (APA):
Schavemaker, P. E. (2018). *Initiation rites in molecular biology*. University of Groningen.

Initiation Rites in Molecular Biology

Paul E. Schavemaker

2018

university of groningen

Initiation Rites in Molecular Biology

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. E. Sterken
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Friday 18 May 2018 at 12.45 hours

by

Paulus Eelke Schavemaker

born on 25 June 1989
in Terschelling

Supervisor

Prof. B. Poolman

Assessment Committee

Prof. S.J. Marrink

Prof. P.J.M. van Haastert

Prof. C. Mullineaux

Contents

Preface

This thesis is a three parter. The first part (Ch. 1, 2 and 3) is about diffusion of proteins in prokaryotes. Chapter 1 serves as an introduction and chapters 2 and 3 describe our efforts to measure diffusion rates in prokaryotic cells. We have shown, among other things, that protein diffusion rates in *Lactococcus lactis* are strongly responsive to abrupt volume changes; and that diffusion rate strongly depends on the protein charge. The second part (Ch. 4 and 5) is about the production of membrane proteins in bacteria. Here, chapter 4 serves as an introduction. Chapter 5 details our efforts to measure, in bacterial cells, the production rates of membrane proteins from single mRNAs. In the third and final part (Ch. 6) I take a breather from the suffocating grasp of my PhD and discuss some issues about science, society, and biology that interest me.

Chapter 1: On the importance of knowing protein diffusion rates in prokaryotes

Everything jiggles! This is one of the foremost facts about the innards of cells that any student of molecular biology should know. This jiggling allows proteins and cells to sample different states and provides meaning to the concept of entropy. One other consequence of perpetual jiggling, and the most relevant for us here, is that molecules move around without using any free energy or (directed) mechanism. This is called diffusion. The type of diffusion I will be discussing is translational diffusion which is the displacement of the center of mass of an object, this is in contrast to rotational diffusion which is the rotation of an object around its center of mass. For a cell, say *Escherichia coli*, to grow and divide it needs all kinds of molecules to find one another: substrates need to find enzymes, transcription factors need to find sites on the DNA, membrane proteins need to find the membrane, etc. Diffusion is what allows this to happen. So it is clear that diffusion is essential. What is not so clear is to what degree diffusion *rate* is important, and that is what I will be exploring here. This text is divided in two main sections. The first, "Old hat", explains some general principles about diffusion including diffusion limitation of reactions, provides a summary of experimentally determined protein diffusion coefficients in prokaryotes, and gives examples of the consequences of diffusion rates in prokaryotic cells. The second part, "New horizons", suggests some principles and experimental directions for further study.

Old hat

General principles of diffusion

As mentioned the perpetual jiggling that goes on inside of cells causes molecules to move around. Following the behavior of such a molecule, by noting its position every so often, reveals that the direction a molecule moves in at every time step is random. And, as a consequence of this, the trajectory the molecule follows is a random walk (Figure 1A). What is not random is the size of the step the molecule makes in each time interval. The step size is determined by the size of the molecule, its interactions with the solvent, and the temperature. The step size (or, more accurately, the step size distribution) is equivalent to the rate of diffusion and is captured by a single parameter, the diffusion coefficient (D). Note that there are exceptions. For example, the diffusion coefficient can be dependent on length scale, which is called anomalous diffusion (Dix *et al.*, 2008). Here we will only consider normal diffusion.

A distribution of molecules over space and its evolution in time are described by the diffusion equation (Phillips *et al.*, 2009a):

$$\frac{\partial c(x,t)}{\partial t} = D \frac{\partial^2 c(x,t)}{\partial x^2} \qquad (1)$$

Here $c(x,t)$ is the concentration of the molecule at position x and time t. D is the diffusion coefficient. This equation describes diffusion in only one dimension. The movement of the particle in one dimension is completely independent of its movement in the other dimensions. One of the solutions of this equation describes how a group of molecules localized to a point spreads out over time (Figure 1B) (Phillips *et al.*, 2009a):

$$c(x,t) = \frac{N}{\sqrt{4\pi D t}} e^{-\frac{x^2}{4Dt}} \qquad (2)$$

The parameters and variables mean the same as above, and N represents the number of molecules. This equation can also be interpreted as the probability distribution for where a single molecule is going to end up after time t. This can be done because diffusing particles do not influence each other. (Note that the molecules they may influence each other, say by changing local viscosity, but then we

are not dealing with normal diffusion anymore.) Taking the weighted mean over the distances in equation 2 and taking into account multiple dimensions (using Pythagoras' theorem) yields:

$$d = \sqrt{2nDt} \qquad (3)$$

Here d is the distance and n is the number of dimensions considered. See Figure 1C.
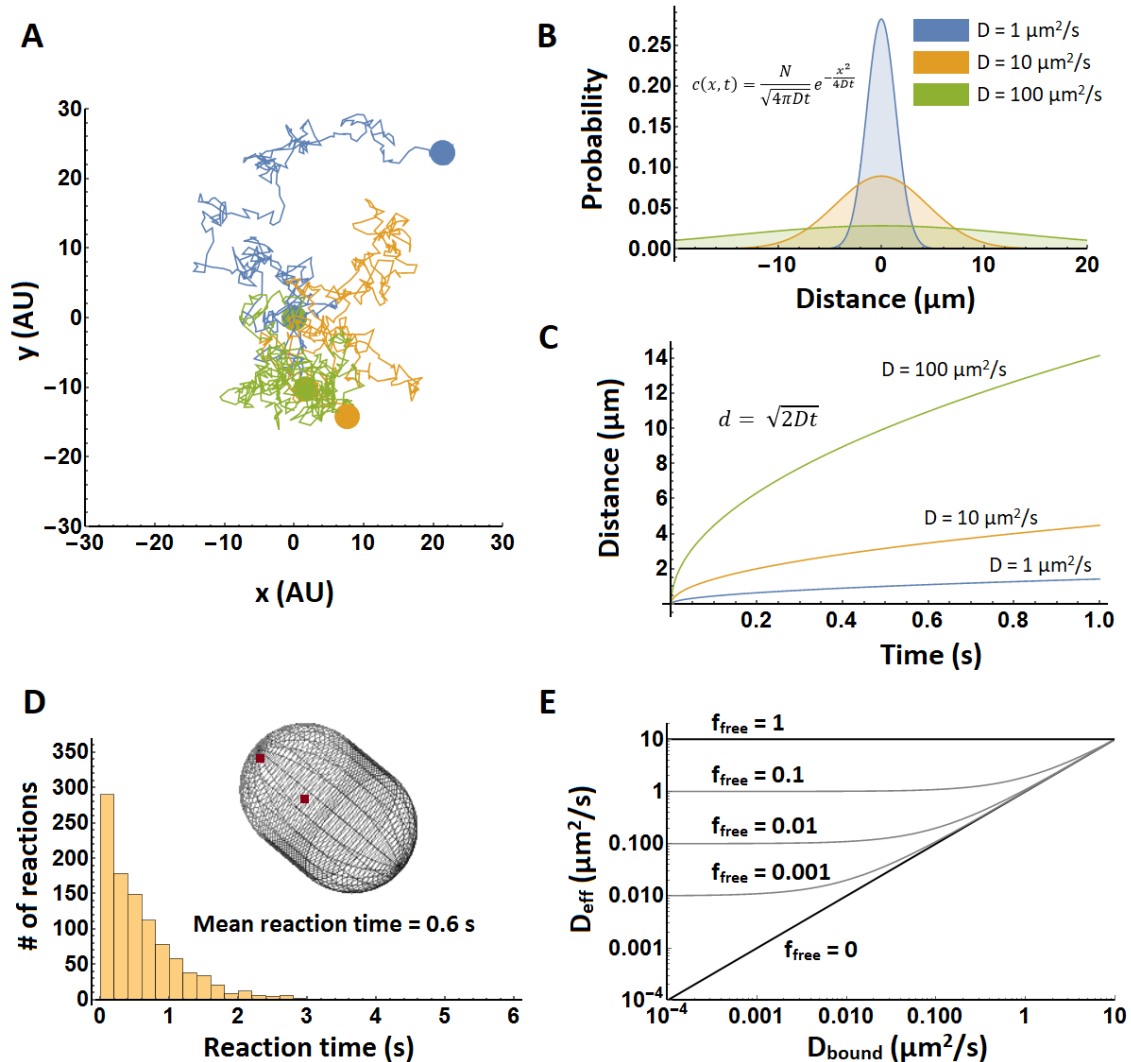
Figure 1: Illustration of diffusion principles. A) Three molecules, each in a different color, undergoing a random walk in two dimensions. Each trajectory consists of 400 steps, and beginning and end are indicated by colored spheres. All three molecules started at position (0,0). B) Probability density in one dimension for the position of a particle after 1 s. Shown are densities for three different diffusion coefficients. Computed from equation 2. C) The mean distance of a molecule in time. Shown for three different diffusion coefficients. Computed from equation 3 with the number of dimensions set to one. D) Simulation of biomolecular reaction times in a spherocylinder of 1.5 μm in length and 1 μm in width. In 1000 separate simulations two particles were positioned randomly in the spherocylinder "cytoplasm" and allowed to diffuse with a diffusion coefficient of 10 μm²/s and react with a k_{on} of 10^9 M⁻¹s⁻¹. Simulations were performed in Smoldyn (Andrews *et al.*, 2010). E) The effective diffusion coefficient of a complex forming protein as a function of bound diffusion coefficient and free fraction. It was plotted using equation 5, with D_{free} = 10 μm²/s. Note that upon binding the free protein takes on the diffusion coefficient of the object it binds to. This means that the top right corner of the graph is somewhat inaccurate.

For two particles with diffusion coefficients of 10 μm²/s to find each other in a 1 μm³ cell takes about 1 s. This can easily be calculated from the bimolecular reaction rate equation:

$$rate = k_{on}[mol1][mol2] \qquad (4)$$

The diffusion limited on-rate constant, k_{on}, is about 10^9 M^{-1}s^{-1} (calculated with Equation 6), and the molecule concentrations are about 1 nM for 1 molecule in 1 µm^3. This yields a rate of 10^{-9} Ms^{-1}, so one molecule reacts in 1 s. A similar result is obtained from a simulation of an association reaction (Figure 1D). Here the reaction was carried out in a spherocylinder of 1.5 µm in length and 1 µm in width. Reaction times are distributed between 0 and 3 s, with the mean at 0.6 s.

Inside a cell cytoplasm a protein can stick to other components of this cell and this can affect its diffusion coefficient. Note that these other components in most cases also diffuse but perhaps at a reduced rate. An effective diffusion coefficient can be calculated as follows (see Chapter 3 of this thesis):

$$D_{eff} = f_{free}\, D_{free} + \left(1 - f_{free}\right) D_{bound} \qquad (5)$$

Here D_{eff} is the effective diffusion coefficient, f_{free} is the free fraction of your protein of interest, D_{free} is the diffusion coefficient when the protein is free, and D_{bound} is the diffusion coefficient when it is bound. The result can be seen in Figure 1E.

The prime principle: diffusion limited reactions

Diffusion matters. To what degree the rate of diffusion matters depends on whether the reactions within a cell are diffusion limited. Diffusion limitation leads to changes in the rate of a process when the diffusion coefficients of the involved proteins change. The rate of an association reaction between two molecules depends on the concentrations of the participants and the on-rate constant, k_{on}, as shown in equation (4). If a reaction is diffusion limited k_{on} depends on the diffusion coefficients of the two proteins. For two spherical proteins with a completely reactive surface area the diffusion limited k_{on} is given by this equation (Schreiber *et al.*, 2009):

$$k_{on,diff} = 4\pi D R \qquad (6)$$

Here D is the sum of the diffusion coefficients and R is the sum of the radii of the two proteins. As an example let's take a protein with a radius of 0.005 µm for which the diffusion coefficient is about 10 µm^2/s in the *E. coli* cytoplasm and 100 µm^2/s in dilute solution. Putting these values in equation (6) yields k_{on}'s of about 10^8 M^{-1}s^{-1} in the cytoplasm and 10^9 M^{-1}s^{-1} in dilute solution. (Note that the right side of the equation has the unit µm^{3}s^{-1} and the left side M^{-1}s^{-1}. This is because on the right side we are dealing with single molecules and on the left side we are dealing with moles. You can convert µm^3/s into M^{-1}s^{-1} by dividing by 10^{15} to convert the volume and then multiplying by Avogadro's number (6×10^{23}).) Most proteins are not reactive over their entire surface and more realistic diffusion limited on-rate constants are 10^5-10^6 M^{-1}s^{-1} (Schlosshauer *et al.*, 2004; Schreiber *et al.*, 2009). The diffusion limit is best understood as a range rather than a point. Because having multiple binding sites on a protein or having electrostatic interactions steer an association can increase the k_{on} beyond 10^5-10^6 M^{-1}s^{-1}.

We just now considered diffusion limitation when only two proteins are involved, and we assume these proteins are both constantly present and reactive. When we deal with an interaction in the cell we also have to take into account other processes. For example: the synthesis of proteins, the release from other complexes, transport over membranes, or the cycling through conformational states of one of the binding partners. Let's go through an example in some detail.

If two proteins, A and B, are reactive over their entire surface they will form a complex as soon as they hit. If they have small reactive patches on the surface they will hit each other many times before

forming a complex. In both cases the faster the proteins bump into each other the faster the complex is formed. The situation changes when protein A cycles through two states, of which only one (active) is able to form the complex. As an illustration we let A spend an average of 10 s in the inactive state and 10 s in the active state. The average time for A and B to find each other is 1 s. The inactive A is hit by B on average ten times before it switches to the active state. When protein A finally does switch to the active state, B binds on average in 1 s. This gives a reaction time of 11 s. If B were diffusing twice as fast this reaction time would only go down to 10.5 s. If B were diffusing twice slower the reaction time would be 12 s. Thus the diffusion rate of B is (relatively) inconsequential and the reaction is not diffusion limited. This is only true if the active period of A is significantly longer than the time for A and B to bump into one another. If the active period is 0.5 s, B could miss its chance to interact with A, and only manage to react in the next active period. This makes the reaction-time sensitive to the diffusion rate of B; though the reaction time will be much longer than the time A and B need to find each other. A real world example of the diminished importance of the cytoplasmic diffusion rate is the binding of the transcription inhibitor LacI to its DNA target site. In the search process for its proper binding site LacI non-specifically binds the DNA and scans it, which takes up about 90 % of the search time (Li *et al.*, 2011). If the LacI would diffuse infinitely faster through the cytoplasm this would only reduce the total search time down to 90 % of the actually measured time.

Diffusion limitation may be different for catalysis and protein-protein interactions. Diffusion limitation has been described as resulting in concentration gradient of reactants (Berg *et al.*, 1985). Say you have an enzyme (the sink) and a homogenous distribution of reactants (the source). A diffusion limited enzymatic reaction will deplete the surroundings of the enzyme so you have a substrate gradient. The use of this picture is not clear when complex formation is concerned. Let us consider a pair of proteins, A (sink) and B (source). When one of B finds A the complex is formed and A is gone. So there is no sink around which to form a gradient. Of course A needs to be synthesized and the spot of synthesis may become the sink. However, proteins are typically made from multiple ribosomes and mRNA, which are located all over the cell. Making it difficult for gradients to form. The use of the gradient description depends on the biological context. It could work well for the process of translation where association between molecules leads to a reaction (Zhang *et al.*, 2010; Klumpp *et al.*, 2013), or when a membrane has the function of a sink (Schulz *et al.*, 2001).

What are the diffusion coefficients of proteins in cells?

Here I provide an overview of translational diffusion coefficients of proteins in prokaryotes. For comparison I have also included diffusion coefficients of a small molecule in *Escherichia coli*, proteins in some eukaryotes, and proteins in dilute solution. See Table 1. Diffusion rates have been measured for proteins in the cytoplasm, plasma membrane, periplasm, and outer membrane. Most diffusion rates have been determined for proteins in *Escherichia coli*, but a decent amount of data is also available for the bacteria *Caulobacter crescentus*, *Pseudomonas aeruginosa*, and *Lactococcus lactis*. Most diffusion rates have been determined with fluorescence recovery after photo-bleaching (FRAP), some are determined by single particle tracking (SPT) or fluorescence correlation spectroscopy (FCS). For a short description of these techniques see (Mika *et al.*, 2011).

The values represented in Table 1 are means or medians over populations of cells. Typically there is considerable variation in the diffusion coefficient between cells (Konopka *et al.*, 2009; Mika *et al.*, 2014). We illustrate this in Figure 2 where we show histograms of the diffusion coefficients of GFP, β-galactosidase-GFP, and LacS-GFP. Also, not all diffusive processes can be described by a single diffusion coefficient. In some cases the molecules are confined (Fukuoka *et al.*, 2007) or exhibit anomalous diffusion (Golding *et al.*, 2006).

Comparison of the GFP (or mCherry) diffusion rates in the cytoplasm of *C. crescentus* (8 μm^2/s), *P. aeruginosa* (4 μm^2/s), *L. lactis* (7 μm^2/s), and the archaeon *Hfx. volcanii* (5.5 μm^2/s) reveals some differences. However, all of these diffusion coefficients fall within the *E. coli* range (3-14 μm^2/s). So it is not clear whether these differences are real, they may be due to measurement error, or different growth and measurement conditions. What's more, since these organisms live in different environments it is not even clear how to make a general comparison of these diffusion rates, or if that's even desirable. Continuing the comparisons. In *E. coli* and *L. lactis* the diffusion rate of β-galactosidase-GFP is virtually the same. There does appear to be a difference in the diffusion coefficient of the ribosome between *E. coli* (0.04 μm^2/s) and *C. crescentus* (0.0002-<0.0011 μm^2/s). However, there is some inconsistency in the *C. crescentus* numbers for the free ribosomes suggesting that at least some of the values are not entirely accurate. For membrane proteins we have again an agreement between *E. coli* and *L. lactis* which have similar diffusion coefficients for membrane proteins with 12 transmembrane helices. And again we have a disagreement between *E. coli* (0.18-0.22 μm^2/s) and *C. crescentus* (0.012 μm^2/s) for membrane proteins with 4 transmembrane helices.

It is not just the isolated values listed in Table 1 that matter, we also need to consider how diffusion values systematically vary in different contexts and with protein properties (Figure 2B-G). Protein diffusion rates go down with molecular weight of the protein, both in dilute solution and in the *E. coli* cytoplasm (Figure 2B). This is also seen for membrane proteins in relation to their (membrane embedded) radius in giant unilamellar vesicles (GUVs) and the *E. coli* plasma membrane (Figure 2C). Increasing the salt concentration of the outside medium reduces the water content of *E. coli* cells and increases the volume fraction that is excluded by macromolecules. When the salt concentration is increased slowly (adapted) the diffusion coefficient drops less fast with excluded volume fraction than when this is done swiftly (shocked) (Figure 2D). The drop in diffusion coefficient with osmotic shock severity is less fast for *L. lactis* than for *E. coli* (Figure 2E). The drop in diffusion coefficient with relative cell volume (after shock) is much more severe in *L. lactis* than in *E. coli* (Figure 2F). The diffusion coefficient of different surface modified variants of GFP depends on their net charge, with positive protein diffusing up to a 100-fold slower. This effect is strongest in *E. coli* but is also present in *L. lactis* and the archaeon *Haloferax volcanii* (Figure 2G).

Table 1: Overview of experimentally determined diffusion coefficients.

Molecule	Organism	Diffusion coefficient (D; μm^2/s)	Comments/References
NBD-glucose	*Escherichia coli*	50	0.423 kDa; (Mika *et al.*, 2010)
GFP	Dilute solution	87	27 kDa; (Potma *et al.*, 2001)
GFP	*Dictyostelium discoideum*	24	Cytoplasm; (Potma *et al.*, 2001)
GFP	*Mus musculus*	27	Fibroblast cytoplasm; (Swaminathan *et al.*, 1997)
GFP	*Escherichia coli*	3-14	Cytoplasm; (Konopka *et al.*, 2009; Mika *et al.*, 2011)
GFP	*Lactococcus lactis*	7	Cytoplasm; (Mika *et al.*, 2014)
GFP	*Bacillus subtilis*	>1	Cytoplasm, germinated spores; (Cowan *et al.*, 2003)
GFP	*Bacillus subtilis*	~0.0001	Spore cytoplasm; (Cowan *et al.*, 2003)
GFP	*Caulobacter crescentus*	8	Cytoplasm; (Llopis *et al.*, 2012)
GFP	*Haloferax volcanii*	5.5	Cytoplasm; this thesis Ch. 3
mCherry	*Pseudomonas aeruginosa*	4	Cytoplasm; (Guillon *et al.*, 2013)

TorA-GFP	*Escherichia coli*	9	Cytoplasm, in ΔtatABCDE strain; (Mullineaux *et al.*, 2006)
PtsH-YFP	*Escherichia coli*	3.8	Cytoplasm, 36 kDa, some degradation of the protein; (Kumar *et al.*, 2010)
CheY-GFP	*Escherichia coli*	4.6	Cytoplasm; (Cluzel *et al.*, 2000)
Crr-YFP	*Escherichia coli*	2.0	Cytoplasm, 45 kDa, some degradation of the protein; (Kumar *et al.*, 2010)
NlpA$_{noLB}$-GFP	*Escherichia coli*	2.7	Cytoplasm, 55 kDa; (Nenninger *et al.*, 2010)
TorA-GFP2	*Escherichia coli*	8.3	Cytoplasm, 57 kDa, 2x GFP in tandem; (Nenninger *et al.*, 2010)
AmiA$_{noSP}$-GFP	*Escherichia coli*	7.1	Cytoplasm, 58 kDa; (Nenninger *et al.*, 2010)
CFP-CheW-YFP	*Escherichia coli*	1.5	Cytoplasm, 72 kDa, some degradation of the protein; (Kumar *et al.*, 2010)
MBP-GFP	*Escherichia coli*	2.5	Cytoplasm, 72 kDa; (Elowitz *et al.*, 1999)
torA-GFP3	*Escherichia coli*	6.3	Cytoplasm, 84 kDa, 3x GFP in tandem; (Nenninger *et al.*, 2010)
CFP-CheR-YFP	*Escherichia coli*	1.7	Cytoplasm, 86 kDa, some degradation of the protein; (Kumar *et al.*, 2010)
DnaK-YFP	*Escherichia coli*	0.67	Cytoplasm, 96 kDa, some degradation of the protein; (Kumar *et al.*, 2010)
torA-GFP4	*Escherichia coli*	5.5	Cytoplasm, 111 kDa, 4x GFP in tandem; (Nenninger *et al.*, 2010)
torA-GFP5	*Escherichia coli*	2.8	Cytoplasm, 138 kDa, 5x GFP in tandem; (Nenninger *et al.*, 2010)
HtpG-YFP	*Escherichia coli*	1.7	Cytoplasm, dimer of 198 kDa; (Kumar *et al.*, 2010)
CFP-CheA-YFP	*Escherichia coli*	0.44	Cytoplasm, 250 kDa, some degradation of the protein; (Kumar *et al.*, 2010)
LacI-Venus	*Escherichia coli*	3	Cytoplasm, tetramer of ~260 kDa, freely diffusing, when DNA binding is included D = 0.4 µm^2/s; (Elf *et al.*, 2007)
β-galactosidase	Dilute solution	31	Tetramer of 466 kDa; (Hahn *et al.*, 2006)
β-galactosidase-GFP	*Escherichia coli*	0.7	Cytoplasm, tetramer of 582 kDa; (Mika *et al.*, 2010)
β-galactosidase-GFP	*Lactococcus lactis*	0.8	Cytoplasm, tetramer of 582 kDa; (Mika *et al.*, 2014)
Ribosome	*Escherichia coli*	0.04	Cytoplasm, fully active, includes all states of translation; (Bakshi *et al.*, 2012)
Ribosome (free, 30S)	*Escherichia coli*	0.6	Cytoplasm, freely diffusing, 1 MDa; (Bakshi *et al.*, 2012)
Ribosome (bound)	*Escherichia coli*	0.055	Cytoplasm, bound fraction; (Sanamrad *et al.*, 2014)
Ribosome (free, 30S or 50S)	*Escherichia coli*	0.4	Cytoplasm, free fraction; (Sanamrad *et al.*, 2014)
Ribosome (bound)	*Caulobacter crescentus*	0.0002-<0.0011	Cytoplasm, obtained from model that includes a bound and free fraction; (Llopis *et al.*, 2012)

Ribosome (free, 50S)	*Caulobacter crescentus*	0.018-0.042	Cytoplasm, obtained from model that includes a bound and free fraction; (Llopis *et al.*, 2012)
Ribosome (free, 50S)	*Caulobacter crescentus*	0.36-0.39	Cytoplasm, after cells were treated with rifampicin or kasugamycin; (Llopis *et al.*, 2012)
Carboxysome	*Synechococcus elongatus*	0.000046	Cytoplasm, constrained movement; (Savage *et al.*, 2010)
mRNA	*Escherichia coli*	0.001-0.03	Cytoplasm, diffusion is anomalous, mRNA in complex with many copies of MS2-GFP; (Golding *et al.*, 2004; Golding *et al.*, 2006)
DNA	*Escherichia coli*	0.0004-0.0007	Chromosomal loci, apparent D as DNA doesn't move freely; (Reyes-Lamothe *et al.*, 2008)
PvdS-eYFP	*Pseudomonas aeruginosa*	1	Cytoplasm, PvdS is a sigma factor; (Guillon *et al.*, 2013)
PvdA-eYFP	*Pseudomonas aeruginosa*	0.5	Cytoplasm; (Guillon *et al.*, 2013)
PvdQ-mCherry	*Pseudomonas aeruginosa*	0.2	Periplasm; (Guillon *et al.*, 2013)
FpvF-mCherry	*Pseudomonas aeruginosa*	0.2	Periplasm; (Guillon *et al.*, 2013)
GFP	*Escherichia coli*	2.6	Periplasm; TorA signal sequence removed upon export to periplasm (Mullineaux *et al.*, 2006)
MotB-GFP	*Escherichia coli*	0.0075-0.0088	Plasma membrane, freely diffusing, dimer; (Leake *et al.*, 2006)
TatA-GFP	*Escherichia coli*	0.13	Plasma membrane; (Mullineaux *et al.*, 2006)
Tar(1-397)-YFP	*Escherichia coli*	0.22	Plasma membrane, 4 transmembrane helices; (Kumar *et al.*, 2010)
Tsr(1-218)-YFP	*Escherichia coli*	0.18	Plasma membrane, 4 transmembrane helices; (Kumar *et al.*, 2010)
LacY-YFP	*Escherichia coli*	0.027	Plasma membrane, 12 transmembrane helices; (Kumar *et al.*, 2010)
MtlA-YFP	*Escherichia coli*	0.028	Plasma membrane, 12 transmembrane helices; (Kumar *et al.*, 2010)
Tar-YFP	*Escherichia coli*	0.017	Plasma membrane, 12 transmembrane helices; (Kumar *et al.*, 2010)
TetA-YFP	*Escherichia coli*	0.09	Plasma membrane, 12 transmembrane helices; (Chow *et al.*, 2012), see also discussion in (Mika *et al.*, 2014)
NagE-YFP	*Escherichia coli*	0.020	Plasma membrane, 16 transmembrane helices; (Kumar *et al.*, 2010)
FliG-GFP	*Escherichia coli*	0.0049	Attached to flagellum basal body; (Fukuoka *et al.*, 2007)
BcaP-GFP	*Lactococcus lactis*	0.02	Plasma membrane, 12 transmembrane helices; (Mika *et al.*, 2014)

LacSΔIIA-GFP	*Lactococcus lactis*	0.02	Plasma membrane, 12 transmembrane helices; (Mika *et al.*, 2014)
PleC-eYFP	*Caulobacter crescentus*	0.012	Plasma membrane, freely diffusing, 4 transmembrane helices; (Deich *et al.*, 2004)
Lipopolysaccharide	*Salmonella typhimurium*	0.02	Outer membrane; (Schindler *et al.*, 1980)
BtuB	*Escherichia coli*	0.05-0.10	Outer membrane, 22-stranded β-barrel, when disconnected from its binding partner TonB D = 0.27 µm^2/s; (Spector *et al.*, 2010)
OmpF	*Escherichia coli*	0.006	Outer membrane, trimer, 16-stranded β-barrel, diffusion is restricted to area with 100 nm diameter; (Spector *et al.*, 2010)
LamB (λ-receptor)	*Escherichia coli*	0.15	Outer membrane, LamB appears to be tethered and diffusion is restricted to area with 50 nm diameter; (Oddershede *et al.*, 2002)

Figure 2: Systematic variation of diffusion coefficients with protein and environment properties. A) Variation of diffusion coefficient within a population of cells for the proteins GFP and β-galactosidase-GFP (tetramer) in the cytoplasm, and LacSΔIIA-GFP in the membrane of *Lactococcus lactis* (Mika *et al.*, 2014). B) The dependence of diffusion coefficient on molecular weight in dilute solution (Tyn *et al.*, 1990) and the *Escherichia coli* cytoplasm (Elowitz *et al.*, 1999; van den Bogaart *et al.*, 2007; Konopka *et al.*, 2009; Kumar *et al.*, 2010; Mika *et al.*, 2010; Nenninger *et al.*, 2010; Bakshi *et al.*, 2012). C) The dependence of diffusion coefficient on radius of the membrane spanning part of membrane proteins in giant unilamellar vesicles (GUVs) (Ramadurai *et al.*, 2009) and in the *Escherichia coli* plasma membrane (Kumar *et al.*, 2010). The radii for the proteins studied in the *E. coli* membrane are calculated from the number of transmembrane helices (Kumar *et al.*, 2010) and the radius of a single helix peptide reported in (Ramadurai *et al.*, 2009). D) The dependence of diffusion coefficient of cytoplasmic GFP on excluded volume fraction in adapted and shocked *Escherichia coli* cells (Konopka *et al.*, 2009). E) The dependence of the diffusion coefficient of cytoplasmic GFP on shock medium osmolality after osmotic shock for *Escherichia coli* and *Lactococcus lactis* (Konopka *et al.*, 2009; Mika *et al.*, 2014). The growth medium had the same osmolality as the first points on the graph. F) The dependence of the diffusion coefficient of cytoplasmic GFP on the relative cell volume after osmotic shock in Escherichia coli and *L. lactis* (Mika *et al.*, 2014). G) The dependence of the cytoplasmic diffusion coefficient of GFP variants on their net charge in *Escherichia coli*, *Lactococcus lactis* and *Haloferax volcanii* (see Chapter 3 of this thesis). There is no data for -30 GFP in *L. lactis*.
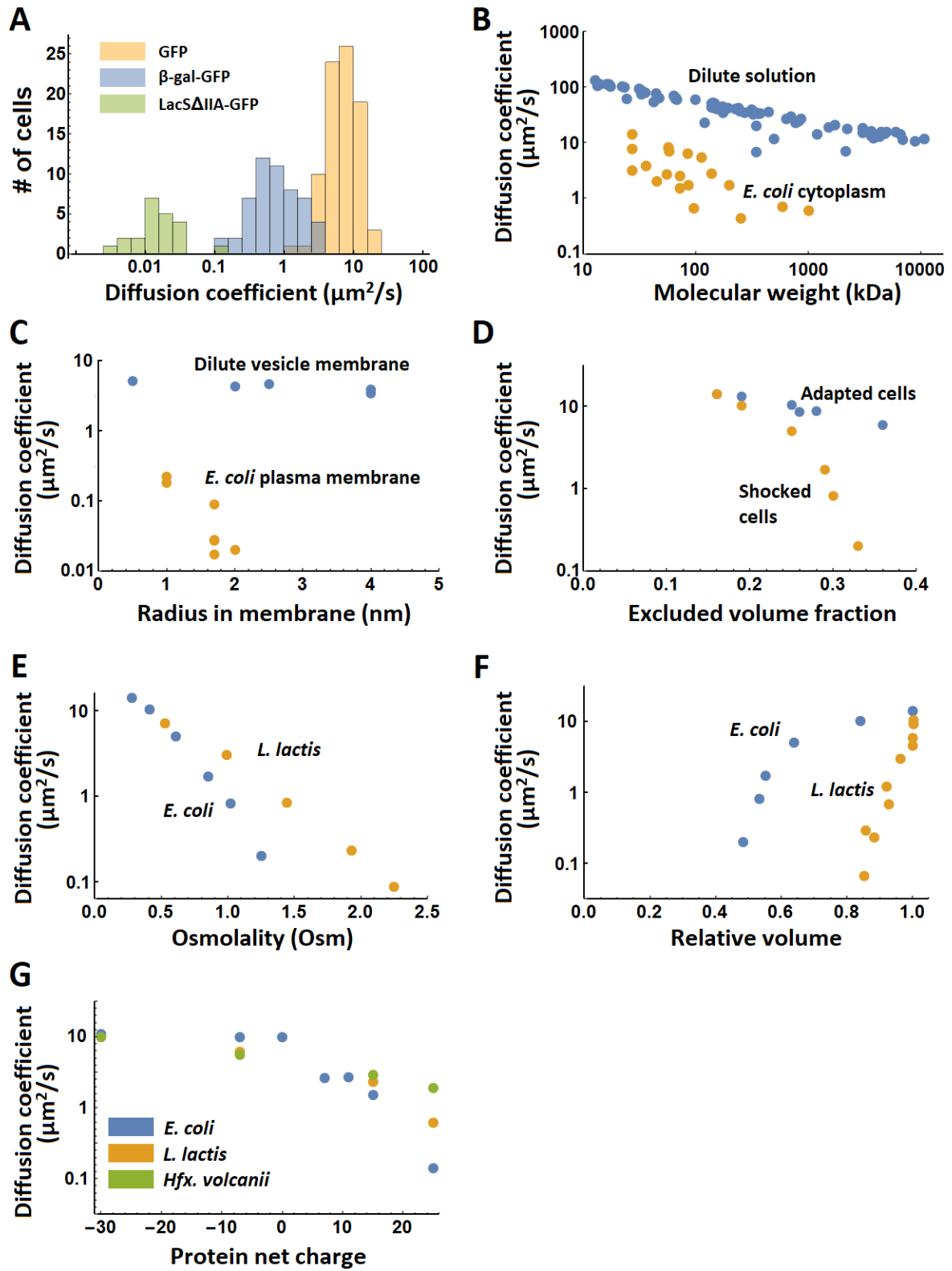
Examples of diffusion limitation in prokaryotes

We can, in our mind's eye, lower the diffusion coefficient of a protein indefinitely; doing this would cause any reaction to become diffusion limited at some point. So a study of diffusion rates and diffusion limitation of processes is pertinent. Despite many claims of the importance of diffusion rates in prokaryotic cells, there are not many examples where this has actually been demonstrated. Here I will summarize some cases where diffusion limitation seems to occur. Some more discussion of diffusion processes in prokaryotes can be found in (Soh *et al.*, 2010).

The on-rate constant of Barnase-Barstar goes beyond the diffusion limit. Consider a pair of proteins that is able to form a complex. The diffusion limited k_{on} starts at 10^5-10^6 $M^{-1}s^{-1}$ (Schlosshauer *et al.*, 2004; Alsallaq *et al.*, 2008). Yet some protein pairs manage to have an on-rate constant of 10^8-10^{10} $M^{-1}s^{-1}$ (Schreiber *et al.*, 1993; Wallis *et al.*, 1995; Alsallaq *et al.*, 2008). One of these pairs is Barnase-Barstar from *Bacillus amyloliquefaciens*. Barnase is an extracellular ribonuclease which is bound by Barstar in the cytoplasm to prevent damage to RNA (Buckle *et al.*, 1994). The fact that the reaction is electrostatically steered, and that the on-rate constant is two orders of magnitude higher than the non-electrostatic diffusion limit, suggests that diffusion rate is important for this reaction. Another protein pair with a very high on-rate constant is ColicinE9-Im9. ColicinE9 is a secreted toxin with DNase activity. Again its binding partner, Im9, is used to prevent damage in the cytoplasm where ColicinE9 is made (Wallis *et al.*, 1995). Note that the increase in on-rate constant could be there to make the complex bind more tightly rather than increase the on-rate per se. A direct determination of diffusion limitation has not been carried out.

There are other bacterial proteins that form complexes with high k_{on}'s, though this is not always demonstrated with the physiological binding partner. SecB from *E. coli* was shown to interact with BPTI with a k_{on} of 5×10^9 $M^{-1}s^{-1}$ (Fekkes *et al.*, 1995).The chaperone complex GroEL interacts with the unfolded state of barnase with a k_{on} of 0.35-1.8x10^8 $M^{-1}s^{-1}$ (Gray *et al.*, 1993; Perrett *et al.*, 1997), MBP with a k_{on} of 0.9-7.0x10^7 $M^{-1}s^{-1}$ (Sparrer *et al.*, 1996), and DHFR with a k_{on} of 3×10^7 $M^{-1}s^{-1}$ (Clark *et al.*, 1997). It is however not clear whether these GroEL interactions are really diffusion limited because the unfolded proteins provide many more interaction opportunities than folded proteins so the limit of 10^5-10^6 $M^{-1}s^{-1}$ may not apply.

Translation and cell growth rate are limited by charged tRNA availability. To grow, cells need to produce proteins. As such, the rate of growth could be limited by the rate of protein production. The total rate of protein production is set by the number of ribosomes and how fast they can start and end the production of one protein, and, more important to this discussion, by how fast they can elongate the proteins. (Note that in individual cases protein production can be limited by ribosome binding site strength rather than elongation rate.) Elongation consists of the arrival of ternary complex, a complex that consists of aa-tRNA, EF-Tu and GTP, and its processing by the ribosome. Using a computational model of the translation process it was found that if many ribosomes (100) are translating a stretch of the same amino acids the rate per codon was decreased because of diffusion limitation. The effect was exacerbated when the diffusion rate was decreased after simulating an osmotic shock (Zhang *et al.*, 2010). It is not clear whether this diffusion limitation is present at actual cellular conditions and amino acid sequences.

In another study (Klumpp *et al.*, 2013) translation was also found to be diffusion limited in its rate. In the calculations Michaelis-Menten kinetics was assumed for amino acid incorporation. The K_M was calculated under the assumption that the reaction is diffusion limited. They estimated a diffusion coefficient of 1 μm^2/s for the ternary complex, from which they determine the k_{on} for binding of ternary complex to ribosome to be 10^7 $M^{-1}s^{-1}$. The rate of going from the ternary complex being bound

to amino acid chain elongation, k_{elong}, is set at 30 s^{-1}. From this they calculate that the diffusion limited K_M is 3 μM. This is then compared to the concentrations of tRNA in *E. coli*, 3-30 μM. The fact that the concentrations of tRNA are equal or higher than the diffusion limited K_M is taken to mean the process functions at a diffusion limited rate. There are some problems with this calculation. The estimate of the diffusion limited k_{on} is made assuming that Equation 6 is valid. To derive this equation it is assumed that the molecules that react can have any orientation upon collision and react immediately. This is unlikely to be the case. Diffusion limited k_{on}'s are also not necessarily single values as electrostatic interactions may steer the interaction and make the reaction faster. Another issue is the estimate of k_{elong}. All else being the same if k_{elong} would be ten times higher the calculated K_M would be ten times higher and the argument would not make sense. They don't actually know the value of k_{elong}. However, let's for the moment assume that the calculations are correct. They next made a model that takes into account allocation of resources to different parts of the proteome. The translation speed is limited by the ternary complex association rate, which depends on both k_{on} and the concentrations of the ternary complex and the ribosome. Putting more resources in increasing the concentration of ternary complex limits the number of resources that can be put into ribosomes. The cell growth rate is a function of both translation speed and ribosome concentration. Thus cell growth rate and allocation of resources are influenced by ternary complex diffusion rate.

Something else to consider here is what environment the *E. coli* cell optimized its macromolecule diffusion coefficients and copy numbers for. If it frequently encounters osmotic stress (and thus large changes in excluded volume and ionic strength) it may tune itself differently than when osmotic stress is rare.

The combination of cell size and protein concentration in prokaryotic and eukaryotic cells is optimized to facilitate rapid diffusion. Say you hold the number of proteins in an *E. coli* cell constant but would vary its cell size, making it smaller or larger. If you make the cell smaller the distances that need to be overcome by diffusion are smaller but the obstacles (crowding) are more severe so diffusion is slower. If you make the cell bigger the distances become larger but diffusion becomes faster. This scenario has been turned into a quantitative model which shows that for prokaryotes the cell diameter is predicted to be 1.1 μm and for eukaryotes 15.7 μm at the smallest characteristic diffusion times (Soh *et al.*, 2013). It is claimed that these diameters are comparable to the typical sizes of the cells of these groups of organisms. Which means that the combination of cell size and macromolecule concentration is optimized for rapid diffusion, which means that there are diffusion limited processes in these cells. Note that this prediction of cell size depends on the number of proteins in these cells which are $3x10^6$ for prokaryotes and $8x10^9$ for eukaryotes. In the study it is mentioned that there model provides an argument for determining what the sizes of prokaryotic and eukaryotic cells should be. However no argument is provided that the number of proteins ought to be $3x10^6$ and $8x10^9$. Another problem with their estimate is that they take the characteristic distance that diffusion needs to bridge to be the size of the cell. For many reactions the targets are probably much closer.

Differences in diffusion rate leading to functional differences

Above I discussed some cases in which diffusion rates limit rates of other processes in the cell. These consequences of the diffusion rates are essentially efficiency improvements, they do not arbitrate on the existence of phenomena. Here I will give two examples in which diffusion rates make a functional difference. Phenomena that wouldn't exist were it not for certain diffusion rates.

The min system oscillation in Escherichia coli. Cell division in *E. coli* creates two equal size daughter cells. A key protein in cell division is FtsZ which forms a ring in the middle of the cell that helps to pull the cell envelope inward. The position of the FtsZ ring is determined, among other things, by the min

system (Loose *et al.*, 2011). The min system consists of the proteins MinC, MinD, and MinE. MinC inhibits FtsZ ring formation and does so only when bound to MinD. MinD and E form an oscillator that moves MinC, D, and E from one cell pole (bound to the membrane) to the other with a periodicity of about a minute. Because of this oscillator MinC spends the least time in the mid cell region so that the FtsZ ring can form. An important feature necessary to create oscillations in space is the fact that when MinD is membrane bound it has a slower diffusion rate than when it is free in solution to move to the other cell pole, that is, diffusion rates determine whether the spatiotemporal oscillation can exist.

Stable cytoplasmic protein gradients in small cells. A group of proteins can spread within seconds through a cell of several micrometers in length. Because of this, it is not likely that stable protein gradients can form over the length of the cell. However, it has been shown theoretically that protein gradients can form under special circumstances (Lipkow *et al.*, 2008). The system we consider consists of three proteins in a cell: a kinase at one of the cell poles, a phosphatase throughout the cytoplasm, and a substrate protein that can cycle between a phosphorylated and unphosphorylated state. For the system to be able to form a gradient of the substrate protein the diffusion coefficient of it two states must be different. In a 5 µm long cell, with a kinase rate constant of 10 µm/s (the system is one-dimensional hence the m rather than m³), a phosphatase rate constant of 1 s⁻¹, and diffusion coefficients of 0.3 µm²/s and 10 µm²/s for phosphorylated and unphosphorylated forms yields a tenfold concentration gradient of the substrate protein over the length of the cell. Again the difference in diffusion coefficients allows the phenomenon to exist.

New horizons

Consequences of electrostatic steering and ionic strength

Earlier I presented the case of the barnase-barstar complex formation. The diffusion limitation that this reaction labors under has been stretched by electrostatic interactions. Yet it is well known that the on-rate of this particular electrostatic interaction, and also others, diminishes with increased ionic strength (Stone *et al.*, 1989; Schreiber *et al.*, 1993; Wallis *et al.*, 1995). This means that organisms with low internal ion concentrations, such as *Escherichia coli* (Shabala *et al.*, 2009), are less affected by diffusion limitation than organisms with high internal ion concentrations, such as *Haloferax volcanii* (Pérez-Fillol *et al.*, 1986). Does this mean that organisms such as *Hfx. volcanii* are unable to use toxin-antitoxin systems like barnase-barstar? How does this affect transcription factor binding to DNA, or the assembly of ribosomes?

On temperature and crowding

All prokaryotes for which protein diffusion coefficients are known function in a small range of temperatures. How does the diffusion coefficient change if you go from 0-100 °C? We can get an estimate from the Stokes-Einstein equation:

$$D = \frac{k_B}{6\pi R} \times \frac{T}{\eta(T)} \qquad (7)$$

Here D is the diffusion coefficient, k_B is the Boltzmann constant, R is the Stokes radius of the protein, T is the absolute temperature, and $\eta(T)$ is the viscosity at temperature T. We want to know how D changes from 0-100 °C, and thus need to consider only $T/_{\eta(T)}$. For 0 °C we fill in T = 273 K and $\eta(273)$ = 1.8x10⁻³ kg s⁻¹m⁻¹, for 100 °C we fill in T = 373 K and $\eta(373)$ = 0.28x10⁻³ kg s⁻¹m⁻¹. This yields a ~9 fold faster diffusion coefficient at 100 °C. In this calculation I used the viscosity of water; thus for this result to apply to cells we have to assume that the cytoplasmic viscosity will vary as the water viscosity. This need not be the case. I have also assumed that the Stokes-Einstein equation applies in the cytoplasm

when varying over the temperature. It has been shown that the Stokes-Einstein equation does not hold in the cytoplasm for the relation between diffusion coefficient and Stokes radius (Mika *et al.*, 2011). Nonetheless the significant increase in D as calculated here portends the actual impact of temperature on the diffusion coefficient. If there indeed is such a difference in diffusion coefficients what does this mean for how cells are organized along this spectrum of temperatures? From this calculation follows the (conditional) prediction that cells at higher temperatures have higher cytoplasmic concentrations of macromolecules.

Direct measurements of diffusion limitation

All examples of diffusion limitation discussed above are indirectly determined, and rely heavily on modelling. It would be helpful to have a method for directly determining the diffusion limitation of a process. That is to vary the diffusion coefficient of one of the actors in the process and then observe whether the rate of the process changes. This is difficult to do because changing the diffusion rate can also change other aspects of the cell. Take the example of an osmotic shock which indeed changes the diffusion coefficient (Konopka *et al.*, 2009; Mika *et al.*, 2014), but firstly it does so for all big molecules, secondly it increases ion concentrations of the cytoplasm, and thirdly it increases the excluded volume and therefore affects rates and equilibria of all kinds of processes. There is a more precise option which builds upon the work presented in Chapter 3. Variants of GFP with different net surface charges diffuse at different rates. Thus one could pick a process of interest and fuse a whole range of different GFPs to one, or more, of the proteins in this process; and then observe what effect this has on the rate of the process. This method comes with its own challenges and limitations. (1) The fusion to GFP variants could change the behavior of the actor in other ways than just diffusion. (2) The change in diffusion rate is achieved by binding of GFP variants to ribosomes which also might affect the process. And finally (3) the change in diffusion rate can only be downward. However, by carefully considering which processes to study with this method you could sidestep these problems. Binding to the ribosome can reduce diffusion rates by 100-fold. Using proteins that bind to DNA could reduce the diffusion rate 10000-fold (Reyes-Lamothe *et al.*, 2008).

A note on the use of $d = \sqrt{2nDt}$

This equation is often used to indicate in what time a process can act over what distance. Yet this reflects an average and thus ignores the key characteristic of diffusion: variation of diffusion times for individual proteins. A cell could exploit this variation by using more proteins to send a signal. If you need concentration x at point A for a signal to be effective you could increase the rate by having more signaling proteins start at point B. It would be interesting to see if this principle could in part explain, for example, the concentrations of two component signaling systems (Capra *et al.*, 2012) in the membranes of bacteria.

Cell size and diffusion length scales

The enormous panoply of prokaryotic species has within itself also a great range of cell sizes. With on the smaller end for example the Archaeon *Thermodiscus*, with a volume of 3×10^{-3} μm^3, and the bacterium *Mycoplasma pneumoniae*, 5×10^{-3} μm^3. On the larger end we have the bacteria *Epulopiscium fishelsoni*, 3×10^6 μm^3, and *Thiomargarita namibiensis*, 2×10^8 μm^3 (Schulz *et al.*, 2001). Somewhat counterintuitively both small and large sized could pose challenges for diffusion. For large size the challenge is obvious; nutrients have to reach parts of the cell from outside of the cell, and proteins have to reach parts of the cell from the chromosome (via mRNA). In *Epulopiscium fishelsoni* and *Thiomargarita namibiensis* this appears to be solved by having many chromosomes, and having them packed against the membrane of the cell. The challenge for the small cells derives from their DNA.

Escherichia coli has 4.6 Mbp of DNA in a single chromosome (Blattner *et al.*, 1997) and has a volume of about 1 µm^3 (Taheri-Araghi *et al.*, 2015); *Mycoplasma genitalium* has 0.58 Mbp of DNA (Fraser *et al.*, 1995) and has a volume of about 0.01 µm^3 (Taylorrobinson, 1995) (here I assume spherical shape for *M. genitalium*, in reality the cells are pear shaped). The chromosome copy number in *E. coli* depends on growth rate (Stokke *et al.*, 2012), as does its volume (Taheri-Araghi *et al.*, 2015). For the following I'm assuming that the chromosome copy numbers are the same for *E. coli* and *M. genitalium*. The *M. genitalium* volume is 100 times smaller than that of *E. coli*, whereas its genome is only 8 times smaller; leading to a 12.5 times higher concentration of DNA. In *E. coli* the DNA constitutes 3.1 % of dry weight, compared to 55 % for protein and 20.4 % for RNA (Phillips *et al.*, 2009b). So DNA makes up 3.9 % of the macromolecules. Multiplying 3.9 % by 12.5 gives 49 % (that is an extra 45 %), so if the protein and RNA content is still the same we have 1.45 times the amount of macromolecule in *M. genitalium* than in *E. coli*. The consequence that this (potential) difference in volume exclusion has on diffusion rates is unclear. For example when excluded volume is altered by osmotic shocks the effect on diffusion rate appears to be totally different in *E. coli* than in *L. lactis* (see Figure 2F). Of course the distance between any point in the cytoplasm and the outside of the cell is smaller in *M. genitalium* than in *E. coli*, and therefore diffusion is more effective in delivering molecules. However, this distance benefit (in travel time) scales only with the power two (here 21-fold; see Equation 3) whereas the increase in DNA excluded volume scales with the power three (here 100-fold). No studies of diffusion rate in prokaryotes have looked at its variation, or lack thereof, along the cell size axis.

In the previous paragraph I mentioned the travel of a molecule from the membrane to somewhere in the cytoplasm. Some characteristic value can be assigned to this travel distance. For example the average distance of a point of cytoplasm to the cell membrane, which will be somewhat less than half the radius. There may also be other distances to consider, for example the average distance between a gene and the membrane or a point in the cytoplasm; the average distance between ternary complex and ribosomes; or the average distance between some position in the cytoplasm and the tip of the stalk of *Caulobacter crescentus* (Young, 2006). All these various distances, and the travel times associated with them could be limiting for some process. This should be taken into account when dealing with diffusion limitation in prokaryotes. Consider a *Thiovulum majus* cell which has a diameter of 18 µm (Schulz *et al.*, 2001). If the limiting factor was the distance from a gene to someplace in the cytoplasm *T. majus* could just increase its number of chromosomes. Many bacteria are known to have increased numbers of chromosomes (Pecoraro *et al.*, 2011). Such a change in the number of chromosomes would be inconsequential if transport from cell membrane to someplace in the cytoplasm is important.

The importance of diffusion rate can't be understood in isolation

Take a look again at Equation 4, the rate of a reaction depends on the concentration of reactants and the on-rate constant. The on-rate constant depends for at least some association reactions on the diffusion coefficient. Thus, for these reactions, the rate can be tuned by changing either the concentration or the diffusion coefficient. This means that if big, and thus slow, proteins need to find each other the cell could simply have more of them around to have the same interaction rate as smaller proteins. This principle is illustrated by the example of diffusion limitation given above, about the association rate of ternary complex with ribosomes. This association rate is capped by limitations on the amount of ternary complex that can be made by a cell before other processes are adversely affected. From this we can glean another principle of diffusion in the context of a cell. If you have a process that needs a lot of proteins, such as ternary complex supplying the amino acids to the ribosome for use in translation, the impact of increasing copy number to increase association rate is tremendous. Whereas if you would want to change the association rate for transcription factor

binding to a site on the DNA, which needs only one copy (if there is one target site), this can be done without much cost. For each protein in the cell we can ask: to what degree is its copy number determined by association rate, and to what degree is it due to other factors in the process it functions in?

From the foregoing paragraph we are led into another question. Is it possible for a cell to have no diffusion limitation? Say we have x amount of stuff in a cell and no reaction is diffusion limited. With more stuff the cell is able to do more things and, for example, grow faster. So you expect there to be evolutionary pressure to increase the amount of stuff in cells, and in so doing use up the free, inconsequential, space along the diffusion rate axis. All the way up to the point that some reactions start to become limiting. This is rather similar to the discussion above about cell size and protein concentration, but looked at from a different angle. If true, this means that there will always be diffusion limitation in cells. We can also turn the argument around and ask: is it possible to have more than one process diffusion limited?

Also discussed in the examples of diffusion limitation was the protein interaction pair barnase-barstar. This reaction is made faster by electrostatic steering. Again this raises questions when dealing with a cellular context. If you wanted to increases all reaction rates in a cell, why not make them all steered by electrostatic interactions? First, one can't necessarily change a protein's surface as it could affect its function directly or its stability. Second. Is it even possible to make electrostatic interactions specific enough so that steering could be done independently for a thousand different interactions? Here again we have limitations laid upon our proteins by the cellular context.

Different parts of cells may be affected differently by diffusion

Any cell is made up of a great number of interlocking and overlapping processes. Protein folding, protein-protein binding, nutrient transport to the cytoplasm, transcription factor binding, structuring the nucleoid, inserting membrane proteins, formation of the Z-ring, Min system cycling, chromosome segregation, cell size maintenance, converting the proteome in response to environmental stress, cell cycle time, etc. For each of these processes we can ask whether they are affected, either in rate or in functional form, by the diffusion rates of their constituent proteins. There are bound to be differences between processes in their susceptibility to diffusion changes. Cell cycle time is dependent on the diffusion rate of the ternary complex, whereas the cycling rate of the Min system is independent of the cytoplasmic diffusion rate of the Min proteins. Processes that require bigger proteins may suffer more from diffusion limitation than processes with small proteins (see Figure 2B). Objects that have a size in the tens of nanometers may also experience other types of mobility (Parry *et al.*, 2014), and processes involving them may thus also be affected. Whether a proteins is folded or disordered also seems to have an effect on its diffusion rate, with a disordered protein diffusing faster than a folded protein in the presence of artificial crowders (Wang *et al.*, 2012). Something discussed earlier was different ranges over which diffusion occurs: translation happens all throughout the cytoplasm with shorter distances between ternary complex and ribosome than, for example, for a two component signaling system that needs to cross the distance between the membrane and a site on the DNA. Different processes are made up of such basic elements in different proportions and may thus be differently affected by changes in diffusion rate. Changes in diffusion rate can happen in real life situations for example after an osmotic shock that reduces the cytoplasmic water content. To know what the impact of such an event is we have to know which processes are vulnerable. More generally we can ask for each process by how many fold the diffusion coefficient needs to go down before this process becomes diffusion limited.

The reach of diffusion limitation

Cellular processes are layered. The association rate of ternary complex binding to a ribosome is involved in the time of incorporation of a single amino acid into a polypeptide chain (chain elongation); the rate of chain elongation figures in the rate of protein production; this in turn determines the rate of accumulation of biomass and cell volume growth; together with other processes this sets the cell cycle time. Here we have layer upon layer upon layer of process. At each step the diffusion limitation that is present at a lower layer could be made inconsequential for higher layers by some other, slower, process coming into the fold. How far a diffusion limited process is affecting processes above it is the reach of this limitation. And this reach should be considered when determining the importance of a diffusion limited protein-protein interactions.

Conclusion

Protein diffusion coefficients have been determined in the prokaryotes *E. coli*, *L. lactis*, *C. crescentus*, *P. aeruginosa*, *Hfx. volcanii*, and others. With *E. coli* being the best studied prokaryote by far. The protein diffusion coefficients have been measured in the cytoplasm, periplasm, plasma membrane, and outer membrane. Various parameters of both proteins and their environment have been compared systematically to the diffusion rate. For example protein size, protein surface charge, cytoplasmic ionic strength, and level of excluded volume in the cytoplasm. Multiple studies have also been carried out on the importance of the diffusion rate in the context of protein toxins, translation, and the level of excluded volume in cells. Yet despite these achievements the role of diffusion rate in prokaryotic cells is still murky. In the future we may look, among other things, into the relation between diffusion rate, excluded volume, and temperature; the relation between diffusion rate, excluded volume, and cell size; the effect of different diffusion length scales on the impact of diffusion rates on processes; and the complex relation between reaction rates, diffusion coefficients, and protein concentrations. We may also want to try and determine diffusion limitation of processes directly by altering the diffusion coefficient of particular proteins and monitoring the rate of whatever process these proteins function in.

References

Alsallaq, R., and Zhou, H. (2008) Electrostatic rate enhancement and transient complex of protein-protein association. *Proteins* **71:** 320-335.

Andrews, S.S., Addy, N.J., Brent, R., and Arkin, A.P. (2010) Detailed Simulations of Cell Biology with Smoldyn 2.1. *PLoS Comput Biol* **6:** e1000705.

Bakshi, S., Siryaporn, A., Goulian, M., and Weisshaar, J.C. (2012) Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. *Mol Microbiol* **85:** 21-38.

Berg, O., and Vonhippel, P. (1985) Diffusion-Controlled Macromolecular Interactions. *Annu Rev Biophys Biophys Chem* **14:** 131-160.

Blattner, F., Plunkett, G., Bloch, C., Perna, N., Burland, V., Riley, M., *et al.* (1997) The complete genome sequence of Escherichia coli K-12. *Science* **277:** 1453-1462.

Buckle, A., Schreiber, G., and Fersht, A. (1994) Protein-Protein Recognition - Crystal Structural-Analysis of a Barnase Barstar Complex at 2.0-Angstrom Resolution. *Biochemistry* **33:** 8878-8889.

Capra, E.J., and Laub, M.T. (2012) Evolution of Two-Component Signal Transduction Systems. *Annu. Rev. Microbiol.* **66:** 325-347.

Chow, D., Guo, L., Gai, F., and Goulian, M. (2012) Fluorescence Correlation Spectroscopy Measurements of the Membrane Protein TetA in Escherichia coli Suggest Rapid Diffusion at Short Length Scales. *PLoS ONE* **7:** e48600.

Clark, A., and Frieden, C. (1997) GroEL-mediated folding of structurally homologous dihydrofolate reductases. *J Mol Biol* **268:** 512-525.

Cluzel, P., Surette, M., and Leibler, S. (2000) An ultrasensitive bacterial motor revealed by monitoring signaling proteins in single cells. *Science* **287:** 1652-1655.

Cowan, A.E., Koppel, D.E., Setlow, B., and Setlow, P. (2003) A soluble protein is immobile in dormant spores of Bacillus subtilis but is mobile in germinated spores: Implications for spore dormancy. *Proc Natl Acad Sci U S A* **100:** 4209-4214.

Deich, J., Judd, E.M., McAdams, H.H., and Moerner, W.E. (2004) Visualization of the movement of single histidine kinase molecules in live Caulobacter cells. *Proc Natl Acad Sci U S A* **101:** 15921-15926.

Dix, J.A., and Verkman, A.S. (2008) Crowding effects on diffusion in solutions and cells. *Annu. Rev. Biophys.* **37:** 247-263.

Elf, J., Li, G., and Xie, X.S. (2007) Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316:** 1191-1194.

Elowitz, M.B., Surette, M.G., Wolf, P.E., Stock, J.B., and Leibler, S. (1999) Protein mobility in the cytoplasm of Escherichia coli. *J Bacteriol* **181:** 197-203.

Fekkes, P., Denblaauwen, T., and Driessen, A. (1995) Diffusion-Limited Interaction between Unfolded Polypeptides and the Escherichia-Coli Chaperone Secb. *Biochemistry* **34:** 10078-10085.

Fraser, C., Gocanye, J., White, O., Adams, M., Clayton, R., Fleischmann, R., *et al.* (1995) The Minimal Gene Complement of Mycoplasma-Genitalium. *Science* **270:** 397-403.

Fukuoka, H., Sowa, Y., Kojima, S., Ishijima, A., and Homma, M. (2007) Visualization of functional rotor proteins of the bacterial flagellar motor in the cell membrane. *J Mol Biol* **367:** 692-701.

Golding, I., and Cox, E. (2004) RNA dynamics in live Escherichia coli cells. *Proc Natl Acad Sci U S A* **101:** 11310-11315.

Golding, I., and Cox, E. (2006) Physical nature of bacterial cytoplasm. *Phys Rev Lett* **96:** 098102.

Gray, T., and Fersht, A. (1993) Refolding of Barnase in the Presence of Groe. *J Mol Biol* **232:** 1197-1207.

Guillon, L., Altenburger, S., Graumann, P.L., and Schalk, I.J. (2013) Deciphering Protein Dynamics of the Siderophore Pyoverdine Pathway in Pseudomonas aeruginosa. *PLoS ONE* **8:** e79111.

Hahn, D.K., and Aragon, S.R. (2006) Intrinsic viscosity of proteins and platonic solids by boundary element methods. *J. Chem. Theory Comput.* **2:** 1416-1428.

Klumpp, S., Scott, M., Pedersen, S., and Hwa, T. (2013) Molecular crowding limits translation and cell growth. *Proc Natl Acad Sci U S A* **110:** 16754-16759.

Konopka, M.C., Sochacki, K.A., Bratton, B.P., Shkel, I.A., Record, M.T., and Weisshaar, J.C. (2009) Cytoplasmic Protein Mobility in Osmotically Stressed Escherichia coli. *J Bacteriol* **191:** 231-237.

Kumar, M., Mommer, M.S., and Sourjik, V. (2010) Mobility of Cytoplasmic, Membrane, and DNA-Binding Proteins in Escherichia coli. *Biophys J* **98:** 552-559.

Leake, M.C., Chandler, J.H., Wadhams, G.H., Bai, F., Berry, R.M., and Armitage, J.P. (2006) Stoichiometry and turnover in single, functioning membrane protein complexes. *Nature* **443:** 355-358.

Li, G., and Xie, X.S. (2011) Central dogma at the single-molecule level in living cells. *Nature* **475:** 308-315.

Lipkow, K., and Odde, D.J. (2008) Model for Protein Concentration Gradients in the Cytoplasm. *Cellular and Molecular Bioengineering* **1:** 84-92.

Llopis, P.M., Sliusarenko, O., Heinritz, J., and Jacobs-Wagner, C. (2012) In Vivo Biochemistry in Bacterial Cells Using FRAP: Insight into the Translation Cycle. *Biophys J* **103:** 1848-1859.

Loose, M., Kruse, K., and Schwille, P. (2011) Protein Self-Organization: Lessons from the Min System. *Annu. Rev. Biophys.* **40:** 315-336.

Mika, J.T., and Poolman, B. (2011) Macromolecule diffusion and confinement in prokaryotic cells. *Curr Opin Biotechnol* **22:** 117-126.

Mika, J.T., Schavemaker, P.E., Krasnikov, V., and Poolman, B. (2014) Impact of osmotic stress on protein diffusion in Lactococcus lactis. *Mol Microbiol* **94:** 857-870.

Mika, J.T., van den Bogaart, G., Veenhoff, L., Krasnikov, V., and Poolman, B. (2010) Molecular sieving properties of the cytoplasm of Escherichia coli and consequences of osmotic stress. *Mol Microbiol* **77:** 200-207.

Mullineaux, C.W., Nenninger, A., Ray, N., and Robinson, C. (2006) Diffusion of green fluorescent protein in three cell environments in Escherichia coli. *J Bacteriol* **188:** 3442-3448.

Nenninger, A., Mastroianni, G., and Mullineaux, C.W. (2010) Size Dependence of Protein Diffusion in the Cytoplasm of Escherichia coli. *J Bacteriol* **192:** 4535-4540.

Oddershede, L., Dreyer, J., Grego, S., Brown, S., and Berg-Sorenson, K. (2002) The motion of a single molecule, the lambda-receptor, in the bacterial outer membrane. *Biophys J* **83:** 3152-3161.

Parry, B.R., Surovtsev, I.V., Cabeen, M.T., O'Hem, C.S., Dufresne, E.R., and Jacobs-Wagner, C. (2014) The Bacterial Cytoplasm Has Glass-like Properties and Is Fluidized by Metabolic Activity. *Cell* **156:** 183-194.

Pecoraro, V., Zerulla, K., Lange, C., and Soppa, J. (2011) Quantification of Ploidy in Proteobacteria Revealed the Existence of Monoploid, (Mero-)Oligoploid and Polyploid Species. *PLoS ONE* **6:** e16392.

Pérez-Fillol, M., and Rodriguez-Vallera, F. (1986) Potassium ion accumulation in cells of different halobacteria. *Microbiologia* **2:** 73-80.

Perrett, S., Zahn, R., Stenberg, G., and Fersht, A. (1997) Importance of electrostatic interactions in the rapid binding of polypeptides to GroEL. *J Mol Biol* **269:** 892-901.

Phillips, R., Kondev, J., and Theriot, J. (2009a). In *Physical Biology of the Cell*. New York: Garland Science, pp. 481-512.

Phillips, R., Kondev, J., and Theriot, J. (2009b). In *Physical Biology of the Cell*. New York: Garland Science, pp. 29-73.

Potma, E., de Boeij, W., Bosgraaf, L., Roelofs, J., van Haastert, P., and Wiersma, D. (2001) Reduced protein diffusion rate by cytoskeleton in vegetative and polarized Dictyostelium cells. *Biophys J* **81:** 2010-2019.

Ramadurai, S., Holt, A., Krasnikov, V., van den Bogaart, G., Killian, J.A., and Poolman, B. (2009) Lateral Diffusion of Membrane Proteins. *J Am Chem Soc* **131:** 12650-12656.

Reyes-Lamothe, R., Possoz, C., Danilova, O., and Sherratt, D.J. (2008) Independent positioning and action of Escherichia coli replisomes in live cells. *Cell* **133:** 90-102.

Sanamrad, A., Persson, F., Lundius, E.G., Fange, D., Gynna, A.H., and Elf, J. (2014) Single-particle tracking reveals that free ribosomal subunits are not excluded from the Escherichia coli nucleoid. *Proc Natl Acad Sci U S A* **111:** 11413-11418.

Savage, D.F., Afonso, B., Chen, A.H., and Silver, P.A. (2010) Spatially Ordered Dynamics of the Bacterial Carbon Fixation Machinery. *Science* **327:** 1258-1261.

Schindler, M., Osborn, M., and Koppel, D. (1980) Lateral Diffusion of Lipopolysaccharide in the Outer-Membrane of Salmonella-Typhimurium. *Nature* **285:** 261-263.

Schlosshauer, M., and Baker, D. (2004) Realistic protein-protein association rates from a simple diffusional model neglecting long-range interactions, free energy barriers, and landscape ruggedness. *Protein Science* **13:** 1660-1669.

Schreiber, G., and Fersht, A.R. (1993) Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. *Biochemistry* **32:** 5145-5150.

Schreiber, G., Haran, G., and Zhou, H. (2009) Fundamental Aspects of Protein-Protein Association Kinetics. *Chem Rev* **109:** 839-860.

Schulz, H., and Jorgensen, B. (2001) Big bacteria. *Annu Rev Microbiol* **55:** 105-137.

Shabala, L., Bowman, J., Brown, J., Ross, T., McMeekin, T., and Shabala, S. (2009) Ion transport and osmotic adjustment in Escherichia coli in response to ionic and non-ionic osmotica. *Environ Microbiol* **11:** 137-148.

Soh, S., Banaszak, M., Kandere-Grzybowska, K., and Grzybowski, B.A. (2013) Why Cells are Microscopic: A Transport-Time Perspective. *J. Phys. Chem. Lett.* **4:** 861-865.

Soh, S., Byrska, M., Kandere-Grzybowska, K., and Grzybowski, B.A. (2010) Reaction-Diffusion Systems in Intracellular Molecular Transport and Control. *Angew. Chem. Int. Ed.* **49:** 4170-4198.

Sparrer, H., Lilie, H., and Buchner, J. (1996) Dynamics of the GroEL protein complex: Effects of nucleotides and folding mutants. *J Mol Biol* **258:** 74-87.

Spector, J., Zakharov, S., Lill, Y., Sharma, O., Cramer, W.A., and Ritchie, K. (2010) Mobility of BtuB and OmpF in the Escherichia coli Outer Membrane: Implications for Dynamic Formation of a Translocon Complex. *Biophys J* **99:** 3880-3886.

Stokke, C., Flatten, I., and Skarstad, K. (2012) An easy-to-use simulation program demonstrates variations in bacterial cell cycle parameters depending on medium and temperature. *PLoS ONE* **7:** e30981.

Stone, S., Dennis, S., and Hofsteenge, J. (1989) Quantitative-Evaluation of the Contribution of Ionic Interactions to the Formation of the Thrombin-Hirudin Complex. *Biochemistry* **28:** 6857-6863.

Swaminathan, R., Hoang, C., and Verkman, A. (1997) Photobleaching recovery and anisotropy decay of green fluorescent protein GFP-S65T in solution and cells: Cytoplasmic viscosity probed by green fluorescent protein translational and rotational diffusion. *Biophys J* **72:** 1900-1907.

Taheri-Araghi, S., Bradde, S., Sauls, J.T., Hill, N.S., Levin, P.A., Paulsson, J., et al. (2015) Cell-size control and homeostasis in bacteria. *Current Biology* **25:** 385-391.

Taylorrobinson, D. (1995) The History and Role of Mycoplasma-Genitalium in Sexually-Transmitted Diseases. *Genitourin Med* **71:** 1-8.

Tyn, M., and Gusek, T. (1990) Prediction of Diffusion-Coefficients of Proteins. *Biotechnol Bioeng* **35:** 327-338.

van den Bogaart, G., Hermans, N., Krasnikov, V., and Poolman, B. (2007) Protein mobility and diffusive barriers in Escherichia coli: consequences of osmotic stress. *Mol Microbiol* **64:** 858-871.

Wallis, R., Moore, G.R., James, R., and Kleanthous, C. (1995) Protein-protein interactions in colicin E9 DNase-immunity protein complexes. 1. Diffusion-controlled association and femtomolar binding for the cognate complex. *Biochemistry* **34:** 13743-13750.

Wallis, R., Moore, G., James, R., and Kleanthous, C. (1995) Protein-Protein Interactions in Colicin E9 Dnase-Immunity Protein Complexes .1. Diffusion-Controlled Association and Femtomolar Binding for the Cognate Complex. *Biochemistry* **34:** 13743-13750.

Wang, Y., Benton, L.A., Singh, V., and Pielak, G.J. (2012) Disordered Protein Diffusion under Crowded Conditions. *J. Phys. Chem. Lett.* **3:** 2703-2706.

Young, K.D. (2006) The selective value of bacterial shape. *Microbiol. and Mol. Biol. Rev.* **70:** 660-703.

Zhang, G., Fedyunin, I., Miekley, O., Valleriani, A., Moura, A., and Ignatova, Z. (2010) Global and local depletion of ternary complex limits translational elongation. *Nucleic Acids Res* **38:** 4778-4787.

Chapter 2: Protein diffusion rates are highly responsive to abrupt cytoplasmic volume changes in *Lactococcus lactis*

We measured translational diffusion of proteins in the cytoplasm and plasma membrane of the Gram-positive bacterium *Lactococcus lactis* and probed the effect of osmotic upshift. For cells in standard growth medium the diffusion coefficients for cytosolic proteins (27 and 582 kDa) and 12-transmembrane helix membrane proteins are similar to those in *Escherichia coli*. The translational diffusion of GFP in *L. lactis* dropped by two orders of magnitude when the medium osmolality is increased by ~1.9 Osm, and the decrease in mobility is partly reversed in the presence of osmoprotectants. We find a large spread in diffusion coefficients over the full population of cells but a smaller spread if only sister cells are compared. While in general the diffusion coefficients we measure under normal osmotic conditions in *L. lactis* are quite similar to those reported in *E. coli*, the decrease in translational diffusion upon osmotic challenge in *L. lactis* is smaller than in *E. coli*. An even more striking difference is that in *L. lactis* the GFP diffusion coefficient drops much more rapidly with volume than in *E. coli*. We discuss these findings in the light of differences in turgor, cell volume, crowding and cytoplasmic structure of Gram-positive and Gram-negative bacteria.

Introduction

The cellular milieu is far different from idealized test tube conditions (Ellis, 2001; Gierasch *et al.*, 2009), with much higher macromolecule concentrations, more interaction partners, a spatially heterogenous and often compartmentalized nature. One of the major differences between *in vivo* and *in vitro* conditions is the crowdedness (and associated molecular complexity) of the cytoplasm and biological membranes (Gershenson *et al.*, 2011). Protein diffusion in this environment is significantly slower than in dilute solutions or idealized *in vitro* systems. For example the diffusion of GFP in water, with a diffusion coefficient (D) of ~90 μm^2/s, is faster than in the cytoplasm of eukaryotic cells (D = 24-27 μm^2/s) (Swaminathan *et al.*, 1997; Potma *et al.*, 2001) or prokaryotic cells (D = 3-14 μm^2/s) (Konopka *et al.*, 2009; Mika *et al.*, 2011b). This difference is often rationalized by the elevated macromolecule crowding of living cells with values reaching on average 200-300 g/l of macromolecules in the cytoplasm of *Escherichia coli* (Konopka *et al.*, 2007). However, by allowing cells to adapt to hyperosmotic conditions, the group of Weisshaar (Konopka *et al.*, 2009) observed that protein mobility is not solely determined by the biopolymer fraction (crowding) of the cell. Membrane protein diffusion in artificial membranes (Ramadurai *et al.*, 2009) is also at least an order of magnitude faster than in the plasma membrane of living cells (Kumar *et al.*, 2010). These examples demonstrate the importance of studying the behavior of proteins in their natural environment.

Most of the data on diffusion in bacterial cells come from the studies carried out in *Escherichia coli* (for reviews see (Konopka *et al.*, 2007; Mika *et al.*, 2011b)). In the highly crowded environment of *E. coli* diffusion is rather slow even when compared with eukaryotic cells. With the exception of the studies carried out on *Bacilli* (Cowan *et al.*, 2003; Cowan *et al.*, 2004), little is known about the diffusion in Gram-positive bacteria. The Setlow group (Cowan *et al.*, 2003; Cowan *et al.*, 2004) has reported on the mobility of proteins and lipid probes in *B. subtilis* spores germinating to vegetative cells but no accurate number for the translational diffusion of GFP in vegetative cells is given.

Although the amount of quantitative data is limited, it is generally thought that the turgor of *E. coli* (~3 atm) (Cayley *et al.*, 2000) is at least an order of magnitude lower than that of Gram-positive bacteria such as *B. subtilis* and *Staphylococcus aureus* (~20 atm) (Whatmore *et al.*, 1990). More recent measurements indicate a turgor pressure for *E. coli* as low as ~0.3 atm (Deng *et al.*, 2011). The higher turgor of Gram-positive bacteria must reflect a higher osmolyte concentration (e.g. K^+ and counter ions) and consequently it will take a larger osmotic upshift to plasmolyse Gram-positive bacteria than Gram-negative cells. Depending on the elasticity of the cell wall, the impact of osmotic stress on crowding and protein diffusion could also be different in these organisms.

By probing diffusion in *E. coli*, we and others have observed that the spread in diffusion coefficients in a population of isogenic cells is large, and larger than predicted from errors in the measurements (Konopka *et al.*, 2006; Mika *et al.*, 2011a). *L. lactis* cells generally grow in pairs. This offers the opportunity to address an intriguing problem of heterogeneity within a culture.

The cell membrane comprises an environment for macromolecules that is very different from the cytoplasm. The high viscosity of the lipid bilayer is a major cause of the at least an order of magnitude slower diffusion in the membrane as compared to the cytoplasm (Saffman *et al.*, 1975; Ramadurai *et al.*, 2009; Mika *et al.*, 2011b). Diffusion coefficients in artificial membranes (low protein to lipid ratio) are up to two orders of magnitude faster than those in the membrane of *E. coli* (Mullineaux *et al.*, 2006; Ramadurai *et al.*, 2009; Kumar *et al.*, 2010), which in part reflects the difference in crowding (excluded volume) but might also be a consequence of differences in the lipid composition and hydrophobic thickness of the membrane (Ramadurai *et al.*, 2010; Ramadurai *et al.*, 2010). The extent of the crowding in the cytoplasmic membrane of *L. lactis* is not known. However, in various biological contexts high numbers of proteins have been found in the membrane, for an overview see (Linden *et al.*, 2012).

26

We determined the diffusion coefficients of proteins in the cytoplasm and plasma membrane of the Gram-positive bacterium *Lactococcus lactis.* To probe the protein diffusion in the cytoplasm, we have used soluble GFP (27 kDa) and β-galactosidase-GFP (582 kDa); to probe the diffusion in the plasma membrane, we used the native transport protein BcaP and the heterologously expressed transport protein LacSΔIIA, each C-terminally tagged with GFP or YPet as fluorescent reporter. Diffusion coefficients are determined with fluorescent recovery after photobleaching (FRAP) and single molecule tracking (SMT). FRAP reports long-range ensemble diffusion, whereas SPT allows detection of confined spaces by tracing the paths of individual molecules. We also studied the similarity in diffusion properties between sister cells. Finally, we determined the turgor and cell volume, and the impact of osmotic stress on the translational diffusion of proteins in the cytoplasm and plasma membrane.

Results

Diffusion of GFP and β–galactosidase-GFP in L. lactis
By using FRAP we determined the diffusion coefficient of GFP and β-galactosidase-GFP in the cytoplasm of *L. lactis* (Figure 1A). The measurements were performed on cells in a chemically defined growth medium (CDM[RP], see experimental procedures) or in phosphate-based media. In Figure 2A, we observe that for any given condition the diffusion of GFP is characterized by a broad spread in *D* values (from about 7-fold to more than 2 orders of magnitude). In non-stressed cells, the median of the diffusion coefficient of GFP is 7 μm^2/s for cells in CDM[RP] (Fig. 2A, black bars, top panel, Table S1) and 4.3 μm^2/s for cells resuspended in phosphate-based medium with the same osmolality as CDM[RP] (Fig. 2A, gray bars, top panel, Table S1). These values are in the same range as those reported for *E. coli:* 3-14 μm^2/s (Konopka *et al.*, 2007; Konopka *et al.*, 2009; Mika *et al.*, 2011b). The expression levels of β-galactosidase-GFP were lower than for GFP and some cells contained aggregates; cells with visible aggregates were not included in the analysis. The overall D_{median} of two independent datasets was 0.77 μm^2/s (number of cells, N_{cells} = 53, see Table S2), which is similar to the value measured previously in *E. coli: D_{median}* of 0.68 μm^2/s (Mika *et al.*, 2010).
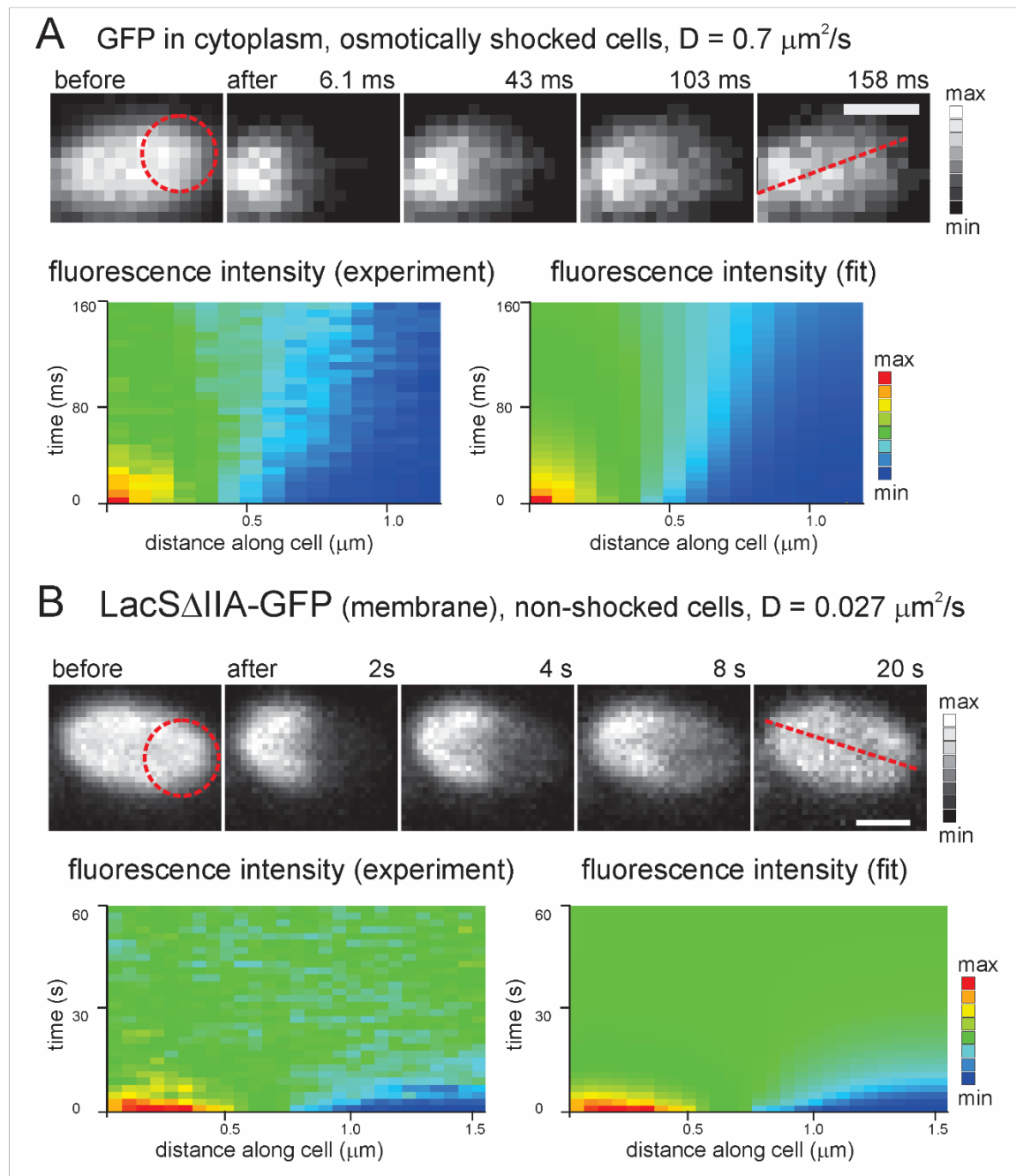
Fig. 1. Fluorescence Recovery After Photobleaching measurements of protein diffusion in the cytoplasm and the membrane of *Lactococcus lactis*. Examples of FRAP measurements of GFP in the cytoplasm (A. an individual cell, osmotically-shocked, D = 0.7 μm²/s) and LacSΔIIA-GFP in the membrane (B. an individual cell under normal osmotic conditions, D = 0.027 μm²/s) of *L. lactis*. Upper panels: normalized fluorescence microscopy images of an individual cell throughout the measurement, with timestamp above the images. Red dotted circle shows the bleaching region at the cell pole. Red dotted line shows the longer axis of the cell along which fluorescence intensities were extracted and used to calculate the diffusion coefficient. The scale bar is 0.5 μm. Lower panels show the normalized fluorescence intensity along the red dotted line (pseudocolored and shown in a rainbow setting) plotted as a function of time, with left showing the actual data and right the fit generated by the analysis software used to calculate the diffusion coefficient.
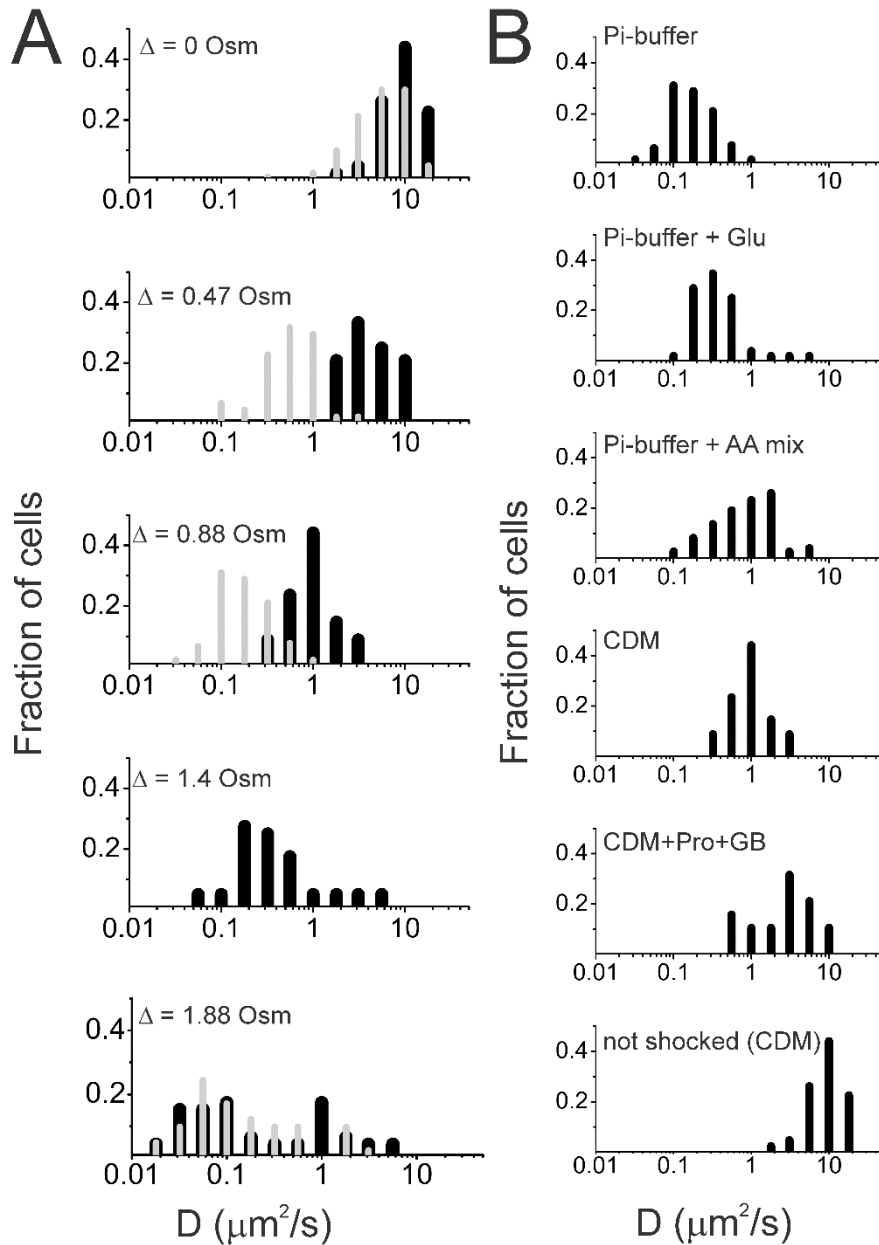
Fig. 2. Diffusion coefficients of GFP in the cytoplasm of _Lactococcus lactis_ cells under different osmotic conditions. (A) Histograms with the distributions of D_{GFP} in _L. lactis_ cells under normal osmotic conditions (Δ = 0 Osm) and upon osmotic upshock (the extent of shock indicated as Δ and expressed in Osm). Black bars: cells grown and measured in CDM^{RP} with varying concentrations of NaCl. Gray bars: cells grown in CDM but measured in phosphate buffers with varying concentrations of NaCl (see experimental procedures for medium composition). The number of cells varied from 24 to 84 individual cells, see Table S1 for details. (B) Histograms of the distribution of D_{GFP} in _L. lactis_ grown in CDM^{RP} and upshifted with NaCl or KCl (Δ osmolality = ~0.88 Osm) in the presence or absence of osmoprotectants. From top to bottom: cells in phosphate buffer with NaCl or KCl; same as above but with 2.7 mM glutamate added; same as above but with the full complement of amino acids present in CDM added (see Experimental Procedures section for details); cells in CDM^{RP} with NaCl; same as above but with 6 mM proline plus 1 mM glycine betaine. Lowest panel: for comparison cells under normal osmotic condition in CDM^{RP} (Δ osmolality = 0 Osm). The number of cells varied from 34 to 84, see Table S6 for details.

Diffusion of proteins in L. lactis membranes

For membrane proteins FRAP was performed in the same way as for the cytosolic proteins (Figure 1B). For the GFP-tagged membrane proteins LacSΔIIA and BcaP, we determined the following diffusion coefficients: for LacSΔIIA-GFP D_{mean} = 0.020 μm^2/s and for BcaP-GFP D_{mean} = 0.019 μm^2/s (Fig. 3A, Table S3, n ≥ 22 cells), which is much slower than the D_{median} of GFP or β-galactosidase-GFP in the cytoplasm of *L. lactis.* These values are similar to those obtained for membrane proteins of similar size expressed in the cytoplasmic membrane of *E. coli*, 0.017 - 0.028 μm^2/s(Kumar *et al.*, 2010). Since FRAP reports long-range lateral mobility, diffusion coefficients will be underestimated if the proteins are confined to specific areas of the membrane. To test this possibility, we performed single molecule tracking (SMT) experiments on LacSΔIIA-YPet. The SMT experiments were done with frame times of 34, 78 and 150 ms, all of which yielded diffusion coefficients from 0.02 to 0.03 μm^2/s with an average of 0.026 μm^2/s (Fig. S1), which is in good agreement with the FRAP data.

In the analysis of the FRAP and SMT data, the contribution of the geometry of the membrane is ignored. To determine whether this assumption is justified, we performed simulations of FRAP and SMT measurements of the diffusion of particles in membranes. For the SMT a set of simulations was done on different cell geometries and with different diffusion coefficients and time steps. The effect of localization error was also analysed. The details are given in the Experimental procedures and Table S4. In the simulations without localization error the step size distribution is not Gaussian, the expected shape for diffusion on a plane, but has a high fraction of short steps. When localization error is introduced the distribution becomes more Gaussian and resembles the data more closely. A difference of more than 2-fold between input and output D was seen only for big a localization error (0.078 μm) and/or very slow diffusion (0.003 μm^2/s). The localization error in our data is ~30 nm. We estimated our localization error by determining the positions of fluorescent beads that were fixed in place and imaged under such conditions that the signal to noise ratio is similar to that of our SMT data. For the FRAP simulations 4000 particles were placed on the membrane of one half of the cell and allowed to diffuse with D values between 0.003 and 0.1 μm^2/s. Analysis was the same as with real data. The D_{out} was in all cases lower than the D_{in} but never by more than 25 % (Table S5). For both FRAP and SMT the effect of the cell geometry on D is smaller than the spread in the data and does not significantly alter our results.
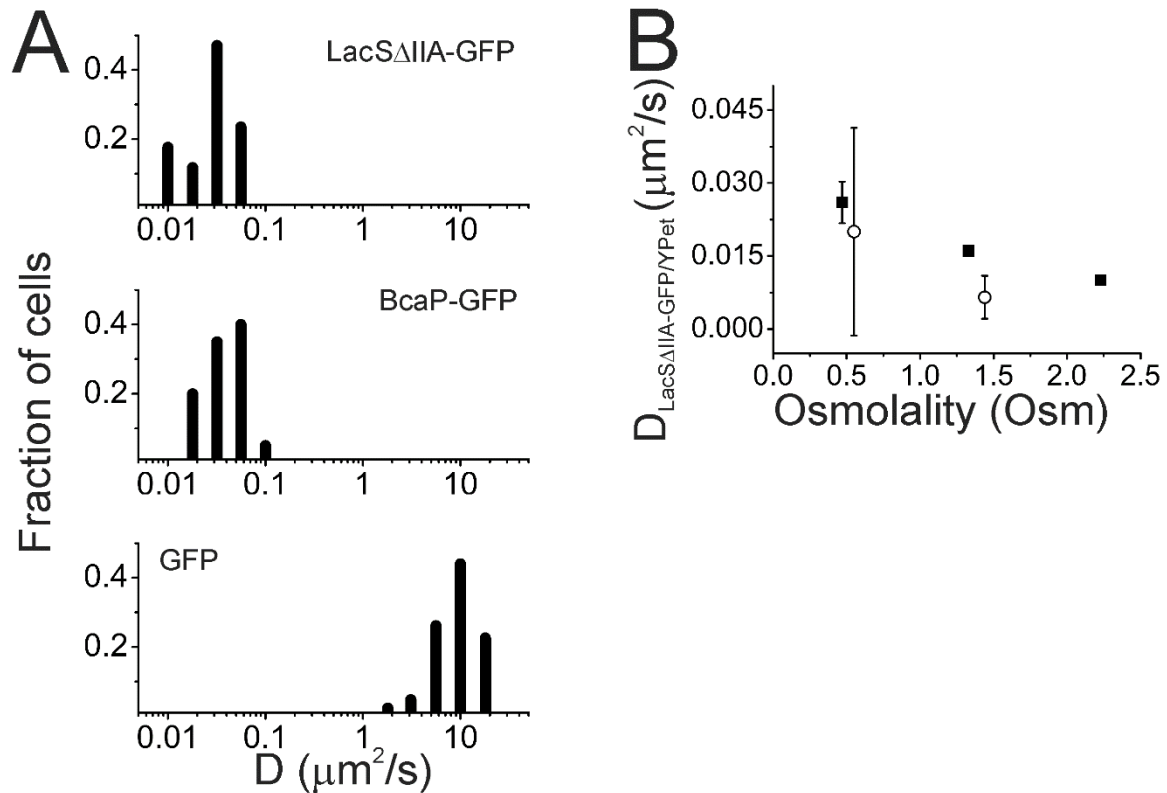
Fig. 3. Protein diffusion in the membrane of *Lactococcus lactis*. A. Histograms of diffusion coefficients of two membrane proteins LacSΔIIA-GFP (top panel) and BcaP-GFP (middle panel), expressed in *L. lactis* cells and grown under normal osmotic conditions in CDM[RP]. The data is compared with the cytosolic GFP (bottom panel). B. The apparent diffusion coefficient of LacSΔIIA-GFP/YPet drops with increased osmotic challenge. Open circles: FRAP data, Black squares: SMT data. The points represent means. The error bars for the FRAP data indicate the standard deviation over about 20 cells. The error bar for the SPT data is the standard deviation over values determined in three independent experiments (with 34, 78 and 150 ms time step in the tracking (see Fig. S1)).

Diffusion rates drop as a consequence of osmotic upshift

Next, we determined the impact of osmotic stress on the mobility of cytoplasmic and membrane proteins. Cells were osmotically-upshifted by resuspending them in CDM[RP] medium or phosphate-based media the osmolality of which was increased by adding NaCl or KCl. In general a drop of diffusion coefficient of GFP is observed with the extent of osmotic upshift (Fig. 2). For cells in CDM[RP], the $\log_{10}$ of the diffusion coefficient drops linearly with the increase in external osmolality (Fig. 2, Table S1, see also Fig. 6 black squares). The D_{median} of GFP dropped from 7 µm^2/s under normal osmotic conditions to 0.087 µm^2/s after osmotic upshift in a medium of about 2.3 Osm. The spread in the diffusion coefficients increases with the strength of the osmotic stress. The drop of GFP diffusion is more pronounced in phosphate-based media than in CDM[RP] (Fig. 2A compare gray bars with black ones, see also Table S1). The rich growth medium (CDM[RP]), unlike the phosphate-based medium, contains osmoprotectants like, glutamate and other amino acids, and the accumulation of these molecules, that help the cell combat effects of osmotic stress, may explain the higher diffusion coefficients (see section on effect of osmoprotectants). The cell volume in CDM[RP] changes by about 15 % when shocked with a 2.3 Osm medium (Fig. S2). The drop in cell volume with increasing osmolality of the medium is linear. With CDM[RP] as the shock medium the volume continues to decrease at higher osmolality down to about 50 % of the initial volume at an osmolality of about 7.5 Osm (Fig. S2). Surprisingly, in phosphate buffer, the diffusion coefficient doesn't drop much beyond 1.4 Osm.

When determining the effect of osmotic upshift on membrane protein diffusion we encountered some technical difficulties that prevent us from drawing firm conclusions. By increasing the medium osmolality from ~0.5 Osm to ~1.4 Osm, the mean diffusion coefficient of LacSΔIIA-GFP/YPet decreased from 0.020 to 0.0065 μm^2/s when probed by FRAP and from 0.026 to 0.016 μm^2/s when probed by SMT. Using SMT a further drop, to 0.010 μm^2/s, is observed in medium with an osmolality of 2.3 Osm (see Fig. 3B). For the FRAP measurements we could not go to higher osmolality because fluorescent foci appeared, which possibly indicates protein aggregation. SMT simulations indicate (Table S4) that when the diffusion coefficient decreases to about 0.003 μm^2/s, a localization error of 36 nm makes it appear as 0.015 μm^2/s (for a frame time of 50 ms). Another problem is that when cells are shrinking due to the osmotic upshift membranes need to undulate, which makes the diffusion path appear shorter than it actually is and leads to an underestimation of the diffusion coefficient. Taking these arguments together we cannot quantify the changes in lateral diffusion, but the impact of osmotic stress on membrane protein diffusion is much lower (if occurring at all) than on the mobility of cytoplasmic constituents.

Osmoprotectants restore GFP diffusion
To quantify the effect of osmoprotectants on the diffusion of GFP in osmotically-stressed cells, we performed FRAP measurements on cells resuspended in media with different (combinations of) osmoprotectants (Fig. 2B, Table S6). Under conditions of moderate osmotic stress (Δosmolality ~0.8 Osm), D_{median} equals 0.11 μm^2/s (Fig. 2B top panel, Table S6) and this value increases to 0.29 μm^2/s upon addition of glutamine (2.7 mM) and to 0.63 μm^2/s when all amino acids present in CDM^{RP} are added. CDM^{RP} gave a D_{median} of 0.84 μm^2/s and CDM^{RP} with omosprotectants proline (6 mM) and glycine betaine (1 mM) gave a D_{median} of 2.4 μm^2/s. Importantly, the substitution of Na^+-ions with K^+-ions in the phosphate-based media did not offer any protection of *L. lactis* against osmotic stress, which is in marked contrast to observations made for *E. coli* (van den Bogaart *et al.*, 2007)*.

Diffusion in Lactococcus lactis sister cells
An intriguing feature of bacteria is the significant spread in observed diffusion coefficients within a population of (isogenic) cells (Elowitz *et al.*, 1999; Konopka *et al.*, 2006; van den Bogaart *et al.*, 2007; Konopka *et al.*, 2009; Mika *et al.*, 2010). The molecular basis for the apparent heterogeneity of the cytoplasm has not yet been established. The spread in diffusion coefficients is higher for osmotically-stressed cells (Konopka *et al.*, 2009). Suggesting that at least part of the variation between cells is real rather than a measurement error.

After cell division *L. lactis* cells stay attached for some time and are likely to have a more similar molecular composition than randomly chosen cells in a population. We thus compared the spread in diffusion coefficients in sister cells to that of the population as a whole. Should the large spread of diffusion coefficients between individual cells in a population be caused by the differences in macromolecular composition of those cells (protein, RNAs and DNA composition and copy numbers), one would expect that two sister cells, created as a result of the division of one mother cell, have a more similar molecular composition and thus more similar diffusion coefficients. To test this hypothesis, we compared the differences in GFP diffusion between sister cells with randomly chosen cell pairs in a population and calculated the Q quotient (see equation 3 in the "statistical data analysis" section of the Experimental Procedures, Table S7 and Fig. 4). If Q<1 than the diffusion coefficients of sister cells are more similar than diffusion coefficients of randomly chosen pairs in the tested population of cells. Similarly, if Q>1 than the diffusion coefficients of randomly chosen cells in the population are more similar than the diffusion coefficients of sister cells. If Q=1 there is no difference between randomly chosen pairs and pairs of sister cells.

As shown in Figure 4A and Table S7, the majority of the data (~75%, n = 494 cells with 217 pairs) has a Q quotient lower than one. This means that overall sister cells are more similar to each other than

two randomly chosen cells. For individual pairs the difference in diffusion coefficient between sister cells can be quite large. This might explain why for some datasets Q > 1. Due to the low number of cells no statistically significant statements can be made on single datasets. To check whether the overall sister cell similarity is significant the median of the Q's was calculated and compared to a distribution of median Q's made by using pairs of cells that were randomly picked from the datasets instead of sister cells (as explained in the Experimental procedures section). This comparison shows that there is only a very small chance that the same median Q value is obtained from the real pairs and randomly picked pairs of cells (see Figure 4B). Less than 0.2 % of the random datasets have a median Q the same or lower than the median Q of the real data; performing the analysis with means instead of medians gives a similar result. Next, we estimated the magnitude of the difference between sister cells by simulating the entire pairs experiment. We used means and standard deviations from the real diffusion data (as determined by FRAP) and calculated median Q values for a range of similarities between sister cells. This gives us the rough estimate that the variation between sister cells is slightly more than half of the total variation in diffusion coefficients (Fig. S3). Based on this we conclude that the sister cells have more similar diffusion coefficients and are thus more homogeneous in cytoplasmic composition than randomly chosen cells in a population.
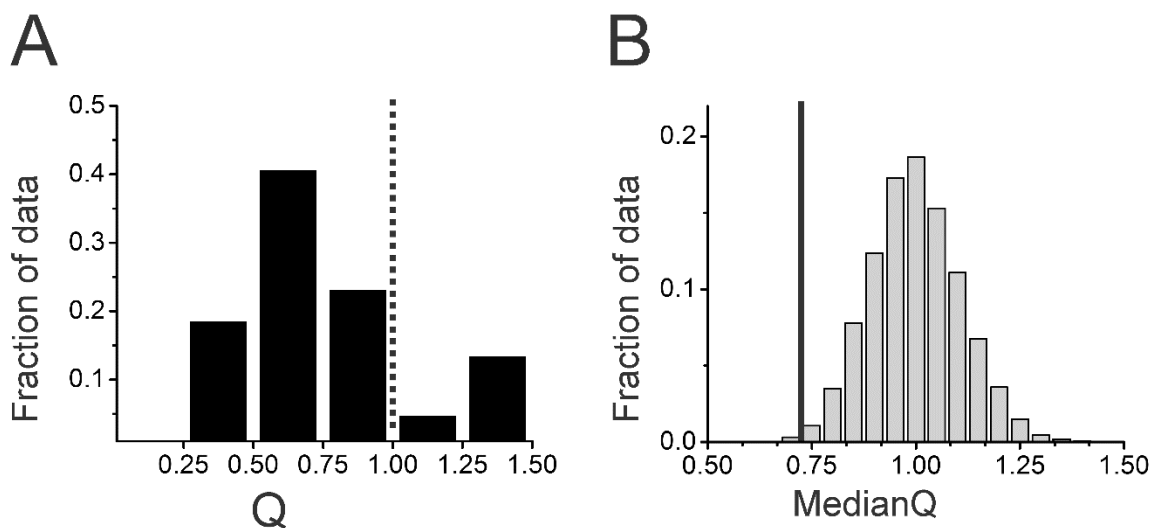
Fig. 4. Similarity of diffusion coefficients of GFP between sister cells of *Lactococcus lactis.* A. Histogram showing the distribution of the Q values over 22 datasets. The Q value indicates the ratio of the median of the differences in the *D* of GFP between cells in a pair of *L. lactis* sister cells (real pairs) and the difference between cells in randomly chosen pairs in the population (artificial pairs), see experimental procedures. The histogram is corrected for the number of cells present in each dataset, $n_{REAL PAIRS}$ = 217; n = 494 cells. See Table S7 for the used datasets and Q values. The dotted line shows Q = 1. If Q < 1, the diffusion coefficient of GFP between sister cells is more similar than throughout randomly chosen cells in the population. B. Histogram showing the distribution of median Q's for randomly picked pairs, details are given in the experimental procedures. The Q value determined from the real pairs, 0.7, is outside the distribution and is indicated by the black line. This means that the similarity between sister cells observed in A is significant.

The relation between cell volume, turgor and the diffusion coefficient

We used fluorescence microscopy to determine the volume of *L. lactis* cells as a function of the osmolality of the medium (Fig. S2) and determined the relation between cell volume and *D*. We observe a drop in *D* of almost a 100-fold with a drop in volume of only about 15 % (Fig. 5). This can be compared to the relation between volume and *D* in *E. coli*. Here the diffusion coefficient drops 3-fold with a volume drop of 40 %. In the next 10 % drop in volume for *E. coli* the D drops in a similar fashion as in *L. lactis* (Fig. 5). The small drop in volume of *L. lactis* after osmotic upshift, as compared to *E. coli*, is probably the result of its higher internal potassium concentration (and counter ions) (Poolman et al., 1987; Konopka *et al.*, 2009). The *E. coli* cell volumes were estimated from the difference in cytoplasm inaccessible and accessible radioactivity reporters (Konopka *et al.*, 2009). We determined the *E. coli* cell volumes by fluorescence microscopy (similar to how the *L. lactis* volumes were determined). Because *E.coli* cells start to plasmolyse at external osmolalities of about 0.6 Osm, only the low osmotic stress range could be probed. We used the osmolality vs *D* data from (Konopka *et al.*, 2009) and our own fluorescence-based measurements to relate volume with *D*; the two datasets show a similar trend.

It is easy to make large errors in the determination of volumes of cells as small as *L. lactis* and *E. coli* because of the optical spread in microscopy. Small differences in fitting the cell outlines can lead to rather big differences in cell volume. Therefore, we determined relative volumes (changes in volumes) rather than absolute ones, which cancels most of the error. We performed calculations to estimate the error when relative volumes are considered (presented in Fig. 5 and Fig. S2). The results are shown in figure S4 and indicate that the difference in cell volume decrease for *E. coli* and *L. lactis* still holds even when large errors (0.4 μm in length and width) would be made in fitting the cell outlines.

The cell volume data can be used to get an estimate of the turgor pressure in *L. lactis* (Whatmore *et al.*, 1990). At high osmotic conditions, the osmolality inside the cell is the same as outside. Combining this information with cell volume data gives the osmolality of the cytoplasm under normal growth conditions. The combination of this and the medium osmolality under normal growth conditions allows calculation of the turgor pressure (see caption of Fig. S3). From figure S3 it is clear that the turgor in *L. lactis* is similar to that of *B. subtilis* (19 atm, which is equivalent to 0.75 Osm) (Whatmore *et al.*, 1990). The turgor in *E. coli* is an order of magnitude lower (Cayley *et al.*, 2000; Deng *et al.*, 2011).
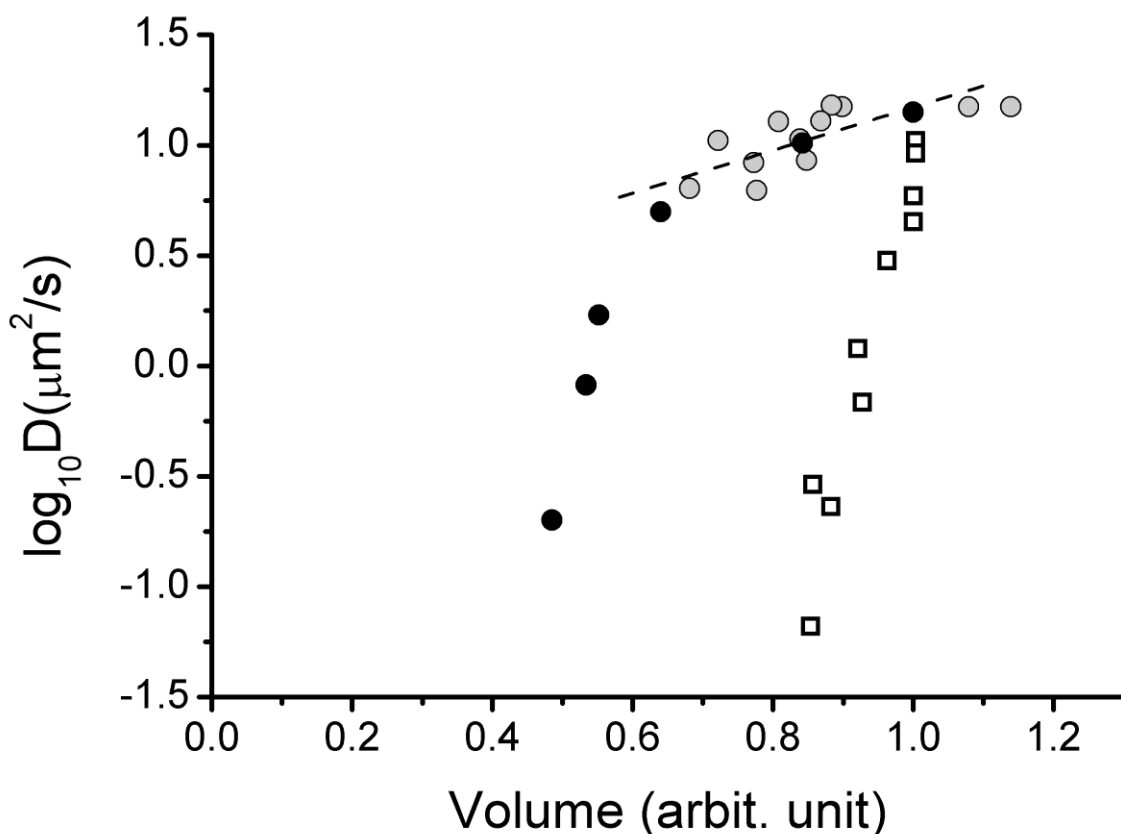
Fig. 5. The relation between cytoplasm volume and diffusion coefficient after osmotic upshift in *Lactococcus lactis* and *Escherichia coli*. Open squares: *L. lactis* (microscopy), black circles: *E. coli* (radioactivity assay), grey circles: *E. coli* (microscopy), Black dashed line: linear fit to *E. coli* (microscopy) volume vs log10(D) data. To plot the *L. lactis* data we combined osmolality vs D with osmolality vs volume measurements. We used a fit of the osmolality vs volume data instead of the data points directly (see fitted line in Fig. S2). The *E. coli* (radioactivity assay) data was taken from (Konopka *et al.*, 2009). To plot the *E. coli* (microscopy) data we combined osmolality vs volume data, which we recorded on the microscope, with osmolality vs *D* data from (Konopka *et al.*, 2009).

Discussion

Significance of measuring diffusion coefficients in living cells

To understand cellular processes knowledge of diffusion coefficients is necessary. However, if a process is reaction-limited rather than diffusion-limited, small changes in the diffusion coefficient will not have much impact. It is not known how many processes in the cell are diffusion limited. However, when cells are osmotically-challenged and the diffusion coefficients drop by more than an order of magnitude a significant amount of processes may become diffusion-limited. Knowledge of diffusion coefficients not only enables back-of-the-envelope calculations that help the understanding of cellular processes (Moran *et al.*, 2010), but also provides actual numbers for computational studies (employed by system biologists). Diffusion coefficients and their effects have been used and studied by various groups (Anderson *et al.*, 2010; Zhang *et al.*, 2010; Halatek *et al.*, 2012; Klumpp *et al.*, 2013; Soh *et al.*, 2013; Balcells *et al.*, 2014).

Comparison of protein diffusion in L. lactis, E. coli and other bacteria

In Figure 6 we compare the diffusion coefficients obtained for *L. lactis* with those measured in *E. coli* cells (Konopka *et al.*, 2006; Konopka *et al.*, 2009; Mika *et al.*, 2011a). It is apparent that under normal osmotic conditions (0.28 Osm for *E. coli* and 0.53 Osm for *L. lactis*) the diffusion coefficients of soluble

and membrane proteins are similar for both organisms. The slow diffusion of proteins reported for *E. coli*, *Pseudomonas aeruginosa* (Guillon *et al.*, 2013) and *Caulobacter crescentus* (Llopis *et al.*, 2012), as compared to eukaryotic cells, is thus also observed for *L. lactis* and may be a general property of Gram-negative and Gram-positive bacteria.

Both *L. lactis* and *E. coli* respond to osmotic stress by a drop in protein diffusion, which is mitigated when the medium contains osmoprotectants. For both organisms a drop in cell size and diffusion coefficient happens even after small osmotic upshift. This suggests that the cell wall, that is initially stretched, causes the cytoplasm to shrink when the turgor pressure is decreased (see also (Koch, 1984)). There are also important differences between the two organisms. We observe that *L. lactis* is less susceptible to osmotic challenge than *E. coli* as it requires higher medium osmolalities to decrease the diffusion of GFP (Fig. 6). This most likely relates to the much higher turgor of *L. lactis*. The diffusion coefficients for GFP in the cytoplasm of osmotically-challenged *L. lactis* cells do not reach values as low as in *E. coli.* Under the highest osmotic shock applied the D_{median} of GFP in *L. lactis* is an order of magnitude higher than in *E. coli* (Fig. 6)*.*

Remarkably, the mobility of GFP in *L. lactis* appears much more sensitive to changes in cell volume than in *E. coli* (Fig. 5)*.* The *L. lactis* cell volume has dropped by only 15 % after a shift in the outside osmolality of about 1.8 Osm, while for *E. coli* the volume drops by 50 % after a shift of only 1.2 Osm. After this 15 % drop of volume in *L. lactis* the GFP diffusion coefficient is lower than after a 50 % drop in volume in *E. coli* (Fig. 5). Because the amount of macromolecules is the same at every volume, our findings imply that in *L. lactis* the mobility of proteins drops faster with relative macromolecule volume fraction than in *E. coli*. The reason for this difference between *L. lactis* and *E. coli* is not clear. It might be that the basal crowding is higher in *L. lactis* or that co-solvents and/or macromolecular interactions differ in the two organisms. Note that the diffusion coefficient at similar biopolymer fractions in *E. coli* can differ more than an order of magnitude (Konopka *et al.*, 2009). When cells are allowed to adapt to hyperosmotic stress conditions the diffusion rate increases again, suggesting that changes in chemical composition allow proteins to move again. The nature of the change in chemical composition is not known. In the future measurements on the composition and concentration of individual macromolecules in *L. lactis* cells, combined with a scaled particle theory like analysis, could lead to an explanation of this striking difference between *L. lactis* and *E. coli*. Depending on the outcome this might have consequences for how one needs to look at processes in the cell. For example, if the basal crowding in *L. lactis* is different from that in *E. coli,* rate and equilibrium constants might be different between the two even for the same processes (Zhou *et al.*, 2008).
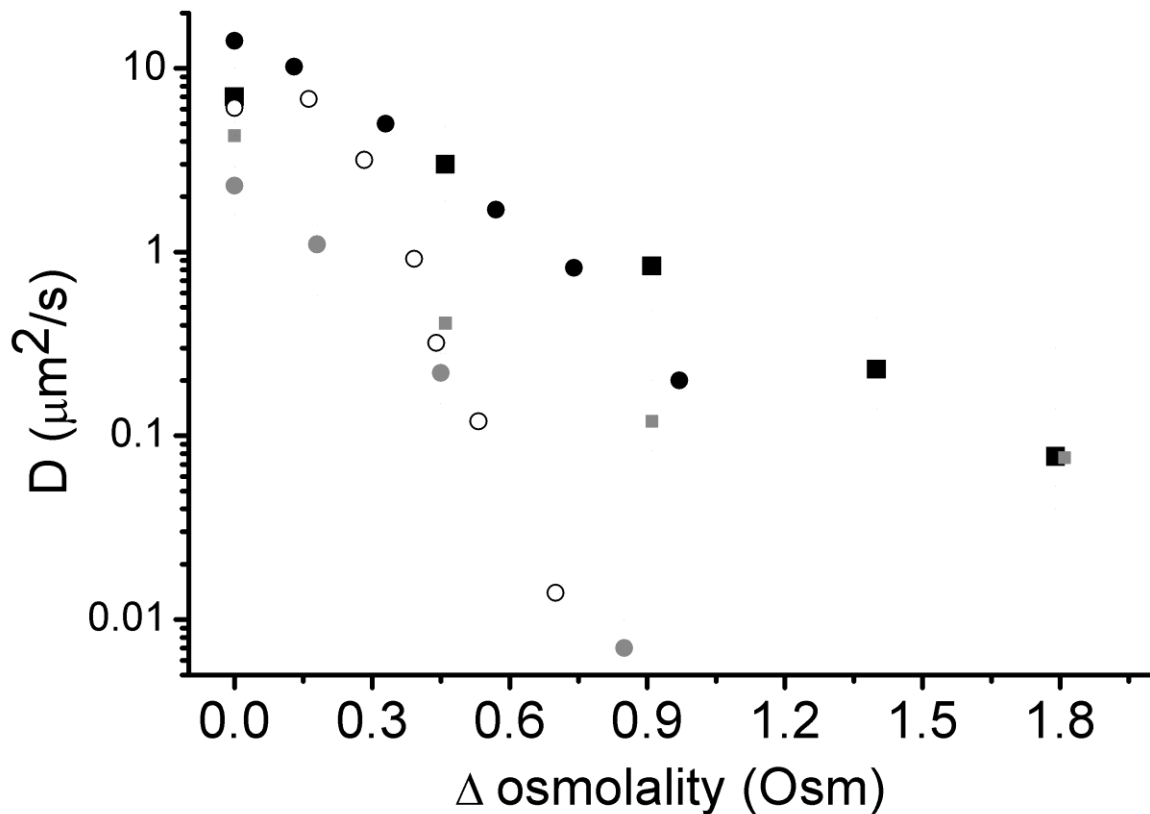
Figure 6. Comparison of GFP diffusion in *Lactococcus lactis* with *Escherichia coli* under different osmotic conditions. GFP diffusion coefficients plotted as a function of the osmolality of the upshock medium. Squares: *L. lactis* cells (data from this study, median values) grown and measured in CDM[RP] at the indicated osmolality (black squares) or grown in CDM[RP] and osmotically-shocked in phosphate buffer (gray squares). Circles: *E. coli* cells (data from literature) grown in a chemically defined, MOPS based, minimal medium (MBM) and shocked in MBM with NaCl but without K[+] (black circles, median values) (Konopka *et al.*, 2009); grown in LB and shocked in phosphate buffer of the indicated osmolality (white circles, mean values (Konopka *et al.*, 2006); gray circles, median values (Mika *et al.*, 2011a)).

Apparent diffusion coefficients at different length scales

Our work indicates that the diffusion coefficient of LacSΔIIA-GFP/YPet is ~0.03 μm^2/s, irrespective of whether the mobility is determined by FRAP or SMT. FRAP probes diffusion over a length scale of 1-1.5 μm, while in our analysis SMT reports on length scales of about 0.1-0.2 μm. Thus, we find no evidence for barriers of diffusion in the membrane of *L. lactis*, at least not in the range of 0.1 μm or beyond. In a recent study the diffusion coefficient of a membrane protein, TetA-YFP, was determined in *E. coli* using both FRAP and fluorescence correlation spectroscopy (FCS) (Chow *et al.*, 2012). The diffusion coefficient determined by FRAP was 0.086 ± 0.017 μm^2/s, which is faster than our estimates (this study) and others using similarly-sized membrane proteins in *E. coli* (ref. Kumar 2010), yet in the same order of magnitude. A value of *D* of 9.1 ± 3.4 μm^2/s was obtained when the mobility was determined by FCS. The authors attribute the discrepancy to the smaller length scale over which FCS reports as compared to FRAP. The combined data is taken as evidence for the presence of diffusion barriers in the membrane of *E. coli.* However, a value of ~9 μm^2/s is even faster than measured for membrane protein diffusion in giant-unilamellar vesicles at very low protein to lipid ratios (Ramadurai *et al.*, 2009). We consider it unlikely that 9 μm^2/s represents a true value of diffusion coefficient of a membrane protein in a living bacterium, as to the best of our knowledge no other measurements of

membrane protein diffusion report values higher than 0.5 µm^2/s (Deich *et al.*, 2004; Nenninger *et al.*, 2014), even when measured by SMT.

Comparison with Bacillus subtilis
The Setlow group investigated protein mobility in *Bacillus subtilis* spores during sporulation (Cowan *et al.*, 2003). When *Bacillus* spores are in the dormant state their water content is extremely low and it is expected that there is very little or no molecular mobility. Based on FRAP measurements on spore-targeted GFP, Cowan and co-workers (Cowan *et al.*, 2003) conclude that soluble proteins are virtually immobile in dormant spores with *D* ~0.0001 µm^2/s and 4 orders of magnitude slower than in germinated spores (D_{GFP} estimated to be > 1 µm^2/s). This is, to the best of our knowledge, the only other example of a macromolecule diffusion study in a Gram-positive bacterium. A direct comparison with our data is not possible because the values reported for germinated *B. subtilis* cells are only estimates. While the severest osmotic upshifts may cause plasmolysis and loss of most of the free water from the cytoplasm, the macromolecules of the cell will be more hydrated than in spores. In fact, the core of dormant spores are in a glass-like state nearly completely devoid of water, which is a likely rationale for the 4-orders of magnitude decrease of GFP mobility in dormant spores as opposed to the 2-orders of magnitude drop in our osmotic upshift studies.

Physiological heterogeneity as a means of managing fluctuating environments
Already in the first studies on protein diffusion in *E. coli* (Elowitz *et al.*, 1999), it was noted that the spread in diffusion coefficients obtained from a population of isogenic cells is fairly large. This phenomenon was subsequently observed for other (macro)molecules (NBD-glucose, (β-galactosidase-GFP)$_4$ (Mika *et al.*, 2010)). In osmotically-stressed *E. coli* cells the spread becomes even larger (Konopka *et al.*, 2006; Mika *et al.*, 2011a), and not all cells respond in the same way, *i.e.* some cells form visual plasmolysis spaces while others do not (Konopka *et al.*, 2006) and in a fraction of the *E. coli* cells the nucleoid seems to form a barrier for diffusion while in others it does not (Mika *et al.*, 2010).

The data reported here also reveal a wide spread in the translational diffusion coefficients of soluble and membrane proteins in *L. lactis*. Also in the case of *L. lactis*, the spread of GFP diffusion coefficients becomes larger when the cells are osmotically challenged. This suggests that at least under stress conditions some of the variation seen is due to real differences between cells. We have also performed double measurements on single cells (Fig. S4), which also suggest that part of the observed variation is real. When looking over most of the data recorded with the GFP expressing strains, we find that sister cells have more similar diffusion coefficients than pairs of randomly picked cells. This also suggests that real variation is seen in the data.

Isogenic bacterial cultures generally display heterogeneity at the individual cell level for multiple physiological parameters (Golding *et al.*, 2005; Lidstrom *et al.*, 2010). This heterogeneity can be beneficial as it affords subpopulations to survive the (stress) conditions that kill the majority of cells (Booth, 2002; Aertsen *et al.*, 2005). An example of such adaptive physiological heterogeneity of isogenic bacterial cultures are persistor cells (Balaban *et al.*, 2004; Lidstrom *et al.*, 2010), *i.e.,* a subpopulation of slowly growing cells that is more resistant to antibiotics and helps these cells to survive under stress conditions (Keren *et al.*, 2004). In lactic acid bacteria physiological heterogeneity was observed in the response of *Lactobacillus plantarum* to acid stress (Ingham *et al.*, 2008). Here, we speculate on the possibility that the heterogeneity in diffusion coefficients has a biological function (i.e. came about by evolutionary adaptation). Assuming that the broad spread of GFP diffusion coefficients reflects heterogeneity of the cytoplasmic structure within populations of isogenic cells, genetic elements that encode this heterogeneity might then be more likely to survive osmotic stress and in time take over the population. Consider the simple case where we have a genotype 1 that encodes A amount of heterogeneity (i.e. spread of diffusion coefficients) and genotype 2 that encodes B amount of heterogeneity. If A is bigger than B, and, due to environmental fluctuations in say osmotic

stress a bigger amount of heterogeneity is beneficial, genotype 1 will be selected for and take over the population. The large spread in diffusion coefficients we observe after cells are osmotically stressed might thus have come about through natural selection. However, the observed variation in diffusion coefficients could be a consequence of physical constraints that relate to the complexity of cells. It might even be that over time, with cells becoming better in regulating the cytoplasmic structure, the variability in diffusion coefficients decreases.

It would be interesting to know what causes the difference in diffusion coefficients between cells. It could be that there are differences in the concentrations of small molecules in cells (e.g. K^+) and/or differences in the degree of volume exclusion by macromolecules. The extent of the true variation in diffusion coefficient within the population under various conditions is still an open question. As are the consequences of this variation for various cell processes both under fast growth and under osmotic stress.

In conclusion: We show that the diffusion coefficients of proteins in the Gram-positive bacterium *L. lactis* and the Gram-negative *E. coli* are similar, *i.e.* when the cells are cultivated in standard media and under optimal conditions for fast growth. In both organisms there is also a big spread in the diffusion coefficients of isogenic cells measured under the same conditions. This spread seems, in part, real variation. We show that the turgor of *L. lactis* is an order of magnitude higher than that of *E. coli,* and, consistently, that it takes a higher medium osmolality to slow down the diffusion in *L. lactis* than in *E. coli.* The changes in cytoplasmic volume as a function of medium osmolality are much smaller in *L. lactis* than in *E. coli*, but, remarkably, protein diffusion is much more sensitive to changes in volume in *L. lactis* than in *E. coli*. We have not yet elucidated the cause of this difference. Potentially, it could have important consequences for our understanding of how processes occur in organisms as different as *L. lactis* and *E. coli.* Differences in crowding or (non-specific) interactions between macromolecules would imply differences in parameter values for folding rates, fold stabilities, interaction rates and interaction affinities of proteins.

Experimental procedures

Strains and plasmids
All experiments on *Lactococcus lactis* were performed on strain NZ9000, carrying pNZ8048 and bearing the genes for GFP, β galactosidase-GFP, BcaP-GFP, LacSΔIIA-GFP or LacSΔIIA-YPet. All genes are under the control of the nisin A promoter (deRuyter *et al.*, 1996). GFP and β-galactosidase-GFP (tetramer) are cytosolic proteins of 30 and 582 kDa, respectively. LacSΔIIA and BcaP are integral membrane proteins predicted to have 12 transmembrane helices (prediction by TMHMM (Krogh *et al.*, 2001)). Sequences of the proteins are given in the supplementary materials. The reference for YPet is the work of (Nguyen *et al.*, 2005). The β-galactosidase sequence used here is from *E. coli* strain MG1655 except that the first eight amino acids (MTMITDSL) were substituted with MGGGFAKG. LacSΔIIA is the integral membrane domain (residues 1-474) of the lactose transporter LacS from *Streptococcus thermophilus* and BcaP is an amino acid transporter from *L. lactis*. For the work on *E. coli* we used an MG1655 strain expressing stGFP (Lawrence *et al.*, 2007) from a pBAD plasmid.

Cell growth and protein expression
L. lactis was grown in the chemically defined medium (CDM) with 1% (w/v) glucose as a carbon and energy source and 5 µg mL^{-1} chloramphenicol to maintain the plasmids, with some modifications (see Supplemental Information for medium composition). Riboflavin was left out of the medium to reduce the fluorescence background during imaging; proline was left out to reduce the osmoprotective capacity. The altered medium is referred as CDM^{RP}. Cells were grown at 30 °C without shaking of the culture. A day before the experiment a culture was inoculated from glycerol stocks. This culture was used to inoculate 4 ml of fresh CDM^{RP} on the day of the experiment up to OD$_{600}$= 0.1. The fresh culture was grown to OD$_{600}$= 0.3-0.5, after which protein expression was induced by adding 1:1000 volume of

L. lactis NZ9700 filtered culture supernatant containing 10 µg/mL nisin A. For the single molecule tracking (SMT) experiments, no inducer was added to keep the expression low. FRAP and SMT measurements were done within 1 and 3 hours after induction or mock incubation.

Sample preparation for FRAP and SMT measurements
The cells were harvested and resuspended either in CDM^{RP} or potassium and sodium phosphate-based media (200 mM sodium or potassium phosphate, pH 6.5, with 1 % (w/v) glucose and the appropriate amount of KCl or NaCl to obtain the desired osmolality). For osmotic upshift experiments the osmolality of the media was increased by the addition of sodium chloride. The osmolality of all solutions was measured by determination of their freezing point, using an Osmomat 030 cryoscopic osmometer (Gonotec, Germany), and indicated in the Supplementary Information (Tables S1, S6 and S7).

To determine the effect of osmoprotectants, the media were further supplemented with either: (1) 6 mM proline plus 1 mM glycine betaine in the case of osmotic upshift in CDM^{RP}, (2) 2.7 mM glutamine (the same concentration as in CDM^{RP}) in the case of experiments in potassium and sodium phosphate-based media or (3) the full complement of amino acids present in CDM^{RP} in the case of osmotic upshift in potassium and sodium phosphate-based media. Subsequently, a small volume of cells (2 µL) was placed onto coverslips coated with 0.1 % (w/v) poly-L-lysine (to hinder cell motion) and FRAP and SMT measurements were carried out for periods not extending 25 minutes after which a fresh sample was prepared.

FRAP measurements
The fluorescence recovery after photobleaching (FRAP) data was recorded on a LSM 710 confocal laser scanning microscope (CLSM, Carl Zeiss, Jena, Germany) with a C-apochromat 40x/1.20 w Korr M27 water immersion objective with NA of 1.2. Both the photobleaching and cell imaging were carried out using the 488 nm laser (at different intensities). Fluorescence emission was collected from 493 to 600 nm. The FRAP protocol is based on the original approach introduced by Elowitz and co-workers (Elowitz *et al.*, 1999) and also described by Mika and co-workers (Mika *et al.*, 2011a), with the exception that a confocal laser scanning microscope (CLSM) is used instead of a wide-field set-up. Briefly, a zoomed–in image of the cell was acquired using low laser intensity (Fig. 1, panels labeled "before"). Subsequently, an area at the pole of the bacterium was photobleached using a diffraction-limited laser beam of high intensity (Fig.1, red dotted circle). Immediately after that a series of images was collected using the low intensity laser beam to capture the fluorescence recovery process (Fig. 1, panels labeled "after"). The speed of the acquisition of fluorescence recovery images was adjusted to the osmotic condition used and the rate of the diffusion process. To capture fast diffusion (recovery complete within 50 ms), the image acquisition speed was increased by reducing the number of pixels per frame to 12x12 and the highest scanning speed of the CLSM was used; image acquisition of 6 ms per frame. When *L. lactis* cells were resuspended in media of high osmolality the acquisition time per frame was extended because the diffusion is slow. In the case of the highest osmotic upshift conditions intervals were introduced between frames to capture the entire recovery process. All FRAP measurements were carried out at 20 ^{o}C.

Diffusion coefficients were calculated from the FRAP measurements as reported before (Elowitz *et al.*, 1999; Mika *et al.*, 2011a). In short, a line was drawn along the longer axis of the cell and fluorescence intensities were extracted for each frame (Fig. 1, red dotted line). To take into account the intrinsic inhomogeneity of the fluorescence intensity along this line, the fluorescence intensity distribution in the image before photobleaching (Fig. 1, panels "before") was used to normalize all the measured distributions. To obtain the diffusion coefficient, home-written software simulated the normalized fluorescence intensity distributions along the given cross-section during the recovery process (Fig. 1, panels with heat maps), and the calculated fluorescence recovery distribution (Fig.1 "fluorescence

40

intensity fit") were compared with the actual data (Fig. 1 "fluorescence intensity experiment"). For the simulation we used the 1-dimensional diffusion equation:

$$(1) \quad \frac{\partial I(x,t)}{\partial t} = D \frac{\partial^2 I(x,t)}{\partial x^2},$$

where I is fluorescence intensity and *D* is the diffusion coefficient; with boundary conditions:

$$(2) \quad \frac{\partial I(x,t)}{\partial x} = 0,$$

at the poles of the cell, corresponding to zero flux of GFP through the cell membrane. For final renormalization, the fluorescence intensity of GFP before photobleaching (Fig. 1 panels "before") was used. The same approach was used for determining the lateral diffusion of membrane proteins (Kumar *et al.*, 2010) (Fig. 1B). The FRAP method for determining membrane protein diffusion coefficients was validated by performing diffusion simulations with the program Smoldyn (Andrews *et al.*, 2010). The cell geometry was estimated by a spherocylinder with a radius and a cylinder length of 0.5 μm. At the start of the simulation the membrane of one half of the cell was homogeneously populated with 4000 particles. Each of these particles performs a random walk with a time step of 0.2 ms. The step sizes for the random walk depend on the diffusion coefficient. We performed simulations with four different diffusion coefficients ranging from 0.003-0.1 μm^2/s. The total simulated time was 40 s and every 2 s the coordinates of all particles were extracted. Simulated microscopy images were made by summing the particles in 100 by 100 nm sized bins. These images and Gaussian blurred versions of them (to take into account the diffraction limit) were analyzed in the same way as the real data. Examples of FRAP measurements are shown in Fig. 1.

Single molecule tracking of membrane protein diffusion
For the single particle tracking we used *L. lactis* NZ9000 expressing LacSΔIIA-YPet fusion. All measurements were performed on a home-build setup based on an Olympus IX-81 microscope with a 60x objective with a NA of 1.49 and equipped with a continuous wave 514 nm laser (Coherent, Santa Clara,CA , USA) to excite YPet. Emitted light from 525 to 555 nm was recorded on an EM-CCD camera (Hamamatsu Photonics, Hamamatsu, Japan). Cells were grown as described above except that no inducer was used. Even in the absence of inducer most cells contain one or a few fluorescent foci that could be tracked. All data were recorded 1 to 3 hours after harvesting of the cells. For imaging, cells were immobilized on poly-L-lysine coated coverslips. Before adding poly-L-lysine the coverslips were cleaned by sonication for 30 min in 1M KOH, 30 min in acetone and again 30 min in 1M KOH, followed by rinsing of the coverslips with distilled water. All measurements were performed at 20 °C. Data was recorded at different frame rates: 34 ms, 78 ms and 150 ms. In all cases the exposure time was 31 ms. For every frame rate, 3 or 4 fields of cells were imaged for 100 frames each. From these datasets cells were selected that showed a few foci. The foci were fitted with a 2D Gaussian function to obtain the position at subpixel resolution. Step size distributions were made separately for x and y coordinates (these coordinates refer to a position on the image). These were fitted with an analytical solution to the 1*D* diffusion equation to extract the diffusion coefficient, *D* (Phillips *et al.*, 2009):

$$(3) \quad p(x) \; = \; \frac{N}{\sqrt{4\pi D \tau}} e^{-\frac{x^2}{4D\tau}},$$

where N and D are fitting parameters, τ indicates the frame rate and x is the step size. Fitting was performed in Mathematica. To extract the diffusion coefficients, it is assumed that the molecules are

diffusing in a plane. In reality the LacSΔIIA-YPet molecules move along a membrane that is curved. To estimate the error arising from the cell geometry and from localization error, we performed simulations of the diffusion of membrane bound particles with Smoldyn (Andrews *et al.*, 2010). In each simulation 1200 particles were positioned randomly over the simulated cell membrane. For each of these particles a random walk was performed with time steps of 0.01 ms. The size of the steps in the random walk depends on the input diffusion coefficient. At regular intervals the coordinates for all particles were extracted for analysis. The analysis was performed in the same way as above except that x and y step sizes were pooled before fitting. To take into account the localization error, the x and y coordinates were modified before analysis by adding a value picked from a Gaussian distribution. A list of simulations and results is provided in the Supplemental Information (Table S4). We performed simulations on spherical and spherocylindrical geometries and with different input diffusion coefficients (0.003-0.3 μm^2/s). The positions of the particles were saved every 0.05-0.25 s. For each of these simulations 4 analyses were performed with an average localization error of 0, 0.019, 0.038 and 0.076 μm, respectively (the localization error distribution was assumed to be Gaussian).

Cell size measurements

The *L. lactis* strain expressing GFP was grown and induced in the same way as for the diffusion measurements. Cells were resuspended in CDM^{RP} medium with varying amounts of NaCl (from 0 to 4 M). Cells were placed on coverslips coated with dichlorodimethylsilane and imaged for no more than 20 min. Dichlorodimethylsilane coverslips were made by sonicating the slides for 30 min in a solution of 5 M KOH in MilliQ and then incubating the slides for 5 min in aceton with 2 % dichlorodimethylsilane. Imaging was done on a Zeiss AxioObserver Z1. Brightfield and fluorescence images (470 nm excitation) were made for 20 fields of cells per condition (N_{cells} = 50 to 1028, mean = 352). The fluorescence images were used to estimate the size of the cytoplasm and the brightfield images were used as a guide for removing particles that were incorrectly selected. Automated cell selection was performed with MicrobeTracker (Sliusarenko *et al.*, 2011).

For *E. coli* cell size determination we used an MG1655 strain expressing stGFP from a pBAD plasmid. A glycerol stock was used to inoculate an LB culture with an osmolality of 0.24 Osm (containing 100 μg mL^{-1} ampicillin). The culture was incubated overnight at 30 °C in a shaker. The next morning this culture was used to inoculate MOPS based minimal medium (0.24 Osm, 100 μg mL^{-1} ampicillin). From the start of the culture 0.1 % L-arabinose was present to induce the expression of stGFP. Incubation at 30 °C in a shaker. The culture was grown to an OD of 0.3-0.4. After that the culture stood at 20 °C for no longer than 1 h. All measurements were done within this time (also at 20 °C). Cells from 1.5 mL of culture were resuspended twice in 250 μL shock medium. The shock media were MBM without glucose and potassium phosphate but with a higher concentrations of NaCl. Cells were deposited on (3-aminopropyl)triethoxysilane coated cover slides and imaged in the same way as indicated above. The cover slides were prepared in same way as the dichlorodimethylsilane slides.

Statistical data analysis

For each experimental condition, the diffusion coefficients were characterized by a rather broad distribution in the population of tested cells (Fig. 2 and 3), similar to what has been observed previously in *E. coli* (Elowitz *et al.*, 1999; Konopka *et al.*, 2006; van den Bogaart *et al.*, 2007; Konopka *et al.*, 2009; Mika *et al.*, 2010). Following previous studies (Konopka *et al.*, 2009; Mika *et al.*, 2010), we have taken the median. To ensure sufficient statistical coverage of this distribution in the FRAP measurements we analysed between 22 and 84 cells per condition. For exact numbers see Supplementary Information (Tables S1, S6 and S7).

To determine whether or not the diffusion coefficients for GFP between sister cells of *L.lactis* were more similar than between randomly chosen cell pairs, we reexamined a large fraction of the diffusion data. For each experimental condition the differences in diffusion coefficients between the sister cells

were taken (referred to as "real pairs") and the median was calculated. Next, a median of the differences of diffusion coefficients of all artificial pairs was calculated for a given condition. The quotient Q was obtained by dividing the first median of diffusion coefficients differences by the latter one:

$$(4) \quad Q = \frac{median(\Delta D_{real\_pair\_\#1}, \Delta D_{real\_pair\_\#2}, \dots)}{median(\Delta D_{artificial\_pair\_\#1}, \Delta D_{artificial\_pair\_\#2}, \dots)}.$$

To determine whether the observed difference between sister cells and randomly chosen pairs was statistically significant, we determined all differences between all cells. Now, instead of determining Q as described in equation 4, ten randomly picked differences (not including real pairs) were compared to all differences. This was done 10^5 times for each dataset and yielded 10^5 distributions of Q values to compare to the real pair distribution. To facilitate this comparison, a new distribution was made for all median values of the distributions of Q values (see Figure 4B). If this distribution is very broad and the median Q we find for the real pairs lies within the main body of this distribution, then the real pair median Q has arisen by chance. If, however, the real pair value is at the edge of the distribution it is unlikely that it is due to chance and there is a real similarity between sister cells.

Acknowledgements

Condition	Osmolality[1] (Osm)	Δ Osmolality[2] (Osm)	D_{MEDIAN} (μm^2/s)	IQR[3] (μm^2/s)	Range[4] (μm^2/s)	n[5]
CDM	0.53	0	**7**	5.1 - 10	1.4 - 18	84
CDM + NaCl	0.99	0.46	**3**	1.9 - 4.8	1.2 - 8.2	24
CDM + NaCl	1.44	0.91	**0.838**	0.48 - 0.92	0.25 - 2.7	34
CDM + NaCl	1.93	1.4	**0.23**	0.14 - 0.44	0.034 - 3.9	40
CDM + NaCl	2.25	1.72	**0.087**	0.041 – 0.72	0.015 - 4.7	46
Pi buffers	0.54	0.01	**4.3**	2.9 - 6.8	0.29 - 15	80
Pi buffers + Salt[6]	1.01	0.48	**0.41**	0.28 - 0.68	0.061 - 1.9	44
Pi buffers + Salt[6]	1.36	0.83	**0.11**	0.068 – 0.21	0.025 – 0.3	41
Pi buffers + NaCl	2.36	1.83	**0.091**	0.037 - 0.29	0.016 - 2.3	41

Table S1. Diffusion coefficients of GFP in the cytoplasm of *Lactococcus lactis* at standard osmotic conditions of cell growth and upon osmotic upshift in CDM[RP] or phosphate-based media.

1) Osmolality of the medium determined as described in the Experimental Procedures section
2) Osmotic upshift: difference in osmolality between the measurement condition and the osmolality of the growth medium
3) The interquartile range (IQR), giving values containing the middle half of the data
4) Minimal and maximal *D* values measured
5) Number of individual cells measured
6) Salt is either KCl or NaCl, since there was no difference between data sets, they were combined

Condition	Osmolality[1] (Osm)	D_{MEDIAN} (μm^2/s)	IQR[2] (μm^2/s)	Range[3] (μm^2/s)	n[4]
CDM	0.53	**0.78**	0.49 - 1.8	0.13 - 2.9	25
CDM	0.53	**0.68**	0.40 - 1.1	0.13 – 2.8	28

Table S2: Diffusion coefficients of β-galactosidase-GFP in the cytoplasm of *Lactococcus lactis* at standard osmotic conditions.

1) Osmolality of the medium determined as described in the Experimental Procedures section
2) The interquartile range (IQR), giving values containing the middle half of the data
3) Minimal and maximal *D* values measured
4) Number of individual cells measured

Protein	D_{MEAN} ($\mu m^2/s$)	IQR[1] ($\mu m^2/s$)	Range[2] ($\mu m^2/s$)	n[3]
LacS – GFP	**0.020**	0.01 – 0.022	0.0038 – 0.11	22
BcaP – GFP	**0.019**	0.013 – 0.024	0.0044 – 0.034	27
GFP[4]	**7.8**	5.1 – 10	1.4 – 18	84

Table S3. Lateral diffusion of GFP-tagged membrane proteins in the plasma membrane of *Lactococcus lactis* at standard osmotic conditions measured with FRAP.

1) The interquartile range (IQR), giving values containing the middle half of the data
2) Minimal and maximal *D* values measured
3) Number of individual cells measured
4) Diffusion coefficient of cytosolic GFP at standard osmotic conditions of growth (CDM medium) is given for comparison

D_{in} (μm²/s)	Simulation time step (ms)	Analysis time step (s)	Geometry	D_{output}/D_{input}			
				Error = 0 μm	Error = 0.019 μm	Error = 0.036 μm	Error = 0.076 μm
0.03	**0.001**	0.05	Spherocylinder	0.80	0.60	1.07	2.40
0.03	**0.01**	0.05	Spherocylinder	0.80	0.60	1.00	2.40
0.003	0.01	0.05	Spherocylinder	0.88	1.73	5.00	18.33
0.01	0.01	0.05	Spherocylinder	0.78	0.88	1.90	6.00
0.03	0.01	0.05	Spherocylinder	0.80	0.60	1.00	2.40
0.1	0.01	0.05	Spherocylinder	0.78	0.51	0.65	1.10
0.3	0.01	0.05	Spherocylinder	0.73	0.47	0.50	0.70
0.03	0.01	**0.05**	Spherocylinder	0.80	0.60	1.00	2.40
0.03	0.01	**0.15**	Spherocylinder	0.83	0.57	0.67	1.13
0.03	0.01	**0.25**	Spherocylinder	0.77	0.50	0.60	0.90
0.03	0.01	0.05	**Sphere**	0.80	0.60	0.97	2.37
0.03	0.01	0.05	**Spherocylinder**	0.80	0.60	1.00	2.40
0.03	0.01	0.05	**Long Spherocylinder**	0.73	0.67	1.00	2.27

Table S4. Results from single molecule tracking (SMT) simulations.

Simulations of the SMT experiment indicate that with our experimentally determined diffusion coefficients (0.03 μm²/s) and localization error (~0.03 μm), we are most likely not more than 2-fold off in estimating the diffusion coefficient when the exact cell geometry is not taken into account. The results are expressed as the ratio of output diffusion coefficient (D_{output}) and input diffusion coefficient (D_{input}). The errors indicated are absolute averages of Gaussian distributions. Details about how the results were obtained can be found in the Experimental Procedures.

		D_{output}/D_{input}	
D_{in} (μm^2/s)	**Raw**	**Gaussian blur (σ=1 pixel)**	**Gaussian blur (σ=1.5 pixel)**
0.003	0.90	0.83	0.87
0.01	0.81	0.95	0.99
0.03	0.80	0.93	0.97
0.1	0.77	0.86	0.92

Table S5. Results from simulations of FRAP on membrane proteins.

The simulations indicate that the assumption that the cell membrane is flat does not significantly affect the determination of the diffusion coefficient by FRAP. The results are expressed as the ratio of output diffusion coefficient (D_{output}) and input diffusion coefficient (D_{input}). Pixel size is 100 nm. Details about how the results were obtained can be found in the Experimental Procedures.

Condition	Osmolality[1] (Osm)	Δ Osmolality[2] (Osm)	D_{MEDIAN} ($\mu m^2/s$)	IQR[3] ($\mu m^2/s$)	Range[4] ($\mu m^2/s$)	n[5]
Pi-based buffers + salt[6]	1.36	0.83	**0.11**	0.068 – 0.21	0.025 – 0.3	41
Pi-based buffers + salt + 2.7mM glutamine	1.41	0.88	**0.29**	0.17 – 0.60	0.027 – 4.8	75
Pi-based buffers + salt + amino acid mix	1.42	0.89	**0.63**	0.35 – 1.3	0.073 – 4.42	73
CDM^{RP} + NaCl	1.44	0.91	**0.84**	0.48 – 0.92	0.25 – 2.7	34
CDM^{RP} + NaCl + 6mM proline + 1mM glycine-betaine	1.37	0.84	**2.4**	1.2 – 3.7	0.32 – 8.3	45
CDM^{RP} [7]	0.53	0.00	**7**	5.1 – 10	1.4 – 18	84

Table S6. Diffusion coefficients of GFP in the cytoplasm of osmotically-stressed *Lactococcus lactis* cells (Δ osmolality ~0.87 Osm) in CDM^{RP} or phosphate-based media and in the presence of osmoprotectants.

1) Osmolality of the medium determined as described in the Experimental Procedures section
2) Osmotic upshift: difference in osmolality between the measurement condition and the osmolality of the growth medium
3) The interquartile range (IQR), giving values containing the middle half of the data
4) Minimal and maximal *D* values measured
5) Number of individual cells measured
6) Salt is either KCl or NaCl, since there was no difference between data sets, they were combined
7) Standard osmotic growth condition, corresponding to that of CDM, is given for comparison

Condition	Osmolality[1] (Osm)	Median $\Delta D_{ARTIFICIAL\ PAIRS}$[2]	Median $\Delta D_{REAL\ PAIRS}$[3]	$n_{REAL\ PAIRS}$	N[4]	Q[5]
CDM	1.4	0.31	0.43	10	26	**1.37**
CDM	2.27	0.25	0.12	7	27	**0.48**
CDM	1.93	0.22	0.12	17	40	**0.52**
KPi + KCl	1.41	0.10	0.07	6	20	**0.69**
NaPi + NaCl	1.38	0.07	0.07	8	20	**1.00**
NaPi + NaCl + glucose	1.38	0.05	0.03	9	21	**0.56**
CDM + proline + glycinebetaine	1.39	1.70	1.30	7	19	**0.76**
KPi+ KCl + glucose	1.34	0.15	0.10	8	20	**0.64**
KPi + KCl + glucose + AA mix	1.36	0.83	0.56	10	21	**0.67**
KPi + KCl + glucose + glutamine	1.37	0.20	0.10	10	22	**0.52**
NaPi + NaCl + glucose + AA mix	1.39	0.59	0.54	12	24	**0.92**
NaPi + NaCl + glucose + glutamine	1.37	0.44	0.30	11	23	**0.68**
NaPi + NaCl + glucose	0.56	0.63	0.20	13	28	**0.32**
NaPi + NaCl + glucose	1.02	0.27	0.19	11	24	**0.70**
NaPi + NaCl + glucose	2.39	0.10	0.13	9	21	**1.34**

CDM	0.51	2.90	2.80	10	20	**0.97**
CDM	0.55	1.60	1.55	10	20	**0.97**
NaPi + NaCl + glucose	0.52	3.45	1.25	10	20	**0.36**
KPi + KCl + glucose	0.55	3.40	1.55	10	20	**0.46**
CDM	0.55	2.00	2.75	9	18	**1.38**
NaPi + NaCl + glucose	0.52	1.70	1.20	10	20	**0.71**
KPi + KCl + glucose	0.55	1.80	1.85	10	20	**1.03**

Table S7. Similarity of diffusion coefficients of GFP in the cytoplasm of sister cells of *Lactococcus lactis* under various conditions for individual datasets.

1) Osmolality of the medium was determined as described in the Experimental Procedures section
2) Median of the differences of D_{GFP} values between sister cells
3) Median of the differences of D_{GFP} values between all possible pairs in a given dataset
4) Number of individual cells measured
5) The quotient (Q) is obtained by dividing the median $\Delta D_{REAL\ PAIRS}$ by the median $\Delta D_{ARTIFICIAL\ PAIRS}$ (see Experimental Procedures section); if Q < 1, the difference in the D_{GFP} of sister cells is smaller than with other cells in the given data set (total population)
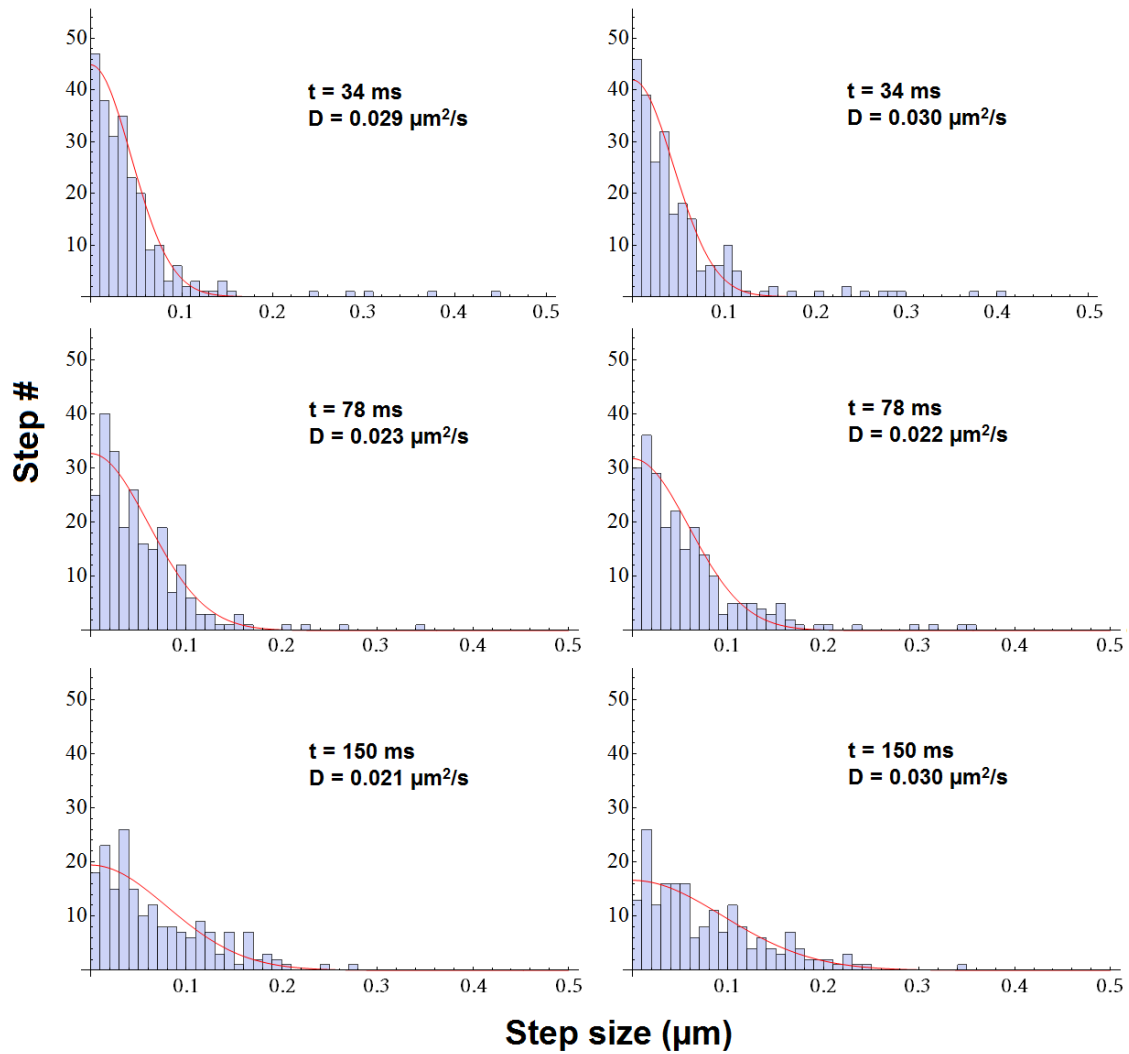
Figure S1: Step size distributions and fits for LacS-YPet diffusion in the plasma membrane of *Lactococcus lactis*, probed by single molecule tracking (SMT).

SMT measurements of the diffusion coefficient of LacS-YPet in the *L. lactis* membrane (D = 0.026 μm^2/s) yield similar values as obtained by FRAP. The plots on the left and right are step sizes along the x- and y-axes, respectively (the x- and y-axes of the images not the cells). The data (in blue) is fitted with a Gaussian function (in red; see Experimental procedures). Time steps and diffusion coefficients are indicated.
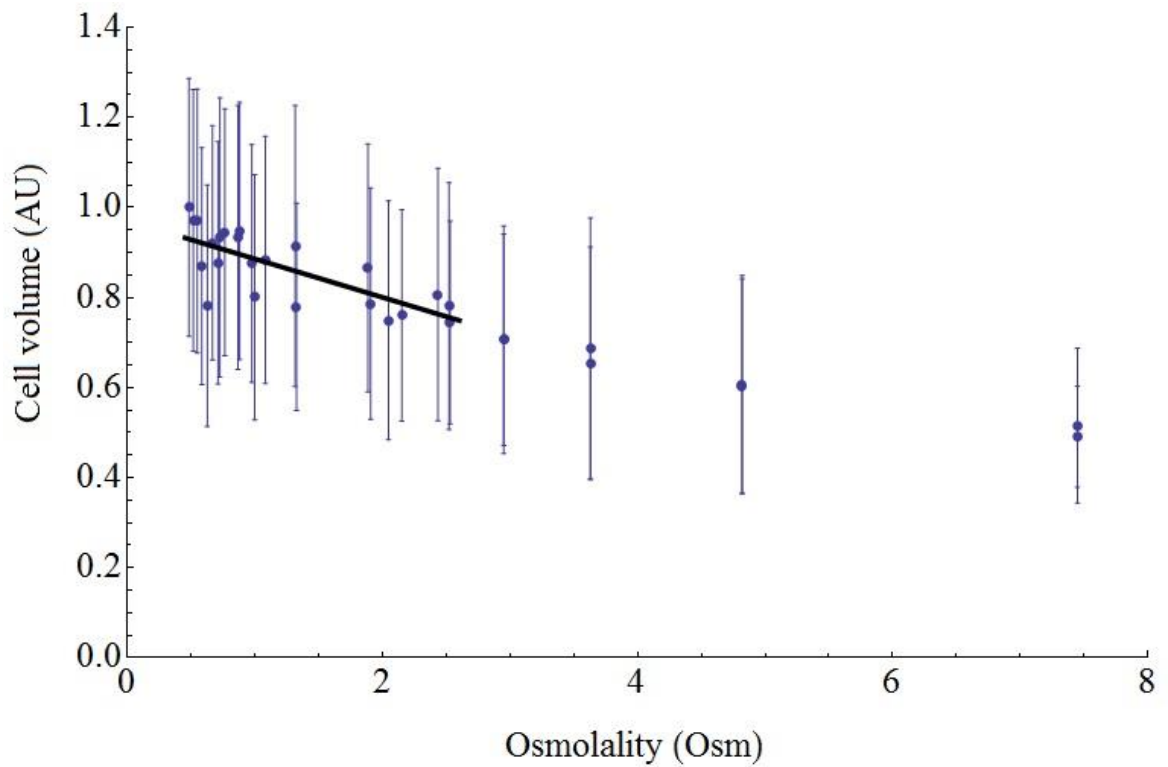
Figure S2: *Lactococcus lactis* **cell volume as a function of the shock medium osmolality.**

We show by fluorescence microscopy that cell volume drops linearly with the osmolality of the growth medium. The blue points indicate mean cell volumes. The error bars are standard deviations. The black line is a linear fit. This fit is used in Figure 5 to plot the volume *versus* the diffusion coefficient. The cell volume data is normalized to 1. The cell volumes were determined by fluorescence microscopy after resuspending cells in growth medium with higher levels of NaCl. Details on data collection can be found in the Experimental procedures section.

Figure S3: Magnitude of the difference in diffusion coefficient between sister cells.

The sister cell experiment was simulated by defining a set of 22 means and standard deviations, which correspond to the real data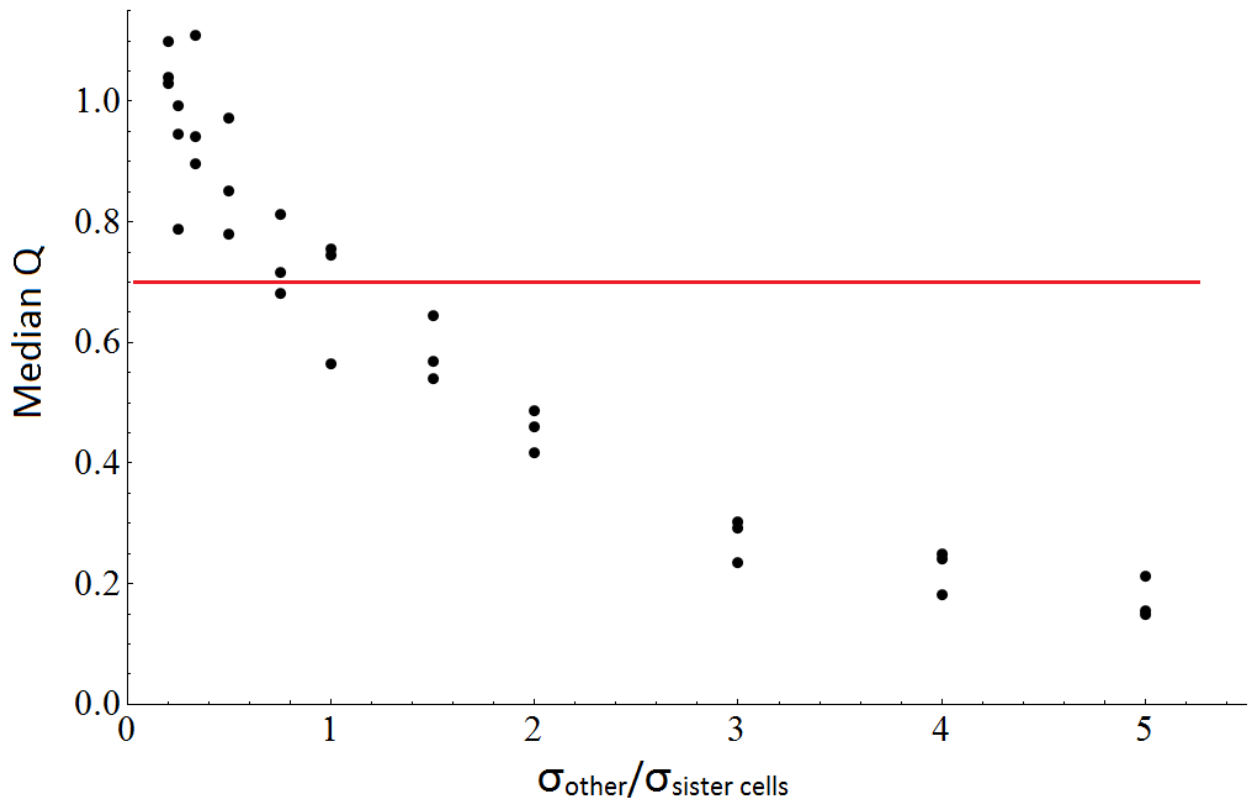sets that are shown in Table S7. We assume that the Log10 of the diffusion coefficients follow Gaussian distributions. We defined two other distributions for each dataset, with standard deviations σ_{other} and $\sigma_{sister\ cells}$, that when convoluted yield the experimentally defined distributions. By varying the ratio $\sigma_{other}/\sigma_{sister\ cells}$, we could vary the difference between sister cells. We performed three simulations for each ratio, each time picking randomly 10 pairs following the defined distributions. The higher this ratio the more similar the diffusion coefficient in the sister cells. The median Q we find in the real data is 0.7 (indicated by red line), which indicates that the variation between sister cells is slightly more than half of the total variation in diffusion coefficients.
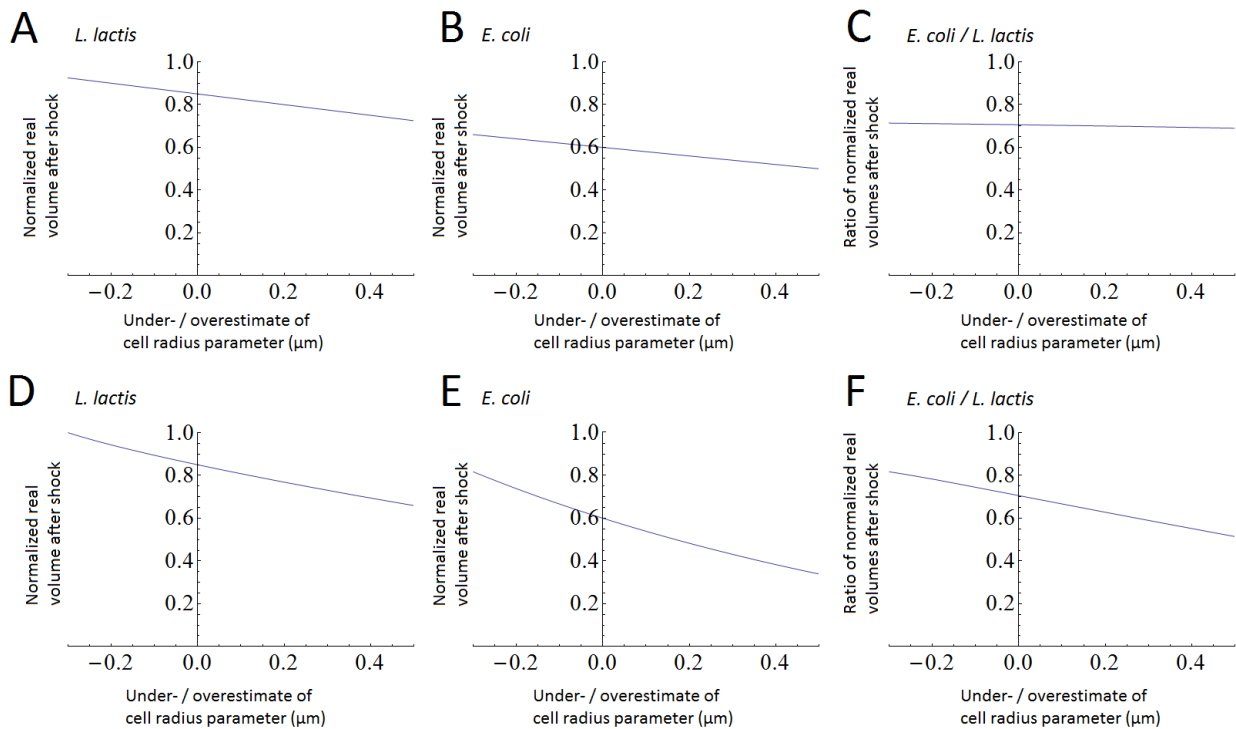
Figure S4: The impact of errors in fitting cell outlines on the determination of relative cell volumes.

The impact of errors in the determination of cell outlines on the determination of relative volumes was investigated on model *L. lactis* and *E. coli* cells. We approximated the geometry of both cells by spherocylinders. For *L. lactis* the radius was 0.3 µm and the cylinder length was 0.4 µm. For *E. coli* the radius was 0.5 µm and the cylinder length was 2 µm. This significantly exaggerates the difference between *L. lactis* and *E. coli*. Under our (normal) growth conditions the cell widths are similar (~0.8 µm) and the lengths are ~1.7 and ~2.1 µm for *L. lactis* and *E. coli*, respectively. We calculated the error in relative volume for the case that the osmotic stress causes only the length of the cell to decrease, panels A, B and C, and when the radius of the spherocylinder decreases (affecting both length and width of the cell), panels D, E and F. For *E. coli* the cells shrink mostly in their length, while for *L. lactis* the cell also somewhat shrink in width. A and B) The normalized real volume after shock was plotted against the error in the measured radius for both *L. lactis* and *E. coli*. We looked at the most extreme volume drops, 0.85 and 0.6 relative volume after osmotic upshift for *L. lactis* and *E. coli,* respectively. Here, the drop in volume is caused solely by changing the length of the cell. C) Ratio of the two lines in A and B. D) Same as in A and B except now the change in volume upon osmotic upshift was caused by a change in the radius (which affects both width and length). F) Ratio of the two lines in D and E.
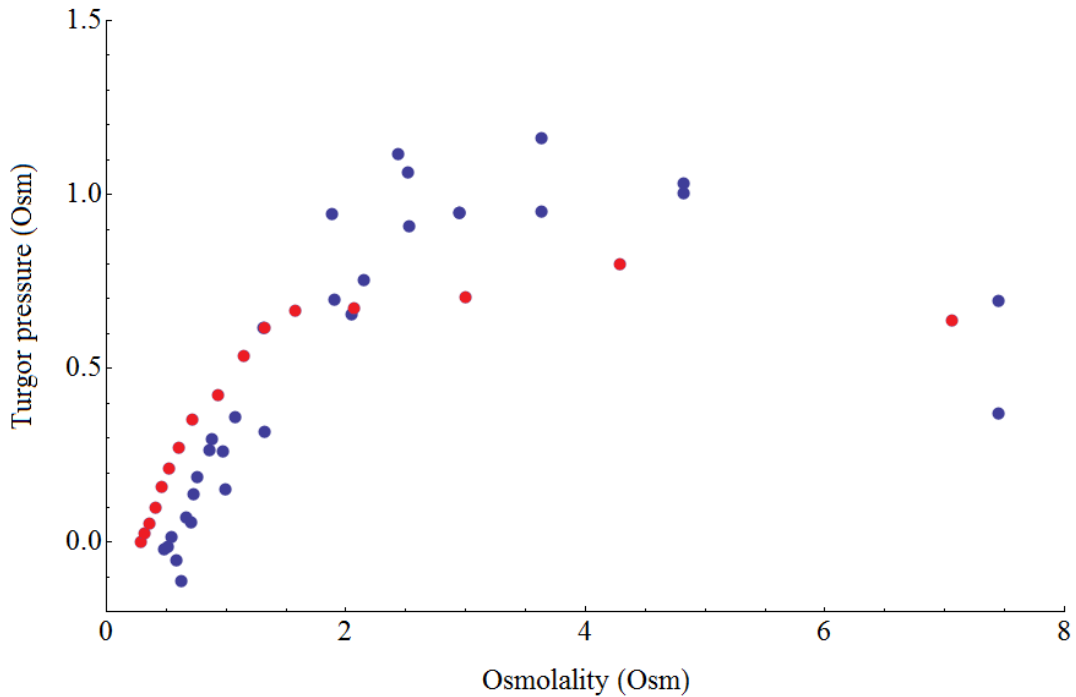
Figure S5: Turgor pressure in *Lactococcus lactis* and *Bacillus subtilis*.

The calculated turgor pressure in relation to the medium osmolality used in the calculation. Red: *B. subtilis* (Whatmore *et al.*, 1990), Blue: *L. lactis* (points were calculated from the osmolality vs volume data shown in Figure S2). At higher osmolalities the cell volume is a linear function of 1/Osm and the calculated turgor pressure should reach a plateau. The turgor pressure at the plateau is the turgor pressure in the cell under growth conditions. The turgor pressure was calculated by:

$$P_{turgor} = Osm_{cyto} - Osm_{gm} = \frac{(Vol_{oum} - NOV) \times Osm_{sm}}{(Vol_{gm} - NOV)} - Osm_{gm}$$ (Whatmore *et al.*, 1990).

where cyto = cytoplasm, gm = growth medium, Vol = cell volume, oum = osmotic upshift medium, and NOV = non-osmotic volume. The non-osmotic volume is taken from the intercept with the y-axis in a 1/Osm vs cell volume plot. The *B. subtilis* turgor pressure is ~0.75 Osm. The *L. lactis* turgor pressure is similar.
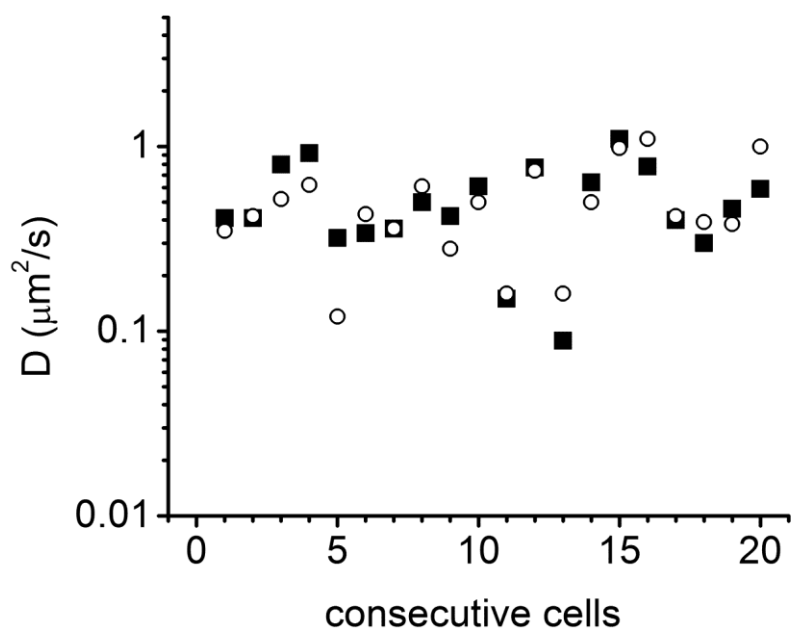
Figure S6. Error of the FRAP measurement.

Dual measurements on the same cells suggest that some of the spread in the FRAP data is real variation rather than measurement error. Plot of diffusion coefficients of GFP measured by FRAP in 20 *L. lactis* cells in phosphate-based media with an osmolality of 1 Osm. Measurement on each cell was performed twice and the outcome of the first measurement is depicted as black square and of the second measurement as open circle.

Supplementary experimental procedures

CDM^{RP} composition.

CDM^{RP} is made by mixing 8 different stock solutions. For 1 L one mixes 920 mL basic medium, 50 mL amino acid mix, 10 mL vitamin mix, 10 mL base mix and 2.5 mL of the metal mixes 1-4. The medium is sterilized by passing it through a 0.2 µm filter. L-cysteine, 0.25 g L^{-1}, is added right before growth. Composition of the stock solutions:

	Stock concentration (g/L)		Stock concentration (g/L)
Basic medium (pH 6.5)		4-aminobenzoic acid	1
Tyrosine	0.27	D-biotin	1
KH$_2$PO$_4$	12.5	Folic acid	0.1
K$_2$HPO$_4$	23.1	Vitamin B12	0.1
Ammonium citrate	0.55	Orotic acid	0.5
Na-acetate x 3H$_2$0	0.92	2-deoxythymidine	0.5
Amino acid mix (pH 7.0)		Inosine	0.5
L-alanine	4.75	DL-6,8 thioctic acid	0.25
L-glutamine	7.8	pyridoxamine dihydrochloride	0.5
L-asparagine	7	pyridoxal HCl	0.2
L-arginine	2.5	Base mix	
L-lysine	8.75	Adenine	1
L-isoleucine	4.25	Uracil	1
L-methionine	2.5	Xanthine	1
L-phenylalanine	5.5	Guanine	1
L-serine	6.75	**Metal mix 1**	
L-threonine	4.5	MgCl$_2$	160
L-tryptophan	1	CalCl$_2$	20
L-valine	6.5	ZnSO$_4$ x 7H$_2$O	2
Glycine	3.5	**Metal mix 2**	
L-histidine	3	CuSO$_4$ x 5H$_2$O	1.2
L-leucine	9.5	CoCl$_2$	0.08
Vitamin mix (pH 7.0)		**Metal mix 3**	
Nicotinic acid	0.2	FeCl$_2$ x H$_2$O	2
Thiamine HCl	0.1	**Metal mix 4**	
Ca-(D+)panthothenate	0.1	MnSO$_4$	14

MBM composition

To prepare MBM the ten solutions listed in the table below are added to 0.36 L MilliQ. We used 0.2 % w/v glucose as carbon source.

	Component	Volume (L)	Concentration (M)
1	MOPS	0.4	1
2	Tricine	0.04	1
3	$FeSO_4$	0.01	0.01
4	NH_4Cl	0.05	1.9
5	K_2SO_4	0.01	0.276
6	$CaCl_2$	0.01	0.0005
7	$MgCl_2$	0.01	0.528
8	NaCl	0.1	5
9	Micronutrients	0.01	
	Ammonium molybdate		0.000006
	H_3BO_3		0.0004
	$CoCl_2$		0.00003
	$CuSO_4$		0.00001
	$MnCl_2$		0.00008
	$ZnSO_4$		0.00001
10	K_2HPO_4	0.01	0.132

Protein amino acid sequences

<u>GFP</u>

MGGGFAQFSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKL
PVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVK
FEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQKNGIKANFKIRHNIED
GSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGM
DELYK **TSENLYFQG**^{Linker} **HHHHHHHHHH**^{Histag}

<u>β-Galactosidase-GFP</u>

MGGGFAKGAVVLQRRDWENPGVTQLNRLAAHPPFASWRNSEEARTDRPSQQLRSLNGEWR
FAWFPAPEAVPESWLECDLPEADTVVVPSNWQMHGYDAPIYTNVTYPITVNPPFVPTENP
TGCYSLTFNVDESWLQEGQTRIIFDGVNSAFHLWCNGRWVGYGQDSRLPSEFDLSAFLRA
GENRLAVMVLRWSDGSYLEDQDMWRMSGIFRDVSLLHKPTTQISDFHVATRFNDDFSRAV
LEAEVQMCGELRDYLRVTVSLWQGETQVASGTAPFGGEIIDERGGYADRVTLRLNVENPK
LWSAEIPNLYRAVVELHTADGTLIEAEACDVGFREVRIENGLLLLNGKPLLIRGVNRHEH
HPLHGQVMDEQTMVQDILLMKQNNFNAVRCSHYPNHPLWYTLCDRYGLYVVDEANIETHG
MVPMNRLTDDPRWLPAMSERVTRMVQRDRNHPSVIIWSLGNESGHGANHDALYRWIKSVD
PSRPVQYEGGGADTTATDIICPMYARVDEDQPFPAVPKWSIKKWLSLPGETRPLILCEYA
HAMGNSLGGFAKYWQAFRQYPRLQGGFVWDWVDQSLIKYDENGNPWSAYGGDFGDTPNDR
QFCMNGLVFADRTPHPALTEAKHQQQFFQFRLSGQTIEVTSEYLFRHSDNELLHWMVALD
GKPLASGEVPLDVAPQGKQLIELPELPQPESAGQLWLTVRVVQPNATAWSEAGHISAWQQ
WRLAENLSVTLPAASHAIPHLTTSEMDFCIELGNKRWQFNRQSGFLSQMWIGDKKQLLTP
LRDQFTRAPLDNDIGVSEATRIDPNAWVERWKAAGHYQAEAALLQCTADTLADAVLITTA
HAWQHQGKTLFISRKTYRIDGSGQMAITVDVEVASDTPHPARIGLNCQLAQVAERVNWLG
LGPQENYPDRLTAACFDRWDLPLSDMYTPYVFPSENGLRCGTRELNYGPHQWRGDFQFNI

SRYSQQQLMETSHRHLLHAEEGTWLNIDGFHMGIGGDDSWSPSVSAXXQLSAGRYHYQLV
WCQK **E**^{Linker} **NLYFQGQFSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKL**
TLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKD
DGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQKNGIK
ANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLE
FVTAAGITHGMDELYK^{GFP} **TS**^{Linker} **HHHHHHHHHH**^{Histag}

LacS∆IIA-GFP

MGGGFA^{N-tail} EKSKGQMKSRLSYAAGAFGNDVFYATLSTYFIMFVTTHLFNTGDPKQNS
HYVLLITNIISILRILEVFIDPLIGNMIDNTNTKYGKFKPWVVGGGIISSITLLLLFTDL
GGLNKTNPFLYLVLFGIIYLVMDVFYSIKDIGFWSMIPALSLDSHEREKMATFARIGSTI
GANIVGVAIMPIVLFFSMTNNSGSGDKSGWFWFAFIVALIGVITSIAVGIGTREVESKIR
DNNEKTSLKQVFKVLGQNDQLMWLSLGYWFYGLGINTLNALQLYYFTFILGDSGKYSILY
GLNTVVGLVSVSLFPTLADKFNRKRLFYGAIAVMLGGIGIFSIAGTSLPIILTAAELFFI
PQPLVFLVVFMIISDSVEYGQWKTGHRDESLTLSVRPLIDKLGGAMSNWLVSTFAVAAGM
TTGASASTITTHQQFIFKLGMFAFPAATMLIGAFIVARKITLTEARHAKIVEELEHRFSV
ATSE **E**^{Linker} **NLYFQGQFSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKL**
TLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKD
DGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQKNGIK
ANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLE
FVTAAGITHGMDELYK^{GFP} **TS**^{Linker} **HHHHHHHHHH**^{Histag}

LacS∆IIA-YPet

MGGGFA^{N-tail} EKSKGQMKSRLSYAAGAFGNDVFYATLSTYFIMFVTTHLFNTGDPKQNS
HYVLLITNIISILRILEVFIDPLIGNMIDNTNTKYGKFKPWVVGGGIISSITLLLLFTDL
GGLNKTNPFLYLVLFGIIYLVMDVFYSIKDIGFWSMIPALSLDSHEREKMATFARIGSTI
GANIVGVAIMPIVLFFSMTNNSGSGDKSGWFWFAFIVALIGVITSIAVGIGTREVESKIR
DNNEKTSLKQVFKVLGQNDQLMWLSLGYWFYGLGINTLNALQLYYFTFILGDSGKYSILY
GLNTVVGLVSVSLFPTLADKFNRKRLFYGAIAVMLGGIGIFSIAGTSLPIILTAAELFFI
PQPLVFLVVFMIISDSVEYGQWKTGHRDESLTLSVRPLIDKLGGAMSNWLVSTFAVAAGM
TTGASASTITTHQQFIFKLGMFAFPAATMLIGAFIVARKITLTEARHAKIVEELEHRFSV
ATSE **E**^{Linker} **SKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKLLCTT**
GKLPVPWPTLVTTLGYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRA
EVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQKNGIKANFKIRHN
IEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSYQSALFKDPNEKRDHMVLLEFLTAAGIT
EGMNELYKEL^{YPet}

BcaP-GFP

MGGGFA^{N-tail} MGFMRKADFELYRDADKHYNQVLTTRDFLALGVGTIISTSIFTLPGQVA
AQFAGPGVVFSYLLAALVAGFVALAYAEMSTVMPFAGSAYSWISVLFGEGFGWIAGWALL
AEYFIAVAFVGSGFSANLQQLLAPLGFQLPKVLANPFGTDGGIVDIISLLVILLSAIIVF
RGASDAGRISQILVVLKVAAVIAFIIVGITVIKPANYHPFIPPHNPKTGFGGFSGIWSGV
SMIFLAYIGFDSIAANSAEAKNPQKTMPRGIIGSLLIAVVLFAAVTLVLVGMHPYSAYAG
NAAPVGWALQQSGYSVLSEVVTAIALAGMFIALLGMVLAGSRLLYAFGRDGLLPKGLGKM
NARNLPANGVWTLAIVAIVIGAFFPFAFLAQLISAGTLIAFMFVTLGIYSLRRRQGKDLP
EATYKMPFYPVLPALGFIGSLFVFWGLDVQAKLYSGIWFLIGIAIYFAYGNRRKKK **E**^{Linker}
NLYFQGQFSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKL
PVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVK
FEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQKNGIKANFKIRHNIED
GSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGM
DELYK^{GFP} **TS**^{Linker} **HHHHHHHHHH**^{Histag}

References

Aertsen, A., and Michiels, C. (2005) Diversify or die: Generation of diversity in response to stress. *Crit Rev Microbiol* 31: 69-78.

Anderson, J.B., Anderson, L.E., and Kussmann, J. (2010) Monte Carlo simulations of single- and multistep enzyme-catalyzed reaction sequences: Effects of diffusion, cell size, enzyme fluctuations, colocalization, and segregation. *J Chem Phys* **133:** 034104.

Andrews, S.S., Addy, N.J., Brent, R., and Arkin, A.P. (2010) Detailed Simulations of Cell Biology with Smoldyn 2.1. *PLoS Comput Biol* **6**(3): e1000705.

Balaban, N., Merrin, J., Chait, R., Kowalik, L., and Leibler, S. (2004) Bacterial persistence as a phenotypic switch. *Science* **305:** 1622-1625.

Balcells, C., Pastor, I., Vilaseca, E., Madurga, S., Cascante, M., and Mas, F. (2014) Macromolecular Crowding Effect upon in Vitro Enzyme Kinetics: Mixed Activation-Diffusion Control of the Oxidation of NADH by Pyruvate Catalyzed by Lactate Dehydrogenase. *J Phys Chem B* **118:** 4062-4068.

Booth, I.R. (2002) Stress and the single cell: Intrapopulation diversity is a mechanism to ensure survival upon exposure to stress. *Int J Food Microbiol* **78:** 19-30.

Cayley, D.S., Guttman, H.J., and Record, M.T. (2000) Biophysical characterization of changes in amounts and activity of Escherichia coli cell and compartment water and turgor pressure in response to osmotic stress. *Biophys J* **78:** 1748-1764.

Chow, D., Guo, L., Gai, F., and Goulian, M. (2012) Fluorescence Correlation Spectroscopy Measurements of the Membrane Protein TetA in Escherichia coli Suggest Rapid Diffusion at Short Length Scales. *PLoS ONE* **7**(10): e48600.

Cowan, A.E., Koppel, D.E., Setlow, B., and Setlow, P. (2003) A soluble protein is immobile in dormant spores of Bacillus subtilis but is mobile in germinated spores: Implications for spore dormancy. *Proc Natl Acad Sci U S A* **100:** 4209-4214.

Cowan, A., Olivastro, E., Koppel, D., Loshon, C., Setlow, B., and Setlow, P. (2004) Lipids in the inner membrane of dormant spores of Bacillus species are largely immobile. *Proc Natl Acad Sci U S A* **101:** 7733-7738.

Deich, J., Judd, E.M., McAdams, H.H., and Moerner, W.E. (2004) Visualization of the movement of single histidine kinase molecules in live Caulobacter cells. *Proc Natl Acad Sci U S A* **101:** 15921-15926.

Deng, Y., Sun, M., and Shaevitz, J.W. (2011) Direct Measurement of Cell Wall Stress Stiffening and Turgor Pressure in Live Bacterial Cells. *Phys Rev Lett* **107:** 158101.

deRuyter, P., Kuipers, O., Beerthuyzen, M., vanAlenBoerrigter, I., and deVos, W. (1996) Functional analysis of promoters in the nisin gene cluster of Lactococcus lactis. *J Bacteriol* **178:** 3434-3439.

Ellis, R. (2001) Macromolecular crowding: obvious but underappreciated. *Trends Biochem Sci* **26:** 597-604.

Elowitz, M.B., Surette, M.G., Wolf, P.E., Stock, J.B., and Leibler, S. (1999) Protein mobility in the cytoplasm of Escherichia coli. *J Bacteriol* **181:** 197-203.

Gershenson, A., and Gierasch, L.M. (2011) Protein folding in the cell: challenges and progress. *Curr Opin Struct Biol* **21:** 32-41.

Gierasch, L.M., and Gershenson, A. (2009) Post-reductionist protein science, or putting Humpty Dumpty back together again. *Nature Chem. Biol.* **5**(11)**:** 774-777.

Golding, I., Paulsson, J., Zawilski, S., and Cox, E. (2005) Real-time kinetics of gene activity in individual bacteria*. Cell* **123:** 1025-1036.

Guillon, L., Altenburger, S., Graumann, P.L., and Schalk, I.J. (2013) Deciphering Protein Dynamics of the Siderophore Pyoverdine Pathway in Pseudomonas aeruginosa. *PLoS ONE* **8**(10): e79111.

Halatek, J., and Frey, E. (2012) Highly Canalized MinD Transfer and MinE Sequestration Explain the Origin of Robust MinCDE-Protein Dynamics. *Cell Rep.* **1:** 741-752.

Ingham, C.J., Beerthuyzen, M., and Vlieg, J.v.H. (2008) Population Heterogeneity of Lactobacillus plantarum WCFS1 Microcolonies in Response to and Recovery from Acid Stress*. Appl Environ Microbiol* **74:** 7750-7758.

Keren, I., Shah, D., Spoering, A., Kaldalu, N., and Lewis, K. (2004) Specialized persister cells and the mechanism of multidrug tolerance in Escherichia coli*. J Bacteriol* **186:** 8172-8180.

Klumpp, S., Scott, M., Pedersen, S., and Hwa, T. (2013) Molecular crowding limits translation and cell growth*. Proc Natl Acad Sci U S A* **110:** 16754-16759.

Koch, A.L. (1984) Shrinkage of Growing Escherichia-Coli-Cells by Osmotic Challenge*. J Bacteriol* **159:** 919-924.

Konopka, M.C., Shkel, I.A., Cayley, S., Record, M.T., and Weisshaar, J.C. (2006) Crowding and confinement effects on protein diffusion in vivo. *J Bacteriol* **188:** 6115-6123.

Konopka, M.C., Sochacki, K.A., Bratton, B.P., Shkel, I.A., Record, M.T., and Weisshaar, J.C. (2009) Cytoplasmic Protein Mobility in Osmotically Stressed Escherichia coli*. J Bacteriol* **191:** 231-237.

Konopka, M.C., Weisshaar, J.C., and Record, M.T., Jr. (2007) Methods of changing biopolymer volume fraction and cytoplasmic solute concentrations for in vivo biophysical studies. *Methods in Enzymology* **428:** 487-504.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J Mol Biol* **305:** 567-580.

Kumar, M., Mommer, M.S., and Sourjik, V. (2010) Mobility of Cytoplasmic, Membrane, and DNA-Binding Proteins in Escherichia coli*. Biophys J* **98:** 552-559.

Lawrence, M.S., Phillips, K.J., and Liu, D.R. (2007) Supercharging proteins can impart unusual resilience. *J Am Chem Soc* **129:** 10110-10112.

Lidstrom, M.E., and Konopka, M.C. (2010) The role of physiological heterogeneity in microbial population behaviour. *Nature Chem. Biol.* **6:** 705-712.

Linden, M., Sens, P., and Phillips, R. (2012) Entropic Tension in Crowded Membranes. *PLoS Comput Biol* **8**(3): e1002431.

Llopis, P.M., Sliusarenko, O., Heinritz, J., and Jacobs-Wagner, C. (2012) In Vivo Biochemistry in Bacterial Cells Using FRAP: Insight into the Translation Cycle*. Biophys J* **103:** 1848-1859.

Mika, J.T., Krasnikov, V., van den Bogaart, G., de Haan, F., and Poolman, B. (2011a) Evaluation of Pulsed-FRAP and Conventional-FRAP for Determination of Protein Mobility in Prokaryotic Cells. *PLoS ONE* **6**(9): e25664.

Mika, J.T., and Poolman, B. (2011b) Macromolecule diffusion and confinement in prokaryotic cells. *Curr Opin Biotechnol* **22:** 117-126.

Mika, J.T., van den Bogaart, G., Veenhoff, L., Krasnikov, V., and Poolman, B. (2010) Molecular sieving properties of the cytoplasm of Escherichia coli and consequences of osmotic stress. *Mol Microbiol* **77:** 200-207.

Moran, U., Phillips, R., and Milo, R. (2010) SnapShot: Key Numbers in Biology. *Cell* **141:** 1262-1262.

Mullineaux, C.W., Nenninger, A., Ray, N., and Robinson, C. (2006) Diffusion of green fluorescent protein in three cell environments in Escherichia coli. *J Bacteriol* **188:** 3442-3448.

Nenninger, A., Mastroianni, G., Robson, A., Lenn, T., Xue, Q., Leake, M.C., and Mullineaux, C.W. (2014) Independent mobility of proteins and lipids in the plasma membrane of Escherichia coli. *Mol Microbiol* **92:** 1142-1153.

Nguyen, A., and Daugherty, P. (2005) Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nat Biotechnol* **23:** 355-360.

Phillips, R., Kondev, J., and Theriot, J. (2009). In *Physical Biology of the Cell*. New York: Garland Science, pp. 481-512.

Potma, E., de Boeij, W., Bosgraaf, L., Roelofs, J., van Haastert, P., and Wiersma, D. (2001) Reduced protein diffusion rate by cytoskeleton in vegetative and polarized Dictyostelium cells. *Biophys J* **81:** 2010-2019.

Ramadurai, S., Duurkens, R., Krasnikov, V.V., and Poolman, B. (2010) Lateral Diffusion of Membrane Proteins: Consequences of Hydrophobic Mismatch and Lipid Composition. *Biophys J* **99:** 1482-1489.

Ramadurai, S., Holt, A., Krasnikov, V., van den Bogaart, G., Killian, J.A., and Poolman, B. (2009) Lateral Diffusion of Membrane Proteins. *J Am Chem Soc* **131:** 12650-12656.

Ramadurai, S., Holt, A., Schafer, L.V., Krasnikov, V.V., Rijkers, D.T.S., Marrink, S.J., *et al.* (2010) Influence of Hydrophobic Mismatch and Amino Acid Composition on the Lateral Diffusion of Transmembrane Peptides. *Biophys J* **99:** 1447-1454.

Saffman, P., and Delbruck, M. (1975) Brownian-Motion in Biological-Membranes. *Proc Natl Acad Sci U S A* **72:** 3111-3113.

Sliusarenko, O., Heinritz, J., Emonet, T., and Jacobs-Wagner, C. (2011) High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. *Mol Microbiol* **80:** 612-627.

Soh, S., Banaszak, M., Kandere-Grzybowska, K., and Grzybowski, B.A. (2013) Why Cells are Microscopic: A Transport-Time Perspective. *J. Phys. Chem. Lett.* **4:** 861-865.

Swaminathan, R., Hoang, C., and Verkman, A. (1997) Photobleaching recovery and anisotropy decay of green fluorescent protein GFP-S65T in solution and cells: Cytoplasmic viscosity probed by green fluorescent protein translational and rotational diffusion. *Biophys J* **72:** 1900-1907.

van den Bogaart, G., Hermans, N., Krasnikov, V., and Poolman, B. (2007) Protein mobility and diffusive barriers in Escherichia coli: consequences of osmotic stress. *Mol Microbiol* **64:** 858-871.

Whatmore, A.M., and Reed, R.H. (1990) Determination of Turgor Pressure in Bacillus-Subtilis - a Possible Role for K+ in Turgor Regulation. *J Gen Microbiol* **136:** 2521-2526.

Zhang, G., Fedyunin, I., Miekley, O., Valleriani, A., Moura, A., and Ignatova, Z. (2010) Global and local depletion of ternary complex limits translational elongation. *Nucleic Acids Res* **38:** 4778-4787.

Zhou, H., Rivas, G., and Minton, A.P. (2008) Macromolecular crowding and confinement: Biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* **37:** 375-397.

Chapter 3: Ribosome surface properties may impose limits on the nature of the cytoplasmic proteome

The cell cytoplasm is a bustling place where much of the molecular motion is diffusive, possibly limiting the tempo of processes. Here, we study an under-examined aspect of protein mobility: the dependence on protein surface properties and ionic strength. We used a set of surface-modified fluorescent proteins (FPs) and determined their lateral diffusion coefficients (D) in the cytoplasm of the bacteria *Escherichia coli* and *Lactococcus lactis*, and the archaeon *Haloferax volcanii*. We find that in *E. coli D* depends on the net charge and its distribution over the protein, with positive proteins diffusing up to 100-fold slower than negative ones. This effect is weaker in *L. lactis* and *Hfx. volcanii* due to electrostatic screening. The decrease in mobility is probably caused by interaction of positive FPs with ribosomes. Ribosome surface properties may thus limit the composition of the cytoplasmic proteome. This finding lays bare a paradox in the functioning of bacterial (endo)symbionts.

Introduction

Many processes in biological cells depend on interactions between macromolecules (proteins and nucleic acids) and thus on the ability of these macromolecules to find each other by translational diffusion. This is especially important in prokaryotes because of the virtual absence of active mechanisms of cytoplasmic transport. It is clear that macromolecules need to diffuse for cells to function. To what extent the actual rate of this diffusion matters depends on the process under consideration and is in many cases unknown. For Brownian diffusion the rate of movement is characterized entirely by the diffusion coefficient, D. The exact value of the diffusion coefficient is important to the rate of a process only if it is diffusion limited, e.g. if the necessary conformational changes in an enzyme are faster than the diffusion of reactants. Arbitrarily lowering a diffusion coefficient, e.g. by osmotic stress, can make a process diffusion limited. Examples of diffusion-limited processes are binding of tRNA complexes to the ribosome, which leads to limitation in cell growth (1); and the binding of barstar to barnase, which we know to be diffusion limited because the proteins are designed to have an increased association rate by electrostatic interactions (2). Because protein diffusion is influenced by the environment we need to determine diffusion coefficients in the context of the cell.

The cytoplasm of cells is not only crowded with macromolecules (3) but also consists of various types of nucleic acids and >1000 types of protein (see proteome analysis below); though only 50 protein types make up 85% of the cytoplasmic proteome of *E. coli* (4). Various studies report on the presence of weak and nonspecific interactions between these components. NMR studies on proteins, either in the *E. coli* cytoplasm (5,6) or cell lysates (7), reveal that there are weak interactions between *E. coli* proteins and proteins cytochrome *c*, ubiquitin, and calmodulin. In a computational study on protein interactions it was found that in *E. coli* more highly expressed proteins are constrained in evolution to be less sticky (8), suggesting that nonspecific interactions are common and consequential. The transient macromolecular interactions *in vivo,* resulting from molecular evolution, are referred to as the quinary structure of proteins (9), and we discriminate these from the more generic nonspecific interactions that occur between molecules without coevolved interfaces.

In this study we set out to study the diffusion coefficients of proteins as a function of their surface properties and thus probe the boundary conditions for the generic nonspecific interactions. Our interest in this was piqued by four datasets from the literature. The first is the scattering of diffusion coefficients in the *E. coli* cytoplasm around a common downward trend when they are plotted against protein molecular weight; the dataset suggests that not only size (and shape) matter (10). Second, the diffusion coefficient of GFP is faster in the cytoplasm of osmotically-adapted *E. coli* cells than in osmotically-upshifted cells, even at similar cytoplasmic macromolecule volume fraction (11). Third, the diffusion coefficient of GFP drops much faster with cell volume (after an osmotic upshift) in *Lactococcus lactis* than in *E. coli* (12). Fourth, the slowing of diffusion in metabolic energy-depleted cells (13-15). In all four cases differential interactions of proteins with their surroundings may play a role, which are grounded in the surface properties of the macromolecules. Besides (possibly) giving insight into these four phenomena, studying the dependence of mobility on protein surface properties adds to our general quantitative understanding of diffusion; complementing studies on the relation between diffusion coefficients and protein size (10,12,16-18), diffusion coefficients and macromolecular crowding (11,12,18-20), and the dynamic structure of the cytoplasm (21,22).

Here, we use a set of GFP variants with a net charge that ranges from -30 to +25; we also studied two variants of +11 GFP that differ in the distribution of the charge over the surface. All diffusion coefficients were determined by fluorescence recovery after photo-bleaching (FRAP). We study these

proteins in the bacteria *Escherichia coli* and *Lactococcus lactis* and the archaeon *Haloferax volcanii*. These three organisms differ in their cytoplasmic ionic strength as shown by measurements on the dominant cation, K^+: *E. coli* (0.2 M) (23), *L. lactis* (0.8 M) (24) (note: *L. lactis* used to be called *Streptococcus cremoris*), and *Hfx. volcanii* (2.1 M) (25); these values are dependent on environmental conditions, but the differences in potassium ion concentration likely report the differences in ionic strength in these prokaryotes. The difference in ionic strength between *E. coli* and *L. lactis* is also reflected in the higher turgor pressure of *L. lactis* (12).

Results

GFP net charge affects its diffusion coefficient in *E. coli*

We performed fluorescence recovery after photo-bleaching (FRAP; see Figure 1a, b) to determine the diffusion coefficients of surface-modified variants of GFP in the *E. coli* cytoplasm. We determined the diffusion coefficient of the -30, -7, 0, +7, +11b, +15 and +25 variants of GFP; see Figure 1c for structural models. The numbers indicate the net charge; the "b" in +11b GFP refers to the distribution of the charge over the surface and will be discussed in more detail below.
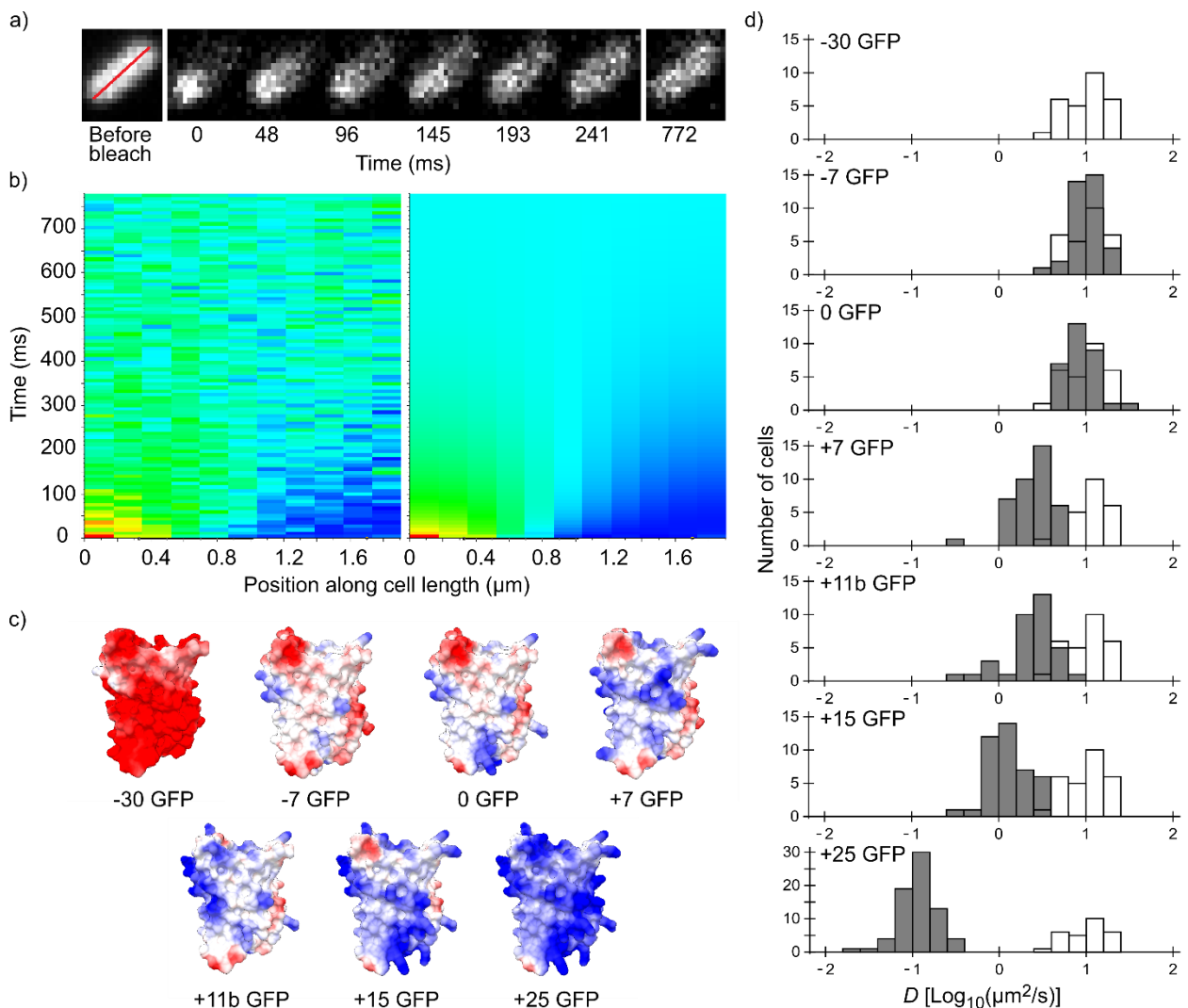
Figure 1: Illustration of the fluorescence recovery after photobleaching (FRAP) method, models of GFP variants and histograms of diffusion coefficients of surface modified variants of GFP in *Escherichia coli*. a) Data from a FRAP experiment. The zero time point is recorded immediately after the bleach. The red line marks the region along which the recovery is analyzed. b) Fluorescence intensity along the red line in time, for data (left) and the

fit to that data (right). The data is fitted with a numerical variant of the 1D diffusion equation. From the fit we obtain the diffusion coefficient. c) Structural models of the surface-modified GFP variants, based on the superfolder GFP structure (PDBID:2B3P). The colors indicate the charge. d) Histograms of diffusion coefficients of GFP variants in *E. coli* over a population of cells. For comparison, the histogram for the -30 GFP variant is shown in white in every plot.

For each variant we measured the diffusion coefficients on cells from at least three independent cultures, and for each cell we obtained a single diffusion coefficient. For each GFP variant we plotted the histogram of diffusion coefficients over the population of cells (Figure 1d). The -30, -7 and 0 variants of GFP all have the same mean diffusion coefficient of 10-11 $\mu m^2/s$ (for table of mean diffusion coefficients see Supplementary file 1B). At +7 GFP the diffusion coefficients start to drop, down to a mean value of 0.14 $\mu m^2/s$ for +25 GFP.

For the +15 and +25 GFP variants we observed heterogeneous fluorescence in some cells. This ranged from a somewhat higher fluorescence at the poles to a clear ring around the outskirts of the cells (Figure 2-figure supplement 1a). For +15 GFP we were able to get rid of the heterogeneities by inducing for a shorter period of time. For +25 GFP, we excluded cells with extreme heterogeneities (see Figure 2-figure supplement 1a). The cells with only slightly inhomogeneous fluorescence had similar diffusion coefficients to cells with homogeneous fluorescence and were included in the data (Figure 2-figure supplement 1b). The heterogeneities are probably due to exclusion of large complexes (+25 GFP forming clusters of ribosomes) from the nucleoid (see below).

Why does the diffusion coefficient drop with net positive charge? The first thing to realize is that almost nothing in the cytoplasm of the *E. coli* cell truly stands still. The membrane rearranges itself continuously and the DNA has some diffusive motion and rearranges itself during the cell cycle. If +25 GFP would stick to another average sized protein it would not move more slowly than the combination of the two can diffuse. If +25 GFP were to bind some bigger complex, like β-galactosidase (~500 kDa) with a diffusion coefficient of ~1 $\mu m^2/s$, it would diffuse with a similar rate as β-galactosidase.

Based on a census of elements present in *E. coli* cells this means that for +25 GFP to diffuse with a diffusion coefficient as low as 0.14 $\mu m^2/s$ it needs to bind to membrane proteins, DNA and/or ribosome-mRNA complexes. For membrane proteins with 12-14 transmembrane helices in *E. coli* and *L. lactis*, D is 0.02-0.03 $\mu m^2/s$ (12,16). Describing the motion of DNA with a diffusion coefficient is somewhat of a stretch, as its parts do not move freely, but apparent values of 0.000035-0.00007 $\mu m^2/s$ have been reported (26). In fast growing cells, we expect ribosomes and mRNA to be associated most of the time (27), and D is 0.04 $\mu m^2/s$ when a one component model is used for fitting the data (28). In another study free and bound ribosomes were discriminated and D values of 0.40 (~15 %) and 0.055 $\mu m^2/s$ (~85 %) were found (27). So membrane proteins, DNA and ribosomes-mRNA all have diffusion coefficients low enough to cause the drop in mobility of +25 GFP. Fluorescence images show that at most fluorescence is located in the cytoplasm, leaving DNA and/or ribosome-mRNA as the most likely (major) binding partners. We can't rule out that some of the GFP binds to the membrane.

Cytoplasmic ion concentrations counteract the drop in diffusion coefficient

We also determined diffusion coefficients of the GFP variants in *L. lactis* and *Hfx. volcanii* (Figure 2 and Supplementary file 1B). For *L. lactis* the mean diffusion coefficient of -7 GFP (6.2 $\mu m^2/s$) is lower than for *E. coli* (10 $\mu m^2/s$). However, D drops less with positive net charge so that for +25 GFP the mean diffusion coefficient is higher in *L. lactis* (0.61 $\mu m^2/s$) than in *E. coli* (0.14 $\mu m^2/s$). This can be explained by the higher cytoplasmic ionic strength of *L. lactis,* reducing the affinity of positive GFPs to

hypothetical negatively-charged binding partners. For *Hfx. volcanii* the drop in diffusion coefficient is even less steep, with the mean diffusion coefficient dropping from 5.5 $\mu m^2/s$, for -7 GFP, to 1.9 $\mu m^2/s$, for +25 GFP. The shallower drop in diffusion coefficient with net positive charge as compared to both *L. lactis* and *E. coli* can again be explained by a difference in ionic strength. Another possible contribution to the high +25 GFP diffusion coefficient in *Hfx. volcanii* is the presence of more negative proteins than in *E. coli* and *L. lactis*, which may titrate GFP away from its slower binding partner (see proteome analysis below). We also note that in *Hfx. volcanii* the diffusion coefficient of -30 GFP is higher than of -7 GFP, 10 $\mu m^2/s$ compared to 5.5 $\mu m^2/s$, which may be caused by a less negative binding partner for GFP in *Hfx. volcanii* than in *E. coli*.
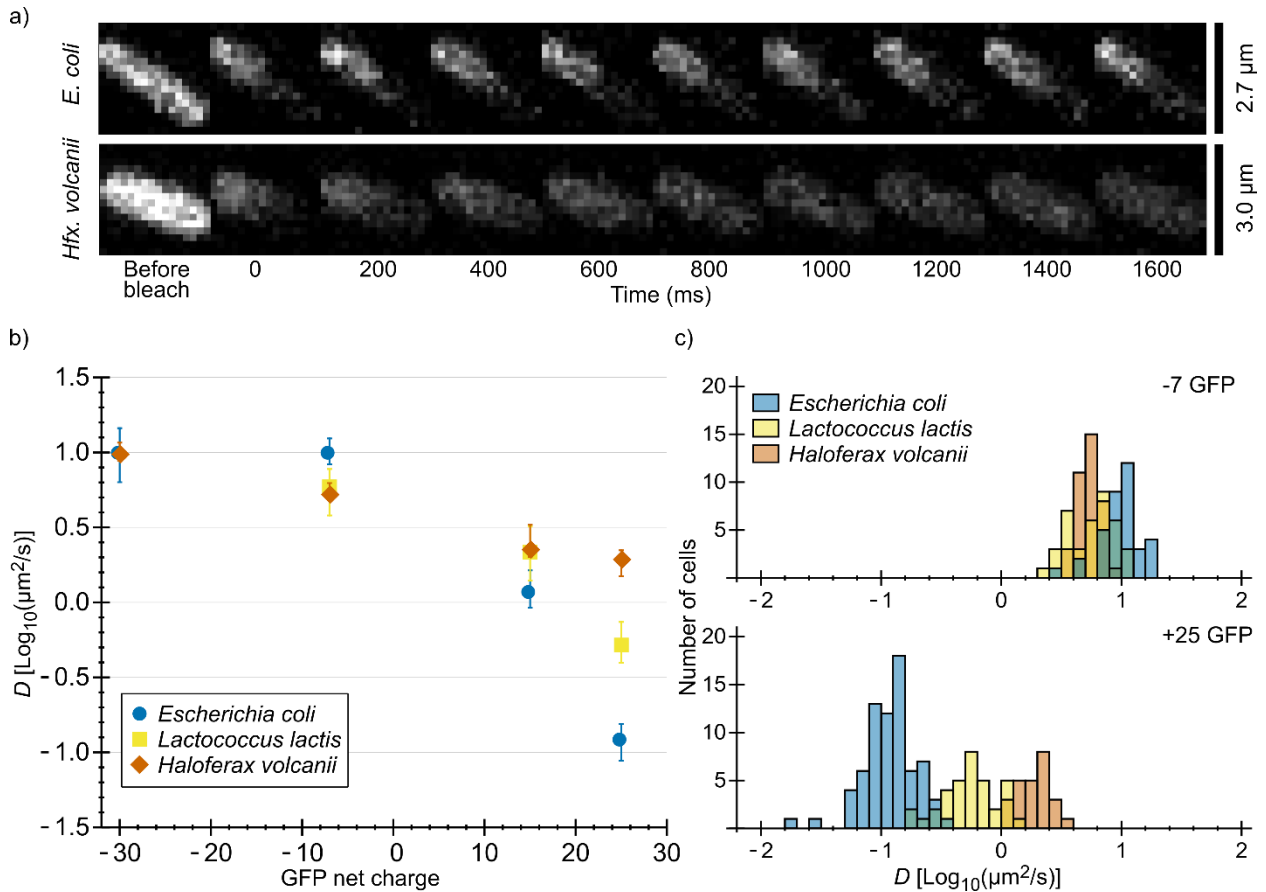
Figure 2: Comparison of diffusion coefficients of surface-modified variants of GFP in *E. coli*, *L. lactis* and *Hfx. volcanii*. a) Example FRAP data for *E. coli* and *Hfx. volcanii* cells expressing +25 GFP. We chose cells of comparable size so that the diffusion rate can be compared visually. b) The GFP diffusion coefficient plotted against its net charge in all three organisms. The points indicate medians and the error bars show the interquartile range. c) Histograms of GFP diffusion coefficients for the -7 and +25 variants in all three organisms.

Figure 2-figure supplement 1: Unequal fluorescence distribution for +25 GFP in *E. coli* and *Hfx. volcanii*. (A) Pictures of *E. coli* and *Hfx. volcanii* (in box with sol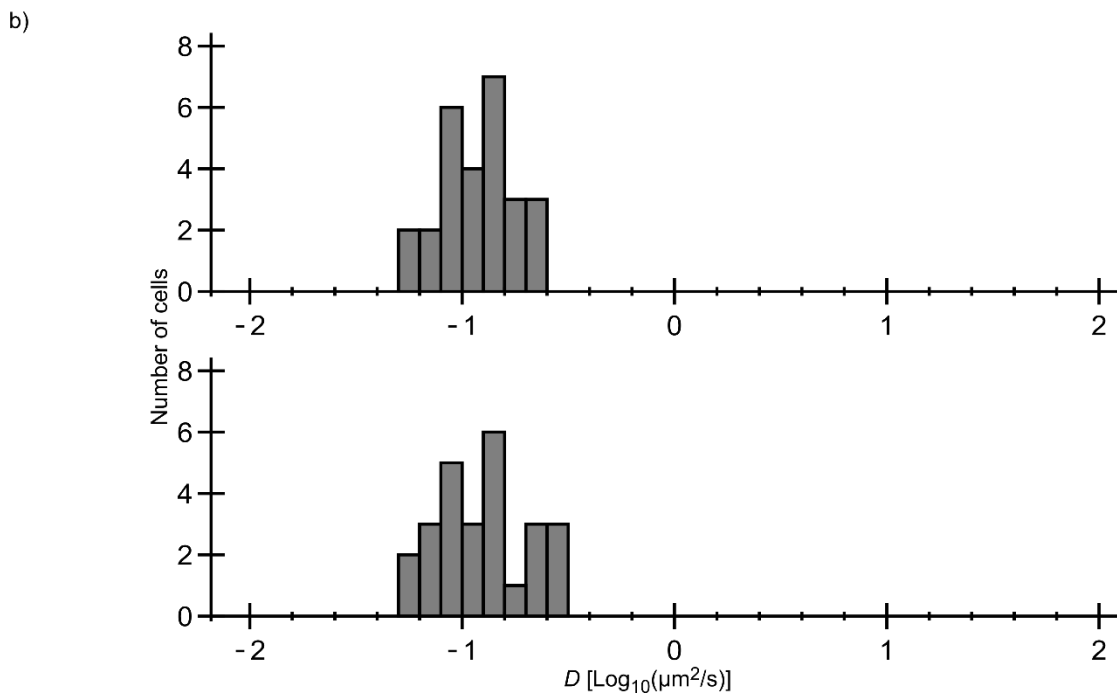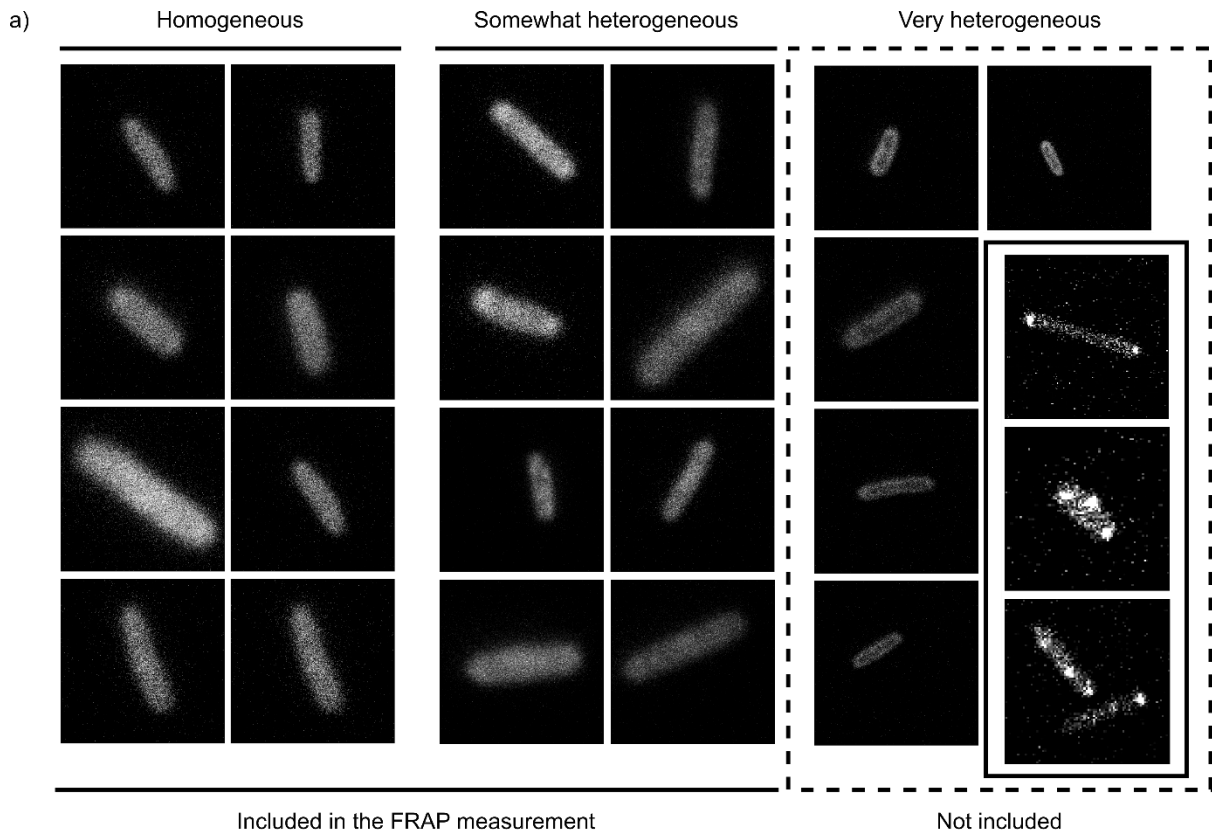id border) with different degrees of heterogeneity in the distribution of fluorescence. (B) Comparison of diffusion coefficients of +25 GFP in *E. coli* cells with a homogeneous (top) and a somewhat heterogeneous (bottom) distribution of fluorescence.

The effect of osmotic upshift on protein diffusion in *E. coli*

Next, we determined the diffusion coefficient of -30, -7, +15 and +25 GFP in *E. coli* after resuspending the cells in medium with a higher osmolality. It is known that the GFP diffusion coefficient drops drastically after an osmotic upshift (11,12,18). We now observe what happens after combining two causes for slowed diffusion: osmotic upshift (increased crowding) and protein surface charge. The cells were grown at an osmolality of 0.28 Osm and resuspended in media of 0.55 or 1.2 Osm (adjusted with NaCl). In 0.55 Osm medium the diffusion coefficients did not change much (Figure 3a and Supplementary file 1B), similar to what was observed before for wildtype GFP (11). In 1.2 Osm medium the diffusion coefficients of all variants dropped (Figure 3a and Supplementary file 1B). The degree of the drop is 56-, 28-, 16-, and 7-fold (between medians) for -30, -7, +15 and +25, respectively, and this difference may be a consequence of the increased cytoplasmic ion concentration that accompanies the osmotic upshift. Thus, the fold-change is less for the positive proteins because the electrostatic screening may compensate partly for the increased crowding effect.
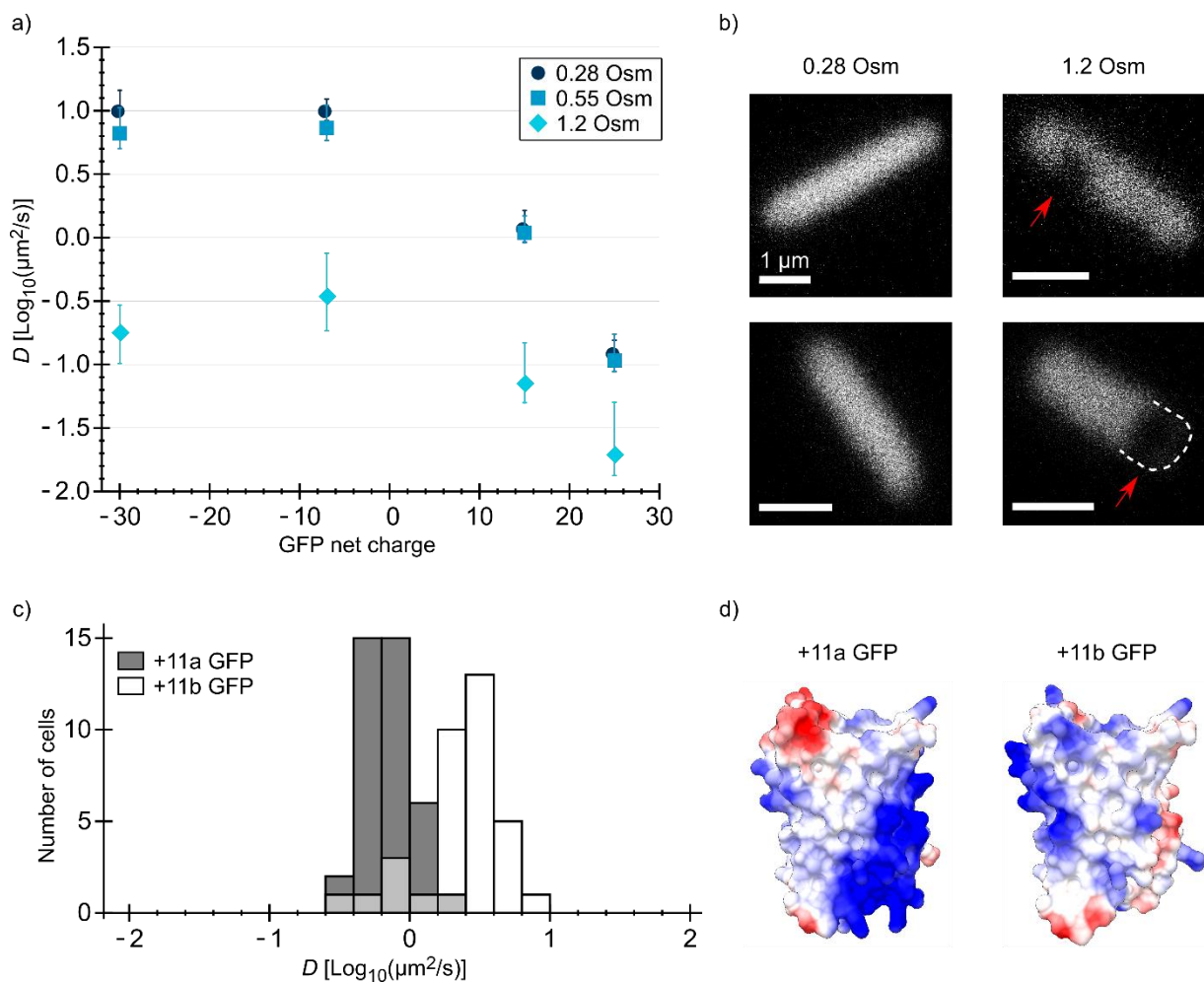
Figure 3: Diffusion coefficients of surface-modified variants of GFP at different osmotic stress and charge distribution effects. a) GFP diffusion coefficients as a function of their net charge and degree of osmotic stress in *E. coli*. The points indicate medians and the error bars show the interquartile range. Discs: *E. coli* cells resuspended in medium with the same osmolality as the growth medium (0.28 Osm); data from Figure 2b. Squares: cells resuspended in 0.55 Osm; Diamonds: cells resuspended in 1.2 Osm. b) Microscopy images of -7 GFP fluorescence of cells resuspended in 0.28 Osm (left panel) and 1.2 Osm medium osmolality (right panel).

Red arrows indicate invaginations which appear after rapid osmotic upshift. Scale bars are 1 μm. c) Histogram of diffusion coefficients for +11a (grey bars) and +11b GFP (white bars) in *E. coli*, measured at growth osmolality (0.28 Osm). The FRAP data only includes cells with homogeneous fluorescence. d) Structural models of +11a and +11b GFP variants. The colors indicate the charge distribution; the same protein faces are shown.

The distribution of surface charge affects the diffusion coefficient in *E. coli*

For -30, -7, 0, +7, +11b, +15 and +25 GFP, the charge is distributed more or less equally over the surface of the protein. We also determined how a more localized charge affects the diffusion. The 11a variant of GFP has the positive charge unequally distributed, compared to +11b GFP (Figure 3d). We determined diffusion coefficients for the 11a and 11b variants by FRAP. Histograms of the diffusion coefficients over populations of cells are shown in Figure 3c. The mean diffusion coefficient of +11b GFP is 2.7 $μm^2/s$ and that of +11a GFP is 0.76 $μm^2/s$ (Supplementary file 1B). So, it clearly matters how the net charge is distributed over the GFP surface. For +11a GFP we also see heterogeneous distributions, similar to +15 and +25 GFP, which could be prevented by inducing for a shorter amount of time.

The +25 GFP variant does not co-localize with DNA

To find out whether the positive GFP variants are bound to DNA or ribosomes-mRNA or both, we first determined the co-localization between GFP and DNA. We expressed +25 GFP in *E. coli*, labelled the nucleoid with DRAQ-5 and shrunk the nucleoid with chloramphenicol. We compared the fluorescence profile, along the length of the cells, of +25 GFP with that of DRAQ-5 (Figure 4a and Figure 4-figure supplement 1). In all cells the distribution of +25 GFP matched the dimensions of the cells. In nine cells (Figure 4-figure supplement 1 a-i) out of 46 the nucleoid had shrunk and in all these cells +25 GFP did not co-localize with DNA. In some cells +25 GFP was occluded from the DNA, which has been seen before for ribosomes (28). In the other cells the nucleoid did not shrink and the DNA and +25 GFP overlapped (see Figure 4-figure supplement 1 j-p). We conclude that DNA is not the major binding partner for +25 GFP.

DNA is not needed for the decrease of +25 GFP diffusion rate

Next, we determined if DNA affects the mobility of +25 GFP by analyzing the diffusion of -7 GFP and +25 GFP in DNA-free regions of *E. coli* cells. We created DNA free regions large enough for FRAP measurements by first treating cells with cephalexin, to elongate the cells, and then with chloramphenicol, to shrink the nucleoid. We visualize the position of DNA by adding DRAQ-5. Only a fraction of the cells had enough GFP fluorescence for FRAP and sufficient DRAQ-5 fluorescence for visualizing the position of DNA, and these were analyzed (Figure 4-figure supplement 2a and b). We find that for both -7 and +25 GFP the diffusion coefficient has dropped after treatment of the cells with cephalexin and chloramphenicol (Figure 4-figure supplement 2c and d). There is a big difference between the mean diffusion coefficient of -7 and +25 GFP after treatment and in the absence of DNA (35-fold). A similar difference between -7 and +25 was found in cells that were not treated and in the presence of DNA (85-fold). This shows that the drop in +25 GFP diffusion rate is not dependent on DNA. Both -7 GFP and +25 GFP are somewhat excluded from DNA (nucleoid) but the effect is largest for +25 GFP; this is another piece of evidence suggesting that +25 GFP binds to ribosome-mRNA and not to DNA.

+25 GFP binds ribosome-mRNA

It has been shown that the 30S ribosomal subunit in *E. coli* increases its diffusion coefficient, from 0.04 μm^2/s to 0.6 μm^2/s, after treatment with rifampicin (28). Rifampicin stops transcription and after adding it to *E. coli* cells the pool of mRNA plummets, with 90 % of the mRNAs having a half time of less than 8 min (29). We determined the diffusion coefficient of +25 GFP in *E. coli* as a function of time after the addition of rifampicin and compared this to the situation without rifampicin (Figure 4b).
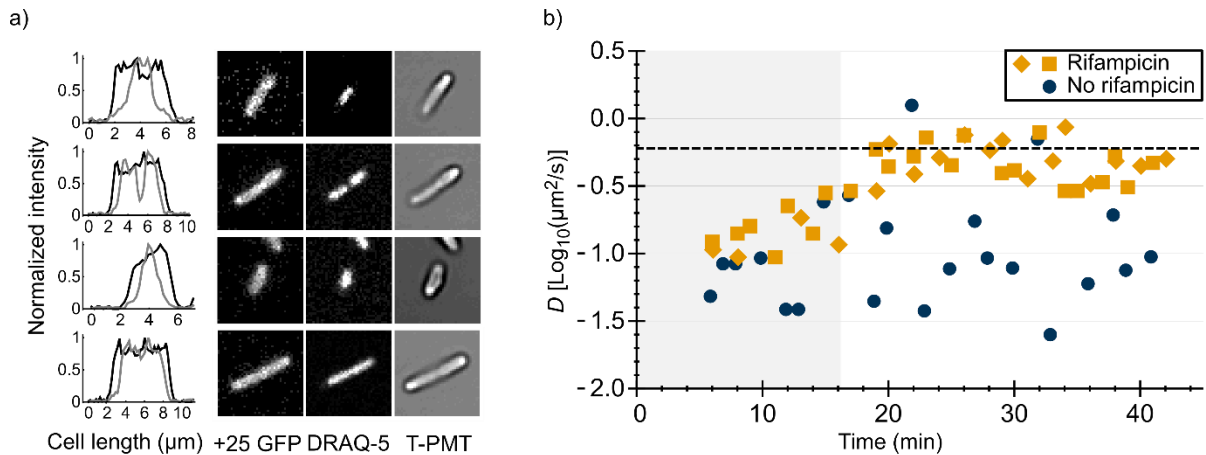
Figure 4: Comparison of distributions of +25 GFP and DNA in *E. coli* and diffusion of +25 GFP in the presence and absence of mRNA. a) Co-localization of +25 GFP and DNA in *E. coli*. The plots indicate the fluorescence profile for +25 GFP (black) and DNA (grey) along the length of the cell, averaged over a 5 pixel wide band. The images show the corresponding cells in the +25 GFP, DRAQ-5 (DNA) and T-PMT channels. The T-PMT image corresponds to the transmitted excitation light during the recording of the fluorescence (it is equivalent to a bright-field image). b) Diffusion of +25 GFP in *E. coli* in the presence and absence of mRNA. At time point zero, DMSO + rifampicin (yellow) or DMSO only (blue) was added to the cells. The squares and diamonds indicate different replicates. The dashed line indicates the diffusion coefficient of the 30S ribosome after the addition of rifampicin. At the transition from the shaded region to the white region, > 75% of the mRNA is gone.

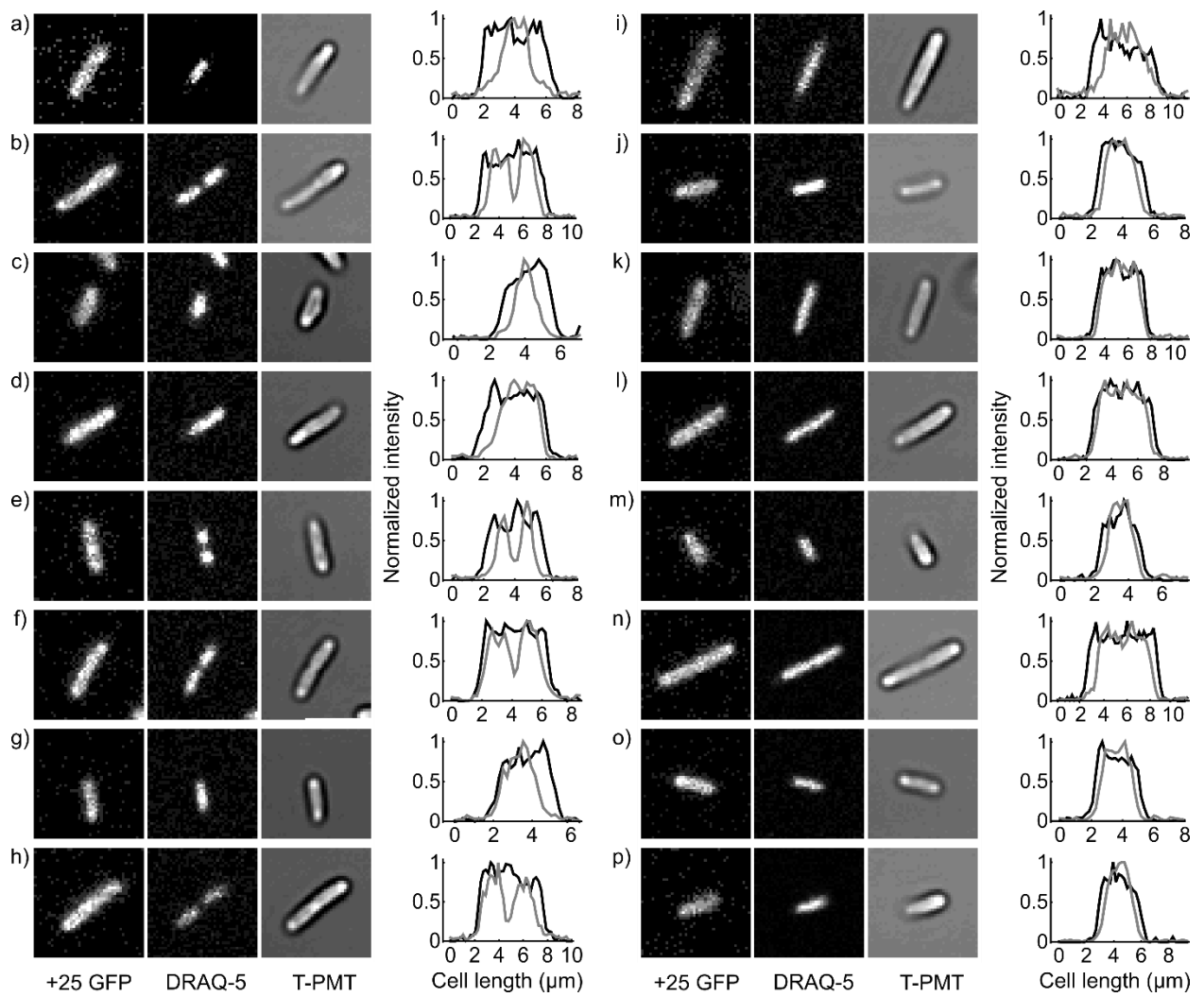**Figure 4-figure supplement 1: Comparison of distributions of +25 GFP and DNA in *E. coli*.** DRAQ-5 reports on the presence of DNA. The T-PMT image is constructed from the transmitted excitation light during the recording of the fluorescence (it is equivalent to a bright-field image). The plots indicate the fluorescence profile for +25 GFP (black) and DNA (grey) along the length of the cell, averaged over a 5 pixel wide band.
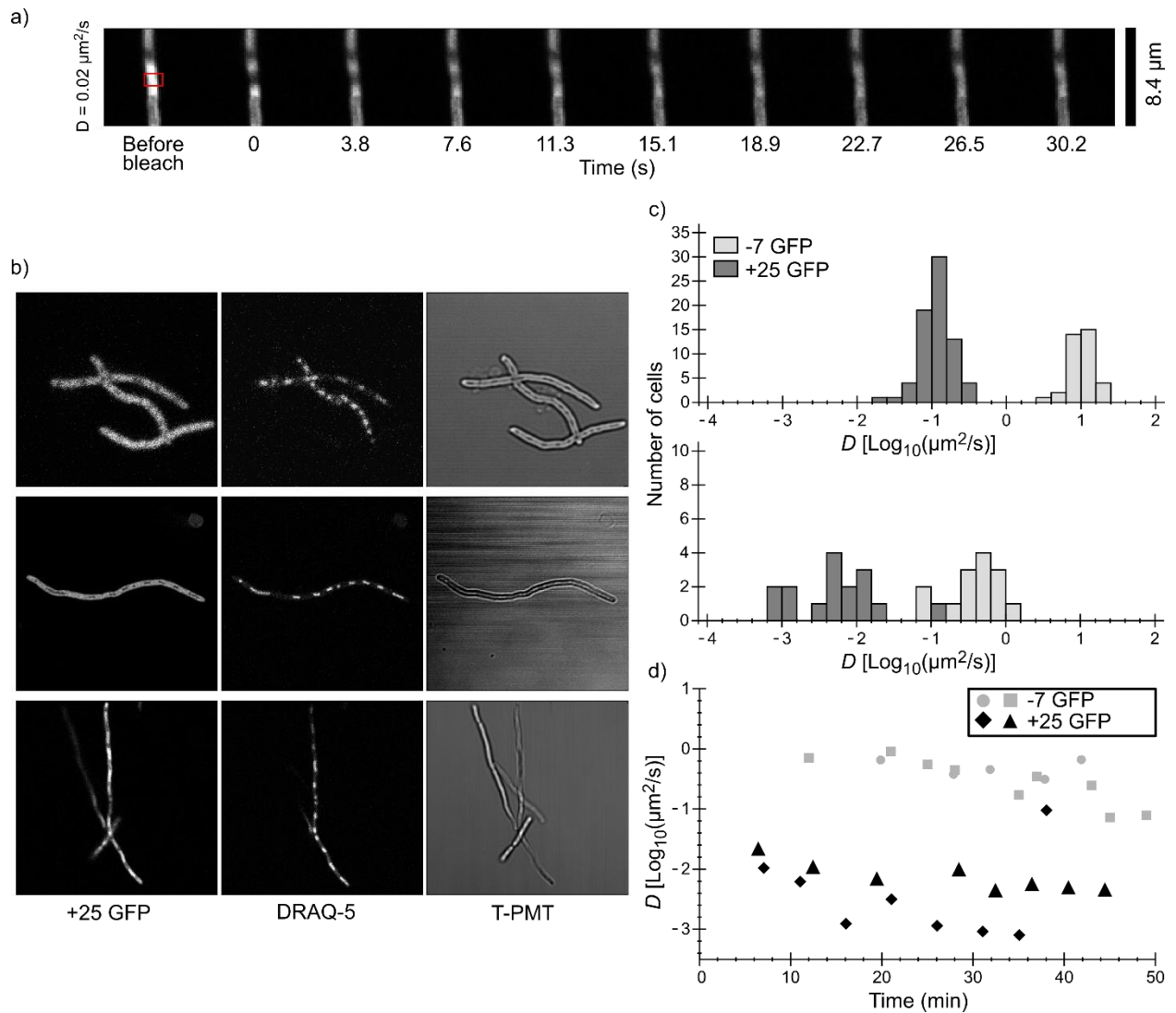
Figure 4-figure supplement 2: Diffusion coefficients of -7 GFP and +25 GFP in DNA-containing and DNA-free regions. To create DNA free regions, *E. coli* cells were elongated by adding cephalexin, and DNA was condensed by treating the cells with chloramphenicol. a) Example of a FRAP experiment on treated cells. The line along which the recovery was analyzed was drawn over the highly fluorescent middle part of the cell. b) Examples of *E. coli* cells after treatment with cephalexin, chloramphenicol and DRAQ-5. DRAQ-5 visualized DNA and the T-PMT is equivalent to a bright field image. c) Comparison of histograms of -7 GFP and +25 GFP diffusion coefficients in DNA-containing (same data as in Figure 1d; top panel) and DNA-free regions (bottom panel). d) Diffusion coefficients of -7 GFP and +25 GFP in treated cells as a function of time. The zero time point indicates when the cells are resuspended in DRAQ-5 free medium right before the FRAP measurements. Each point corresponds to a single cell. Discs, squares, diamonds and triangles indicate different replicates.

We found that after treatment with rifampicin, the diffusion coefficient of +25 GFP increases for 20 min and then levels off. This coincides with the time needed to degrade most of the mRNA. Importantly, the diffusion coefficient after 20 min of rifampicin treatment is close to the value of the 30S and presumably the 50S subunit. We also find that the fluorescence of +25 GFP expressing cells becomes more homogenous after rifampicin treatment. Together, these findings indicate that the

positive GFP variants bind to ribosomes, and that this is the major cause for their slow diffusion. With reasonable confidence we also put aside two other hypotheses: (i) differential partitioning of negative and positive GFPs in different cytoplasmic phases; and (ii) formation of big clusters of positive GFPs with negative proteins. Finally, we find that the variation in diffusion coefficient between cells is smaller in the presence than in the absence of rifampicin, suggesting that part of the spread in the diffusion relates to ongoing transcription.

Co-localization on sucrose gradients shows that +25 GFP binds predominantly to ribosomes

To substantiate our *in vivo* findings on the binding of positive GFPs to ribosomes, we determined whether -7 and +25 GFP co-localize with ribosomes and/or DNA on a sucrose gradient. For this experiment we used (ribosome containing) lysates of -7 or +25 GFP expressing *E. coli* cells. The cell lysates were 200-300 times diluted relative to the cytoplasmic contents. The contents of the lysates were separated by centrifugation on a linear sucrose gradient and we determined the presence of the GFPs, by fluorescence spectroscopy, in fractions taken along the length of the gradient. We also determined the presence of ribosomes by electron microscopy. We observed a clear difference in the position of -7 and +25 GFP along the gradient, with the peak of the +25 GFP distribution coinciding with the presence of ribosomes (Figure 5a and c). We do not know the exact DNA content of the lysates, so we performed two more experiments in which we added pure DNA to the +25 GFP cell lysates before separating their contents on sucrose gradients. We used a DNA/ribosome ratio that is comparable (0.02 g/L DNA, experiment 1), or five times higher to that in the cell (0.12 g/L DNA, experiment 2) (30). The position of DNA along the gradient was determined in a separate experiment in which only the pure DNA was added to a sucrose gradient. Even at the highest concentration of DNA, +25 GFP co-localizes with ribosomes and not with DNA (Figure 5b). From the combination of sucrose gradient experiments and our *in vivo* studies described above, we conclude that +25 GFP binds predominantly to ribosomes.
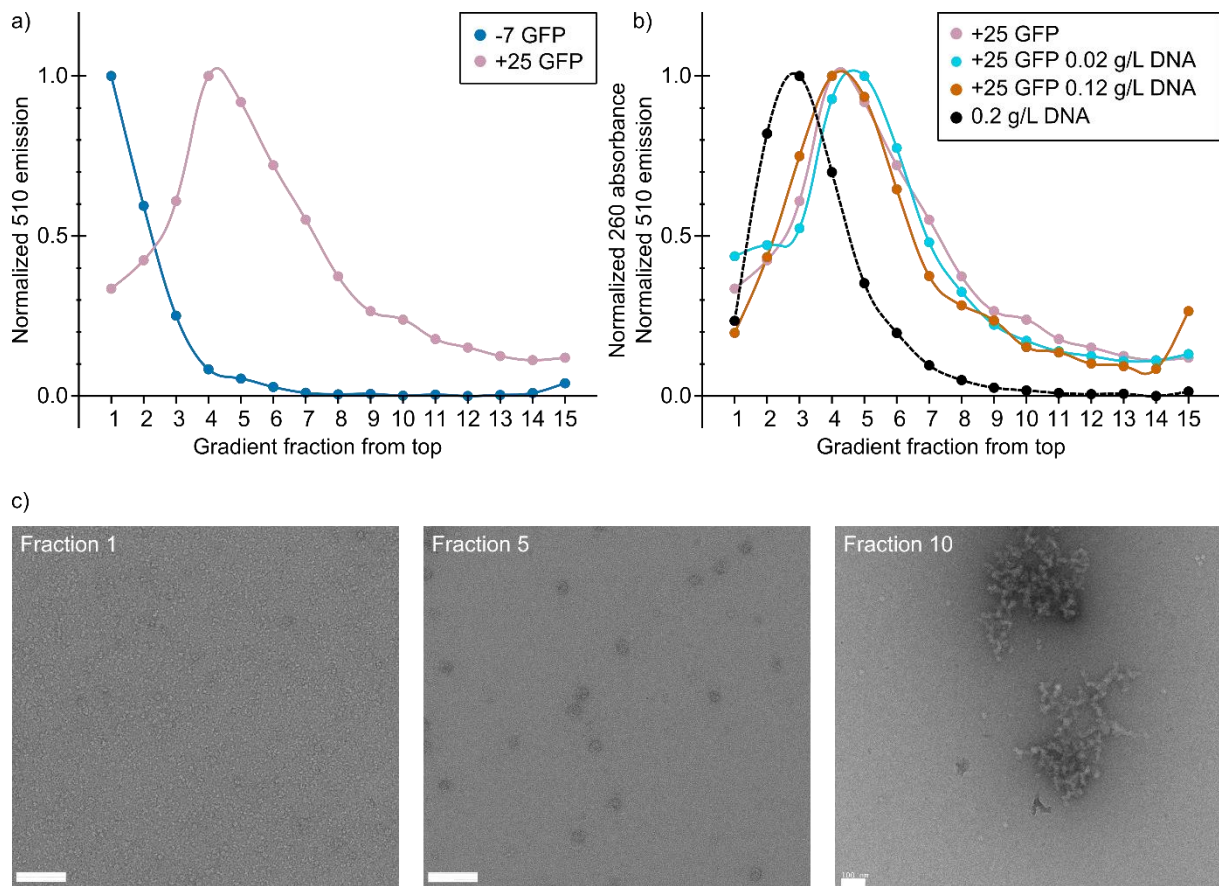
Figure 5: Co-localization of +25 GFP with ribosomes on sucrose gradients. a) Comparison of fluorescence profiles of sucrose gradient centrifugation experiments performed on *E. coli* cell lysates containing either -7 GFP or +25 GFP. The majority of -7 GFP is present in the loaded sample (fractions 1 and 2), while +25 GFP peaks at fractions 4-5, corresponding to 15-18 % (w/v) sucrose. The fluorescence signals were normalized to the highest value, because the absolute fluorescence of +25 GFP is lower than that of -7 GFP. b) Comparison of fluorescence (and absorption) profiles of sucrose gradient centrifugation experiments performed on purified DNA (0.2 g/L), and *E. coli* cell lysates containing +25 GFP with or without additional DNA. c) Transmission electron microscopy images of uranyl acetate-stained fractions from the cell lysate containing sucrose gradients. Fraction 1 lacks distinct large structures, whereas fraction 10 shows large aggregates. Ribosomes, spheres of around 25-30 nm diameter, are visible and peak in fraction 5. The scale bar is 100 nm.

Discussion

The fraction of GFP variants bound to ribosomes in *E. coli*, *L. lactis* and *Hfx. volcanii*

The diffusion coefficient of GFP in cells is a function of free and bound GFP. If the exchange between these two states is longer than the time of the FRAP measurement one will observe two populations. If the two states exchange on a timescale much shorter than the time of the FRAP measurement the diffusion can be described by a single diffusion coefficient (D_{eff}):

$$D_{eff} = f_{free} \, D_{free} + \left(1 - f_{free}\right) D_{bound} \qquad (1)$$

We used eq. 1 to calculate the fraction of free GFP for each of the GFP variants in *E. coli*, *L. lactis* and *Hfx. volcanii* (see methods for derivation of eq. 1). We made a number of assumptions: (1) the exchange between free and bound state is much faster than the FRAP measurement; (2) the highest diffusion coefficient of all variants in a given organism reflects the free state of GFP; (3) GFPs bind

solely to ribosomes; (4) the total number of binding sites on all ribosomes is higher than the number of GFPs; (5) the decrease of diffusion coefficient with net positive charge has the same origin in all three organisms; and, finally (6) the ribosome diffusion coefficient is the same in all three organisms. The justification for these assumptions, as far as they exist, are as follows. (1) We did not observe two-component recovery in the FRAP data; (2) for *E. coli* the GFP diffusion coefficient levels off towards more negative charge; (3) evidence from the rest of the paper makes ribosomes the most likely (substantial) binding partner; (4) the number of binding sites is about 10^6 (see below) it is unlikely that we express that many GFPs because the cell only has about 3×10^6 total proteins; (5) all three organisms have ribosomes, and assuming the same cause for slow diffusion is the most parsimonious; (6) for all three organisms -7 GFP diffusion coefficients are similar suggesting similar crowding etc., for *E. coli* and *L. lactis* the diffusion coefficient are similar also for a big protein complex (ref. 12). The consequence for violating assumption 1 is that there is no D_{eff} to speak of and the whole calculation becomes moot. The consequence for violating assumptions 2, 3, 5, and 6 is that the numerical values coming out of the equation will be different, with the severity of the error depending on the difference in diffusion rates with the actual free diffusion and binding partners. The consequence of violating assumption 4 is that there will be free GFP irrespective of affinity as there are no more binding sites to fill. The results are shown in Figure 6-figure supplement 1. From the analysis we conclude that even in *Hfx. volcanii*, with its high internal ion concentration, a major fraction (0.81) of +25 GFP is still bound to ribosomes.

Next, we estimated the dissociation constant (K_d) of the association between GFP and ribosomes. For *E. coli*, under our growth conditions, the number of ribosomes per μm^3 is about 17000 (31), which corresponds to a concentration of 10 µM. GFP probably binds to the RNA that is exposed on the surface of the ribosome and probably does so nonspecifically. The ribosome has a diameter of 20 nm, which yields a surface area of 1260 nm^2, assuming a spherical shape. About half of this surface area is RNA so we end up with 630 nm^2. The diameter of GFP is 3.5 nm, giving a 9.6 nm^2 cross section. Dividing the ribosome RNA surface area by the GFP cross section gives a maximal number of binding sites of 66. This means that the concentration of binding sites is 660 µM. To calculate K_d we use:

$$f_{bound} = \frac{[binding\ site]}{K_d + [binding\ site]} \qquad (2)$$

This equation is valid when the number of binding sites is significantly higher than the number of GFPs. Using a fraction of bound +25 GFP of 0.99 in *E. coli,* we obtain K_d = 6.7 µM. If we make the assumption that the concentration of ribosomes in *L. lactis* and *Hfx. volcanii* is the same as in *E. coli*, then the K_d for binding of +25 GFP to a ribosome binding site is 65 µM for *L. lactis* and 155 µM for *Hfx. volcanii*.

The relation between diffusion coefficient, GFP net charge and ionic strength

In this section we seek to explain: (i) the relation between D_{eff} and cytoplasmic ionic strength; and (ii) the relation between D_{eff} and net charge of GFP. To explain (i) we first compare our values to literature data on electrostatic interactions between proteins in dilute solution. To make the comparison possible we relate K_d, rather than D_{eff}, to ionic strength. In Figure 6a we plot the K_d *versus* ionic strength of the interaction between barnase and barstar (32), colicinE9 and Im9 (33), and different forms of thrombin and hirudin (34).
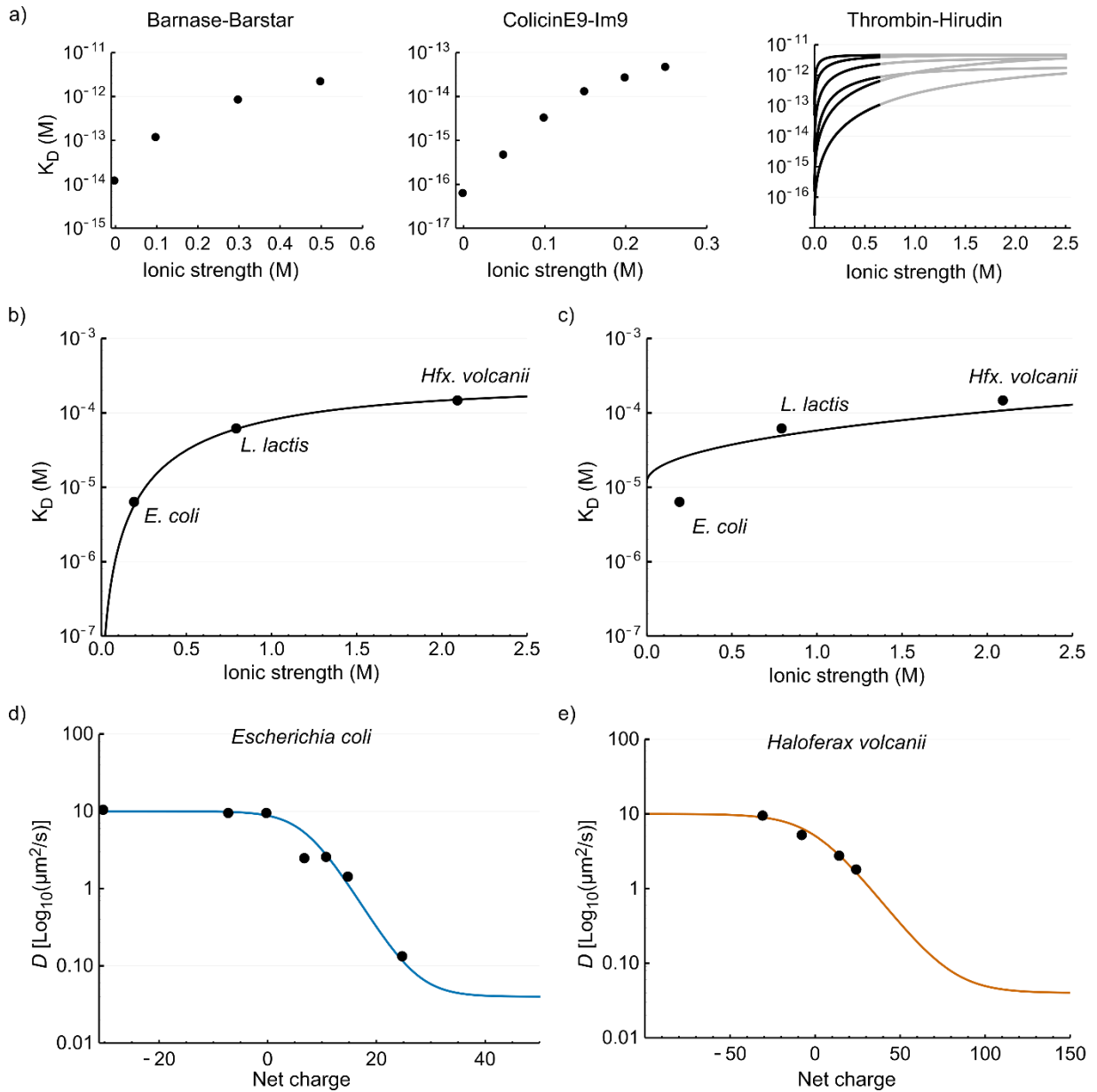
Figure 6: The relation between the diffusion coefficient, net charge and ionic strength. a) The dependence of dissociation coefficients on ionic strength for protein binding pairs barnase-barstar, colicinE9-Im9, and different variants of hirudin binding to thrombin. The data from the literature was in the form of k_{on} and k_{off} and we used $K_D = \frac{k_{off}}{k_{on}}$ to determine the affinity constants. For the thrombin-hirudin interactions we show fits with a combination of equations (3) and (4). The black part spans the data, the grey part is an extrapolation. The charge on hirudin decreases from the bottom to the top line (at the black part). b) The dependence of dissociation coefficients on ionic strength for +25 GFP. The black line is a fit with a combination of equations (3) and (4). c) Same data as in b) but with the non-ionic contribution to the binding free energy fixed at zero during fitting. d) The dependence of diffusion coefficient on GFP net charge for *E. coli*. The line is a fit with equation 5. We did not include +11a GFP because of its different charge distribution. e) The dependence of diffusion coefficient on GFP net charge for *Hfx. volcanii*. The line is a fit with equation 5.
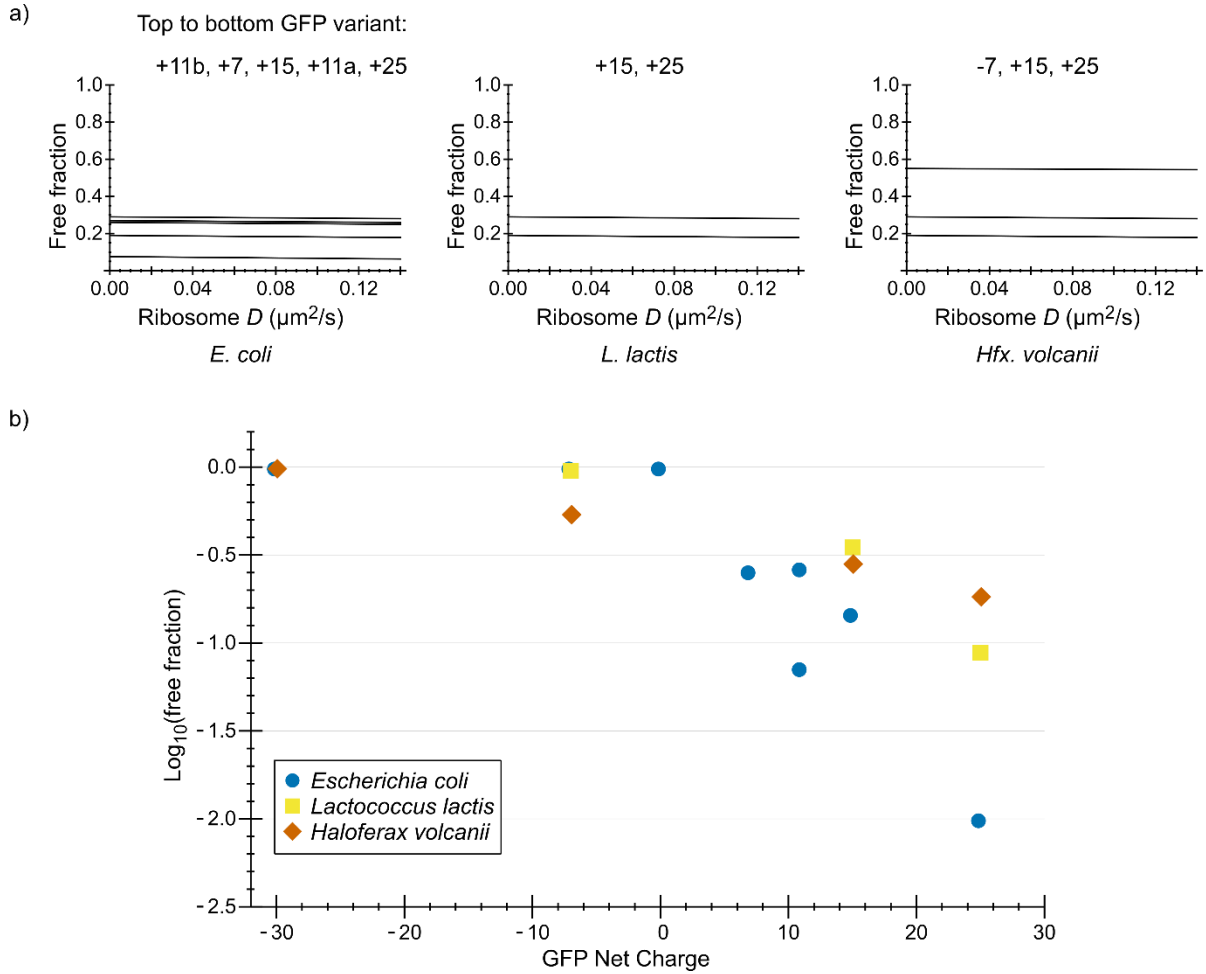
Figure 6-figure supplement 1: Fraction of free GFP variants in *E. coli*, *L. lactis* and *Hfx. volcanii*. Using the data from Figure 2b we calculated the fraction of GFP bound to ribosomes. a) The fraction of free GFP variants as a function of ribosome diffusion coefficient. b) The fraction of free GFP variants as a function of GFP net charge.

The ionic strength dependence of the interaction of +25 GFP with ribosomes (recorded in *E. coli*, *L. lactis* and *Hfx. volcanii*; Figure 6b) is similar to that of the three protein pairs in dilute solutions (Figure 6a); the K_d increases with ionic strength and levels off at higher ionic strength. To get more insight into the interaction we applied a semi-empirical equation that was successfully used to describe the (partly) electrostatic interaction between the proteins thrombin and hirudin (34):

$$\Delta G_b^\circ = \Delta G_{nio}^\circ + \Delta G_{io0}^\circ \frac{e^{-C_1\sqrt{I}}}{1 + C_1\sqrt{I}} \qquad (3)$$

This equation was derived from Debye-Hückel theory and was subsequently modified to account for behavior at higher ionic strength. Here ΔG_b° is the total binding energy, ΔG_{nio}° is the binding energy due to non-ionic interactions, ΔG_{io0}° is the binding energy due to electrostatic interactions in the absence of ions, C_1 is a parameter that depends on the distance between charges and the screening effects that are not due to ions, and I is the ionic strength. To apply equation 3 we need to relate the K_D and ΔG_b°:

$$K_D = e^{\frac{\Delta G_b^\circ}{RT}} \qquad (4)$$

We fitted the +25 GFP interaction data with a combination of equations (3) and (4) and obtained the following parameter values: ΔG_{nio}° = -20 400 J mol^{-1}, ΔG_{io0}° = -28 900 J mol^{-1}, and C_1 = 1.53. We had expected that the non-ionic interaction free energy would be close to zero, but if we impose this conditions the fit becomes bad (Figure 6c). The bending off at higher ionic strengths depends on ΔG_{io0}° being negative; we expect ΔG_{io0}° to be negative given the electrostatic attraction between positive GFP and negatively-charged surfaces of the ribosomes.

The second phenomenon we explain is the relation between D_{eff} and the surface charge of GFP. In the work describing the hirudin-thrombin interaction a number of charge variants of hirudin were used (34). The ΔG_{io0}° depended linearly on the number of charges. This is expected from Coulombs law, assuming that non-linearities do not arise from charge screening. In the rest of the analysis we assume that ΔG_{io0}° indeed depends linearly on the charge and we write: $\Delta G_{io0}^{\circ} = \Delta G_{pc}^{\circ} \times charge$, where ΔG_{pc}° is the free energy change per charge. We can now combine this with equations (1), (2), (3), and (4) to obtain:

$$D_{eff} = \left(1 - \frac{[binding\ site]}{[binding\ site] + e^{\frac{\Delta G_{nio}^{\circ} + \Delta G_{pc}^{\circ}\ charge\frac{\exp(-C_1\sqrt{I})}{1+C_1\sqrt{I}}}{RT}}} \right) (D_{free} - D_{bound}) + D_{bound} \quad (5)$$

We used this equation to fit the data for *E. coli* and *Hfx. volcanii*. We set D_{free} = 10 μm^2/s, D_{bound} = 0.04 μm^2/s (ribosome diffusion coefficient), $[binding\ site]$ = 660 μM, T = 293 K, R = 8.314 J K^{-1} mol^{-1} (gas constant), and I = 0.2 M for *E. coli* and I = 2.1 M for *Hfx. volcanii*. We are left with three fitting parameters: ΔG_{nio}°, ΔG_{pc}°, and C_1. The fits are shown in Figure 6d and e (see Supplementary file 1C for fitting parameters). The model fits the data well; this is more telling for *E. coli* than it is for *Hfx. volcanii*, as the data covers more of the curve. The upper bound for the diffusion coefficient is set by free diffusion and the lower bound by the diffusion of the ribosome. When we use a log scale to represent the diffusion coefficient we see a linear dependence of diffusion coefficient with net charge from 0 to +25 (for *E. coli*). Thus, under the assumption that ΔG_{io0}° depends linearly on the number of charges the model reproduces the linear dependence between the upper and lower bound. Again, we needed to include a non-ionic interaction for a proper fit. There is a discrepancy between the parameter values of the fits shown in Figure 6b, d and e (see Supplementary file 1C). This may be caused by the assumptions made above not holding up. Together the results for the relation between K$_d$ and ionic strength, and D_{eff} and GFP charge, show that the binding of GFP to ribosomes can be described by electrostatic interactions and screening by small ions on top of a base non-ionic interaction component.

Proteome analysis reveals potentially slow proteins

To determine the consequences of our findings we analyzed the proteomes of *E. coli, L. lactis* and *Hfx. volcanii* and four (endo)symbiotic bacteria. We determine (i) how a protein will diffuse in light of the composition of the proteome, for which we need to know both the net charges and the abundances of all proteome components; and (ii) how the ribosomes affect the diffusion of the proteome constituents, for which we need to know the net charge of the proteins. We determined the distributions of pI values and net charges for all proteins in the genome and for only cytoplasmic proteins, in *E. coli*, *L. lactis* and *Hfx. volcanii* (Figure 7a). We also determined the distribution of pI values and net charges for cytoplasmic proteins in *E. coli* taking into account protein copy numbers (Figure 7-figure supplement 1).
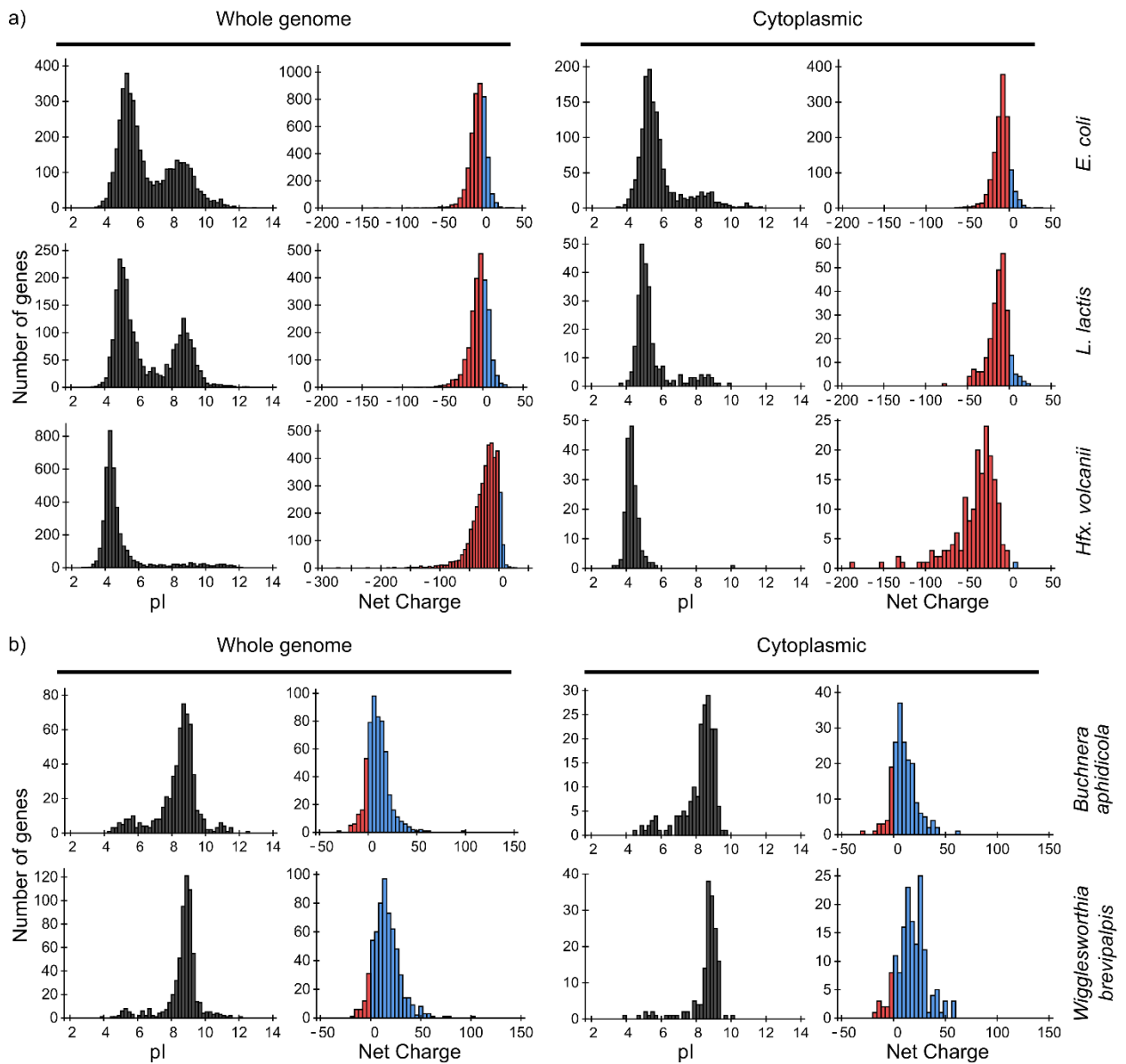
Figure 7: pI and net charge distributions for proteins of *E. coli*, *L. lactis*, *Hfx. volcanii, Buchnera aphidicola* and *Wiggleworthia glossinidia brevipalpis*. The histograms show the number of genes that encode proteins with given pI and net charge. We show distributions over all genes (left panels) and over genes that encode cytoplasmically localized proteins (right panels). a) *E. coli*, *L. lactis*, and *Hfx. volcanii*; b) the two symbionts, *Buchnera aphidicola* and *Wiggleworthia glossinidia brevipalpis*, that have the most positive proteomes (from the four symbionts that we analysed). We used gene ontology annotations from the UniProt database to find the cytoplasmic proteins. In all cases we assumed a pH of 7.5 for calculating the net charge.
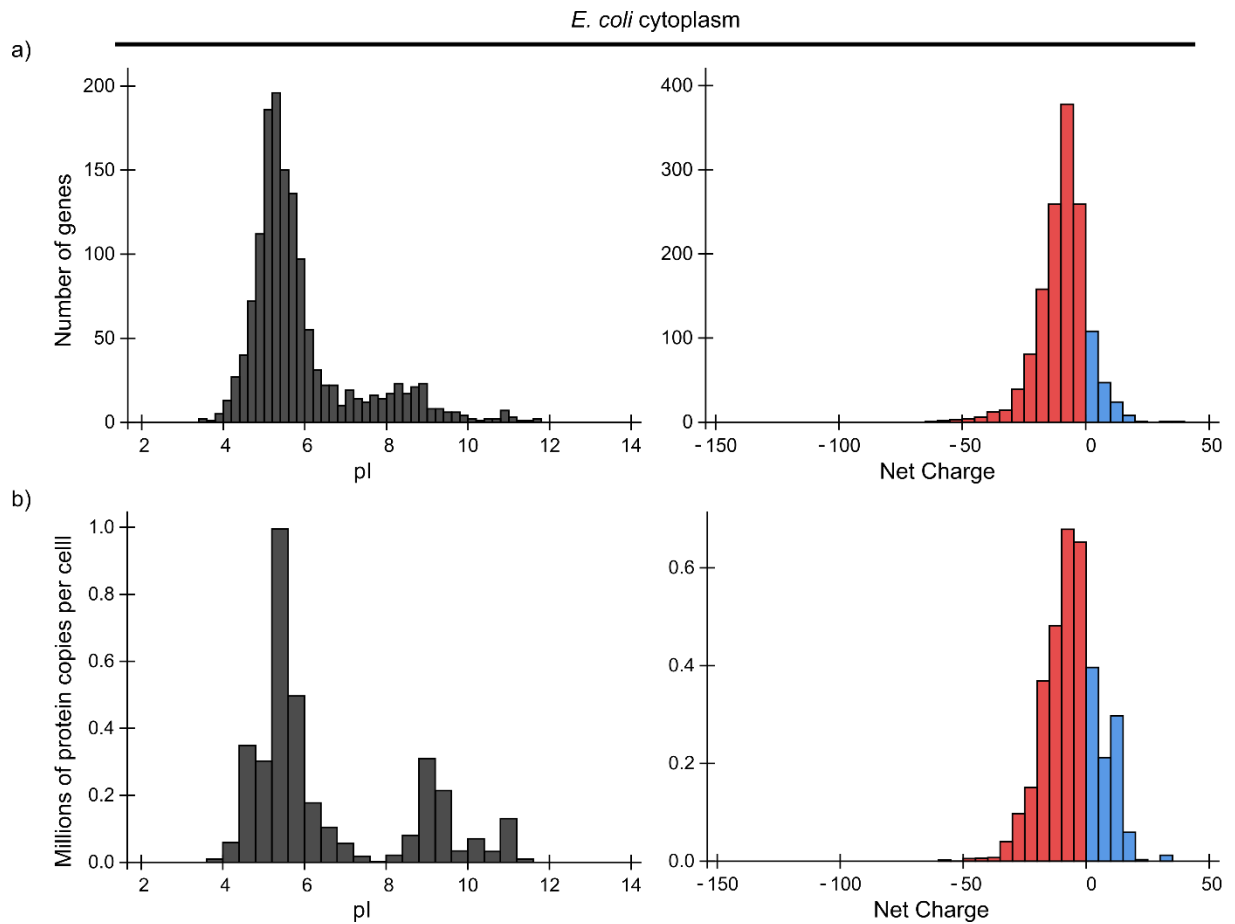
Figure 7-figure supplement 1: pI and net charge distributions for the *E. coli* proteome, taking into account protein abundance. a) Histograms of the number of genes that encode proteins with given pI (left panel) or net charge (right panel) that are located in the cytoplasm of *E. coli*. b) Histograms of the number of protein copies with given pI (left panel) or net charge (right panel) that are located in the cytoplasm of a single *E. coli* cell. We used a pH of 7.5 to calculate the net charge.

Figure 7-figure supplement 2: Protein pI and net charge distributions for *Buchnera aphidicola*, *Blochmannia floridanus*, *Onion yellows phytoplasma*, and *Wigglesworthia glossinidia brevipalpis*. Histograms of number of genes that encode for proteins with given pI and net charge. Left panel: the whole proteome; right panel: cytoplasmic proteome. In all cases we assumed a pH of 7.5 for calculating the net charge.

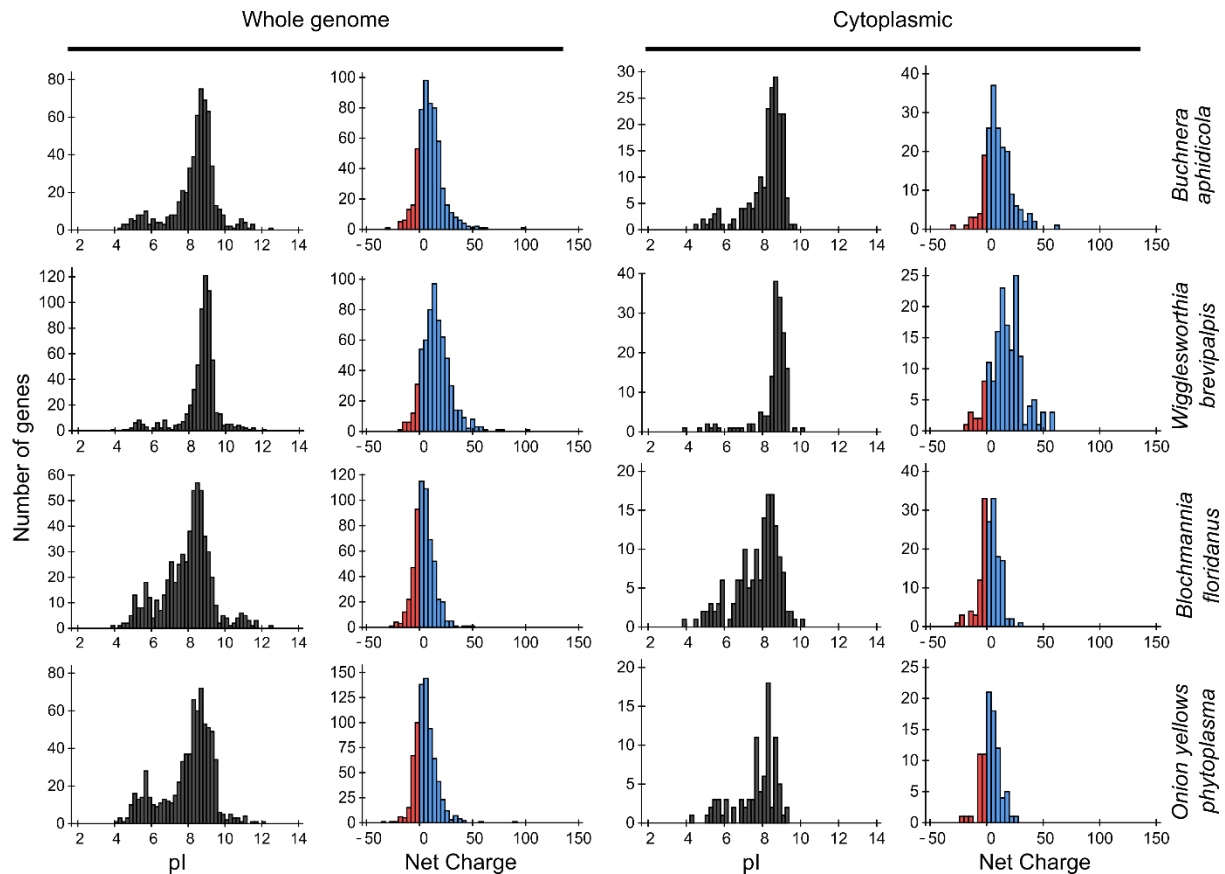From the pI distributions it is clear that in all three organisms the majority of cytoplasmic proteins is acidic and thus negatively charged at the physiological internal pH of 7.5; for internal pH values we refer to (35-37) for *E. coli* (see also BNID 105980 and BNID 106518) (38), and (24) for *L. lactis*. The protein net charge distributions of *E. coli* and *L. lactis* go up to a value of +25, irrespective of whether we take the gene-based distributions or protein copy numbers (for *E. coli*), or whether we take the full or cytoplasmic proteome. The net charge distribution of *Hfx. volcanii* stops at about 0. In *E. coli* 35 cytoplasmic proteins have a net charge higher than +10. These consist of 18 ribosomal, 9 RNA-associated, 5 DNA-associated and 3 uncharacterized proteins. For *L. lactis*, with 7 cytoplasmic proteins that have a net charge bigger than +10, the breakdown is similar. The only *Hfx. volcanii* protein with a net charge bigger than 0 is a ribonuclease, rnp4, with a net charge of +6. Thus, all three organisms essentially have no "free" positive cytoplasmic proteins.

The drop in diffusion coefficient with increased positive charge may partly explain why the proteome is mostly negative in *E. coli*, *L. lactis* and *Hfx. volcanii*. Positive proteins not only diffuse slowly but their binding to ribosomes might also inhibit protein synthesis, that is, by affecting the assembly or the activity of ribosomes. As mentioned, there are 18 ribosomal proteins in *E. coli* that have a net charge of more than +10. If the findings on GFP are transferable to ribosomal proteins, then these ribosomal proteins experience a drop in diffusion coefficient of more than 5-fold. The most extreme cases, RplT

and RplB, with net charges of +24 and +31, would have a drop in diffusion coefficient of 100-fold. The fact that ribosomal proteins themselves are positively charged potentially limits the rate of assembly of new ribosomes. The reason is that ribosomal proteins that have just been synthesized bind nonspecifically to the surfaces of fully assembled ribosomes, and this causes slow diffusion, lower effective protein concentrations and may affect the functioning of ribosomes. An implication of our findings is that the synthesis of cationic ribosomal proteins and the assembly of ribosomes should be highly coordinated and preferably be modular to minimize unwanted side-effects of nonspecific interactions (39).

A high positive net charge is not a guarantee for binding to the ribosome. You could imagine a positive protein that is disordered before binding but ordered upon binding. The reduction in entropy reduces the binding affinity which in turn causes the diffusion coefficient to be high. Another option is that the surface shapes don't match even if the net charges are complementary. This would also lower the affinity and increase the diffusion coefficient.

A conundrum is encountered when you look at the proteomes of the bacteria *Buchnera aphidicola*, *Blochmannia floridanus*, *Onion yellows phytoplasma*, and *Wigglesworthia glossinidia brevipalpis*. All four organisms are (endo)symbionts of plants or insects and have small genomes: 572-730 protein-encoding genes. All of these have very basic proteomes (Figure 7b and Figure 7-figure supplement 2).

For *Buchnera* the number of cytoplasmic proteins with a net charge of more than +10 is 89, out of 190 total cytoplasmic proteins. For *Wigglesworthia* the numbers are 119 out of 158. The number of cytoplasmic proteins with a net charge bigger than +25 is 18 for *Buchnera* and 49 for *Wigglesworthia*. It is unclear how these organisms are able to deal with, or avoid, slow diffusion and ribosomes getting swarmed with positive proteins. It is possible that these organisms have a cytoplasmic pH higher than 7.5, but even at pH 8.5 a large fraction of the proteins is still positive (75% for *Wigglesworthia glossinidia brevipalpis* compared to 90% at pH 7.5). There is also a practical consequence to these findings: to (over)express proteins from any of these four organisms in *E. coli* may be less favorable than in *L. lactis* or *Hfx. volcanii* which have a higher ionic strength.

A few general points on diffusion and binding

A protein can diffuse only as slow as the combination of this protein and the slowest component it binds to in the cell. This is described by equation (1), which can be rewritten in the following way:

$$D_{eff} = f_{free} (D_{free} - D_{bound}) + D_{bound} \qquad (6)$$

If we fill in f_{free} = 0 we get $D_{eff} = D_{bound}$ and if we fill in f_{free} = 1 we get $D_{eff} = D_{free}$; the black lines in Figure 8. If we fill in D_{bound} = 0 we get $D_{eff} = f_{free} D_{free}$, i.e. after some point D_{eff} becomes independent of D_{bound}; this is illustrated by the grey lines in Figure 8. In more concrete terms: 99% binding of +25 GFP to ribosomes (D_{ribo} = 0.04 µm²/s) leads to the same diffusion coefficient as 99 % binding to DNA (D_{DNA} = 0.000035-0.00007 µm²/s).
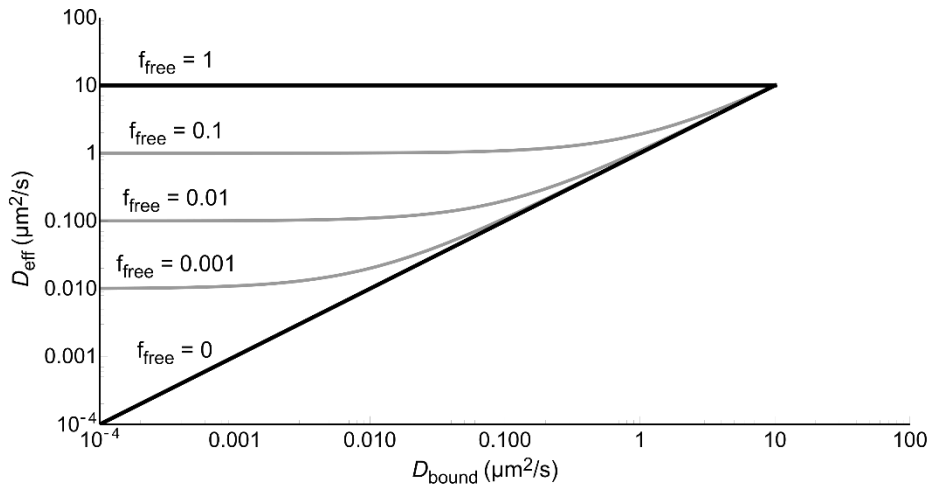
Figure 8: The effective diffusion coefficient as a function of free fraction and the diffusion coefficient of the bound complex. All lines are generated with equation (6) and D_{free} = 10 μm²/s.

Conclusion

We find that the diffusion coefficients of proteins in the cytoplasm of *E. coli* depends on their net charge and the distribution of charge over the protein surface, with positive proteins moving up to 100-fold slower. The diffusion becomes even slower when cells are exposed to an osmotic upshift. In *L. lactis* and *Hfx. volcanii* the slowdown in diffusion with increasing positive surface charge is less than in *E. coli* due to electrostatic screening. The decrease in diffusion rate is caused by binding of positive proteins to ribosomes, with K_D values on the order of μM; showing that non-selected interactions need not be weak. Ribosome surface properties thus limit the composition of the cytoplasmic proteome. These findings are of general value due to the universal presence of ribosomes in cells. Application of these findings to bacterial (endo)symbionts lays bare a paradox in the functioning of these cells.

Materials and methods

Strains used

We use *E. coli* strain MG1655 (40). The GFP variants, -30, -7, 0, +7, +11a, +11b, +15, and +25, were all expressed from an L-arabinose inducible promoter on a pBAD vector with an ampicillin-resistance selection marker. We obtained the genes for -30, -7, +15, and +25 from David Liu's lab at Harvard, and nucleotide sequences of +7, +11a, +11b from David Thompson. See McNoughton et al. (41) for -30, -7, +15, and +25 GFP; see Thompson et al. (42) for +7, +11a and +11b. We designed the 0 GFP ourselves. All GFP variants have an N-terminal histag. For all GFP variants amino acid sequences are available in Supplementary file 1A.

We used *Lactococcus lactis* strain NZ9000 (43), which contains the *nisR* and *nisK* genes which in the presence of the inducer, nisin A, switches on the expression of genes from the *nisA* promoter (44). We cloned the coding segments for the his-tagged versions of -30 GFP, -7 GFP, +15 GFP and +25 GFP behind the *nisA* promoter in the pNZ8048 vector (44). The expression levels of -30 GFP were too low for FRAP analysis.

We used *Hfx. volcanii* strain H1895 (45). We expressed the GFP variants, -30, -7, +15, and +25, from the pTA1228 plasmid, which has a tryptophan-inducible promoter (46). The amino acid sequences are the same as for *E. coli* and *L. lactis*. We optimized the nucleotide sequence for *Hfx. volcanii* by making the codon frequency in the GFP variants the same as for protein coding genes on the *Hfx. volcanii* chromosome. We constructed the pTA1228 bearing the genes for GFP variants in *E. coli*, and transformed the final plasmids in *Hfx. volcanii* by protocols described in the Halohandbook (47). We obtained the strain, plasmid and the codon usage table from Thorsten Allers (University of Nottingham).

Preparation of *E. coli* for FRAP

For each experiment we took a glycerol stock of *E. coli* with one of the GFP variants and stabbed it with a pipette tip to obtain a small amount of cells. These we deposited in 4 mL LB medium containing 0.2 % v/v glycerol and 100 µg/mL ampicillin. We incubated the culture at 30 °C, with 200 rpm shaking. The next day we took 8 µL of the LB culture to inoculate 4 mL MBM containing 0.2 % v/v glycerol and 100 µg/mL ampicillin. The composition of MBM, MOPS based medium, can be found in Neidhardt et al. (48). We adjusted the osmolality of MBM to 0.28 Osm with NaCl. Osmolalities were measured with an Osmomat 030 cryoscopic osmometer (Gonotec, Germany). For the -30 GFP variant we also added 0.4 % L-arabinose (from a 20 % w/v stock in MilliQ), to induce protein expression. Again we incubated at 30 °C, with 200 rpm shaking. The next morning the cultures had reached an OD_{600} between 0.6-1.6, which were diluted to an OD_{600} of 0.18-0.25. At the moment of dilution we also induced the expression of -7 GFP, with 0.1 % w/v L-arabinose, and 0, +7 and +11b GFP, with 0.4 % w/v L-arabinose. We incubated the cultures for a further 2-4 hours, to obtain an OD_{600} of 0.4-0.5, and then performed the FRAP measurements. To avoid aggregation, the +11a, +15 and +25 GFP variants were induced with 0.4 % w/v L-arabinose 1-2 hours before the FRAP measurements. For each GFP variant the experiment was repeated at least 3 times.

Preparation of *L. lactis* for FRAP

For each experiment we took a glycerol stock from the -80 °C freezer and stabbed it with a pipette tip to obtain a small amount of cells. These cells were deposited in 4 mL of growth medium in a culturing tube. The growth medium was CDM in all cases; the formulation of the chemically defined medium (CDM) is given in the supplement of (12), where it is referred to as CDM^RP. There is one difference,

here we also added L-proline (0.68 g/L, final concentration). We include glucose (1 % w/v), as a carbon and energy source, and chloramphenicol (5 µg/mL), to maintain the plasmids. We incubated the cultures at 30 °C, without shaking (*L. lactis* grows semi-anaerobically). In the morning of the next day about 100 uL of culture was added to 4 mL of fresh CDM, to yield an OD_{600} of about 0.1. Simultaneously, we added 4 uL of nisin A solution (filtered supernatant from a *L. lactis* NZ9700 culture). The cultures were incubated at 30 °C. We used the cultures for FRAP measurements at an OD_{600} of 0.38-0.46. We diluted the cultures to keep them from overgrowing. For each GFP variant the experiment was repeated 2-3 times.

Preparation of *Hfx. volcanii* for FRAP

For each experiment we took 2-3 colonies of *Hfx. volcanii*, expressing -30, -7, +15, or +25 GFP, from an Hv-YPC agarose plate and suspended these in 4 mL Hv-YPC medium (47). We incubated the cultures at 42 °C, with 200 rpm shaking. The next morning we diluted the culture to an OD_{600} of 0.2-0.3 and at the same time added 4 mM L-tryptophan, to induce expression of the GFP variants. We incubated the cultures for 2-3 hours at 42 °C, with 200 rpm shaking, before using the cells for FRAP measurements. At the time of the measurements the OD_{600} was 0.3-0.5. For each GFP variant the experiment was repeated 3 times.

Determination of diffusion coefficients

We performed fluorescence recovery after photo-bleaching (FRAP; see Figure 1a, b) on a LSM710 Zeiss confocal laser scanning microscope, following a method originally described by Elowitz et al.(49). Our implementation of this method is described in (12). We started with an overview image containing many cells, from which picked cells that are lying flat, are not undergoing cell division and have no neighbors that would obscure the analysis. We take a high resolution close-up of the cell and its immediate surroundings to see if the cell is fit for measurements. For FRAP we programmed the microscope to take three images, then photo bleach the GFP at one of the cell poles and finally record the recovery of the fluorescence over time. We recorded all images with a 488 nm laser; the same laser was used for bleaching but at a higher power.

For *E. coli* we did the FRAP measurements as follows. We took 400 µL culture and resuspended those cells twice in 300 µL MBM*. MBM* did not have glycerol or ampicillin and has Na^+ instead of K^+. The osmolality of the MBM* was either 0.28, 0.55 or 1.2 Osm. We adjusted the osmolality of the resuspension medium with NaCl. After we resuspended the cells we pipetted 4 µL of these cells on a cover slide. To make sure the *E. coli* cells didn't move we used (3-aminopropyl)triethoxysilane (APTES)-treated cover slides. The slides were prepared as follows. First we cleaned them by sonicating for 1 h in 5 M KOH, rinsing 10 times with MilliQ and blowing off the remaining MilliQ with pressurized nitrogen. We then deposited the slides in acetone that contained 2 % v/v (3-aminopropyl)triethoxysilane. We incubated for 5 min at room temperature, removed the acetone and APTES and rinsed the slides 10 times with MilliQ. Again, the remaining MilliQ was blown off with pressurized nitrogen. After putting our cells on the APTES slide we put an object slide on top, for stability, and put the whole on the microscope stage. The stage temperature was maintained at 30 °C. We used the slide for no longer than 20 min after depositing the cells. For -30, -7 and 0 GFP under normal conditions, that is, no osmotic upshift, we recorded images at 50 time points and with 8x8 pixels, with a 4 ms exposure time and no extra time between exposures. For the positive GFP variants, and for the FRAP experiments after osmotic upshift (slower diffusion), we recorded images at a 100 time points, with 16x16 pixels, and a 8 ms exposure and a 8-100 ms time step between the start subsequent exposures.

86

For *L. lactis* we pipetted 4 uL culture on a cover slide. We put an object slide on top, for stability, and the whole was put on the stage of the microscope (maintained at 30 °C). We made sure that the cells didn't move by using cover slides that were sonicated for 1 h in 5 M KOH, rinsed 10 times with MilliQ and dried by blowing of the remaining MilliQ with pressurized nitrogen. We used cover slides for no longer than 20 min after depositing cells, and cultures for no longer than 1 h after reaching an OD_{600} of 0.38-0.46. We recorded the images in 8x10 pixels and with an exposure time of 5-37 ms, short for -7 GFP and long for +25 GFP. For each cell we recorded the whole FRAP measurement in 50 images, without extra time between exposures.

For *Hfx. volcanii* we pipetted 4 μL culture on a cover slide (non-treated). Then a patch of 1 % agar in 18 % SW was put on top, to immobilize the cells (see the Halohandbook (47) for the composition of SW). We then put the sample on the microscope stage. The stage temperature was maintained at 30 °C. After a couple of minutes the cells stopped moving and we performed our FRAP measurements. We imaged for 20 min on a single sample. We recorded the images in 16x16 pixels with an exposure time of 8 ms and a time step, the time between the start of subsequent exposures, of 8-20 ms. For +15 and +25 GFP we see clear aggregates in some cells (Figure 2-figure supplement 1a). We did not include these cells in the FRAP measurements.

Overall about 10 % of the analyzed cells were too noisy and/or did not show a bleached area, so that a diffusion coefficient could not be determined. Those cells were excluded from the analysis. For the FRAP results of -30 GFP (1.2 Osm shock), -7 GFP and 0 GFP in *E. coli* we excluded around 20 % of the cells. For 0 GFP this was necessary because the fluorescence level was too low in a fraction of the cells; for -30 (shock) and -7 GFP the apparent diffusion was too fast for detection. When we include the excluded cells we obtain the following medians: 0.28 μm^2/s for -30 GFP (shock), 11.3 μm^2/s for -7 GFP and 9.8 μm^2/s for 0 GFP. This can be compared to the values 0.18, 10 and 8.6 μm^2/s reported in Supplementary file 1B.

Co-localization of +25 GFP and DNA in *E. coli*

We grew *E. coli*, containing +25 GFP, in a culturing tube with 4 mL EZ; a rich defined growth medium (from Teknova), to which we added glycerol (0.2 % v/v), as a carbon and energy source, and ampicillin (100 μg/mL), to maintain the plasmid. We used EZ medium because the chloramphenicol didn't condense the DNA in cells grown in MBM. The culture was incubated at 30 °C, with 200 rpm shaking for aeration. The next morning we used 200 uL of this culture to inoculate 4 mL of fresh EZ medium. We incubated the culture for 1 h before adding 0.1 % (w/v) L-arabinose, to induce the expression of +25 GFP. After another hour of incubation, at an OD_{600} of 0.5-0.6, we added DRAQ-5 (2 μM), to visualize the DNA, and chloramphenicol (200 μg/mL), to condense the DNA (28). We imaged these cells between 1 and 1.5 h after adding the DRAQ-5 and chloramphenicol, for which we deposited 10 uL culture on an APTES cover slide, prepared as described above, put an object slide on top and put the whole on the microscope stage. We performed all measurements at 30 °C. We focused on a 200 μm x 200 μm area, containing ~200 cells, and recorded an image in the 488 (+25 GFP) and 633 (DNA) channels and used an exposure time of 3.3 s. We also recorded the transmitted excitation light to obtain a bright-field image. We picked the cells used for analysis from the transmission image to avoid bias. We selected cells that were lying flat and in focus.

Diffusion coefficients in DNA-free regions

We did FRAP measurements on -7 GFP and +25 GFP in DNA free regions in *E. coli*. The cells were grown in a culturing tube with 4 mL LB medium, containing glycerol (0.2 % v/v), as a carbon and energy source, and ampicillin (100 μg/mL), to maintain the plasmid. We incubated the culture overnight at

37 °C, with 200 rpm shaking. We used LB medium, and incubated at 37 °C, to get elongated cells in a reasonable time window upon cephalexin treatment (see below). The next day we made two new cultures, 4 mL LB (same composition as above), by adding 4 µL or 16 µL of overnight culture. We incubated these cultures at 37 °C, with 200 rpm shaking. At an OD_{600} between 0.19-0.26 we added L-arabinose (0.1-0.2 % w/v), to induce GFP expression, and cephalexin (25 µg/mL), to elongate the cells. After a further two hours of incubation we added chloramphenicol (200 µg/mL), to condense the DNA, and DRAQ-5 (10 µM), to stain the DNA. The concentration of DRAQ-5 is above the minimum inhibitory concentration, 5 µM, for growth of *E. coli* MG1655 in EZ medium (28); at lower concentrations of DRAQ-5 the DNA did not stain properly. After a further 30-70 min incubation we took 200-400 µL of culture and resuspended the culture twice in 200 µL LB (with 0.2 % v/v glycerol, 100 µg/mL ampicillin and 200 µg/mL chloramphenicol), to get rid of the DRAQ-5 background fluorescence. We used the following stock solutions: 20 % w/v L-arabinose in MilliQ, 1 mg/mL cephalexin in MilliQ, 20 mg/mL chloramphenicol in ethanol, and 200 µM DRAQ-5 in MilliQ.

We deposited 4 µL of the sample on an APTES cover slide, put an object slide on top and put the whole on the microscope stage. The microscope stage temperature was maintained at 37 °C. The cells were on the stage for no longer than 1 h. We measured the diffusion coefficients of -7 GFP and +25 GFP in these treated cells by the same method as before. The exception is that we do not draw the line for FRAP analysis from pole to pole. We draw it either to where DNA blocks movement of GFP (see Figure 4-figure supplement 2a) or far enough away from the bleach. It is important that the boundary conditions that are used in the analysis are still satisfied, meaning no transfer of particles over the ends of the line. We also make sure to only measure at places where there is no DNA present. We recorded the recovery in 150 time points with 750 ms exposure time for +25 GFP and 23-30 ms for -7 GFP. There was no additional time between exposures. The experiment was performed twice for both -7 GFP and +25 GFP.

Diffusion coefficients in the presence and absence of mRNA

We performed FRAP measurements over time of +25 GFP expressing *E. coli* cells with and without rifampicin treatment. The cells were grown in a tube with 4 mL LB (with 0.2 % v/v glycerol and 100 µg/mL ampicillin). We incubated the culture overnight at 30 °C, with 200 rpm shaking. The next day we used 8 µL of the overnight culture to inoculate a 4 mL MBM culture (with 0.2 % v/v glycerol and 100 µg/mL ampicillin). Again we incubated the culture at 30 °C, with 200 rpm shaking. The next morning, depending on the OD_{600}, the cultures were diluted with more MBM or allowed to continue growing. When the culture reached an OD_{600} of 0.17-0.23 the production of +25 GFP was induced by adding 0.4 % w/v L-arabinose (from a 20 % w/v in MilliQ stock). At OD_{600} of 0.30-0.34 we took 198 µL culture and added either 2 µL DMSO (control) or 2 µL DMSO with 50 mg/mL rifampicin (for a final concentration of 0.5 mg/mL). After mixing we deposited 4 µL sample on an APTES cover slide, put an object slide on top and the put the whole on the microscope stage. The stage was maintained at 30 °C for the duration of the experiment. We measured the diffusion coefficients by FRAP and recorded the time for each measurement. We did each FRAP measurement on a unique cell. The replicates represent separate cultures on the day of the FRAP measurements.

Sucrose gradient centrifugation of *E. coli* cell lysates and purified DNA

For each experiment on *E. coli* cell lysates, with *E. coli* expressing either -7 or +25 GFP, we deposited a small amount of cells from a glycerol stock into 4 mL LB medium (containing 0.2 % v/v glycerol and 100 µg/mL ampicillin). We incubated the cultures at 30 °C with 200 rpm shaking overnight. The next day we took 20 µL of the LB culture to inoculate 10 mL MBM (0.28 Osm, osmolality adjusted with NaCl, containing 0.2 % v/v glycerol and 100 µg/mL ampicillin). Again we incubated overnight. The next

morning the cultures had reached an OD_{600} between 0.8-1.3, and were diluted to an OD_{600} of 0.1-0.16 in 50 mL of fresh medium in a 250 mL flask. We added 0.4 % (w/v) L-arabinose to induce protein expression. We incubated the cultures for 3.5-4 hours, to obtain an OD_{600} of 0.25-0.37. We centrifuged 44 mL of each culture at 5250 g, 20 minutes, 4 °C. From this point onward we did all the work on ice and used cooled buffers. Pellets were suspended in 1 mL of resuspension buffer (20 mM Tris-HCl pH 7.5, 15 mM magnesium acetate, 100 mM ammonium acetate plus 6 mM 2-mercaptoethanol), resulting in around 200-300 fold dilution of cytoplasmic content. To calculate the dilution we assumed a cytoplasmic volume 0.5 fL (50) and 8×10^8 cells in 1 mL of culture of OD_{600} of 1. To each tube containing the suspension we added around 0.2 mg of 106 µm glass beads (Sigma) and lysed the cells using two repetitions of 50 Hz oscillation for 5 minutes (TissueLyser LT, QIAGEN). We cooled the sample on ice in between repetitions. We added PMSF (100 mM in isopropanol stock) to the lysates to a final concentration of 1 mM. Then the lysates were centrifuged at 9000 g, for 2 min at 4 °C. We took the supernatant and centrifuged it at 9000 g for 15 min at 4 °C. We layered 800 µL of the resulting supernatant onto 8 mL of linear 10-40 % sucrose gradient. The sucrose solutions contained 1 mM PMSF and were prepared with the resuspension buffer. We centrifuged the gradients using a swing-out rotor (SW 32.1 Ti, Beckman) at 125 000 g for 80 minutes at 4 °C as described previously (51,52). We recorded a fluorescence profile over the sucrose gradient by dividing the gradient in 600 µL fractions, and measuring the fluorescence intensity for each fraction in a Jasco FP-8300 fluorimeter. We excited with 488 nm and recorded the emission from 500-600 nm (in 5 nm intervals). For the analysis we used the fluorescence emission at 510 nm. To correct for background fluorescence, we acquired spectra of 10, 20, 30 and 40 % sucrose in resuspension buffer with 1 mM PMSF. A linear fit of the 510 nm emission intensities was used to calculate sucrose-caused background values for each fraction. For lysates of *E. coli* expressing -7 and +25 GFP the experiment was carried out two times.

For the samples containing DNA, we dissolved salmon testes DNA in autoclaved MQ to a final concentration of 1 mg/mL. The DNA was added to the cell lysates directly after the second centrifugation step; the sample was incubated for around 30 minutes on ice, before layering it onto the sucrose gradient. To determine the fractionation profile of DNA we layered 0.2 mg/ml DNA onto the linear sucrose gradient, except that the 2-mercaptoethanol and PMSF were omitted. The collected fractions were diluted 1:1 in 20 mM Tris-HCl pH 7.5, 15 mM magnesium acetate, 100 mM ammonium acetate. We determined DNA levels by measuring the absorbance of each fraction from 200-340 nm with 5 nm intervals, using Cary 100 Bio UV-VIS spectrometer. To correct for background absorption we measured fractions of pure MQ sample treated in the same way. For lysates of *E. coli* expressing +25 GFP with addition of DNA the experiment was carried out once for each DNA concentration; the DNA control experiment was also done once.

Electron microscopy

We dialyzed the fractionated cell lysate samples in pre-cooled 20 mM Tris-HCl pH 7.5, 15 mM magnesium acetate, 100 mM ammonium acetate for 1 hour to remove the sucrose. The samples were pipetted on glow-discharged carbon-coated copper grids, excess liquid was removed by blotting and the grids were stained with 2 % uranyl acetate. EM was performed on a Tecnai T20 electron microscope (FEI, Eindhoven, The Netherlands) operated at 200 kV, images were acquired with a 4000 SP 4K slow-scan CCD camera (Gatan, Pleasanton, CA, USA) as described previously (52).

Computational analysis of proteomes

We made histograms of the distributions of pI and net charge of all proteins encoded by the genomes of *E.coli*, *L. lactis* and *Hfx. volcanii*. All protein sequence data was retrieved from UniProt (53). For *E. coli* we used a K12 strain (proteome ID: UP000000318), for *L. lactis* we used strain MG1363 (proteome

ID: UP000000364) and for *Hfx. volcanii* we used strain DS2 (proteome ID: UP000008243). The *L. lactis* MG1363 strain is the parent strain of *L. lactis* NZ9000 that we used for FRAP (43). The *Hfx. volcanii* DS2 is the parent strain of *Hfx. volcanii* H1895 (45). We calculated the pI of each protein based on its amino acid sequence using the Isoelectric Point Calculator by Kozlowski (54). We modified the program to allow for net charge calculations; the pI and net charge values we report are based on the IPC_protein pK$_a$ dataset of the Isoelectric Point Calculator. To calculate the net charge we used a pH of 7.5. To get the distributions in the cytoplasm we took only those proteins that have gene ontology labels cytoplasm and cytosol in the uniprot database(GO:0005737 and GO:0005829). For *E. coli* this yielded 1406 proteins (compared to 4254 proteins in the full genome), for *L. lactis* 253 (2383), and for *Hfx. volcanii* 177 (3987). For *E. coli* we also made pI and net charge distributions in which protein copy numbers are taken into account. To do this we took copy number data from Schmidt et al. (55). Specifically, we took the abundance data for *E. coli* BW25113 cultured in M9 glycerol.

We also made histograms of the distributions of pI and net charge of all proteins encoded by the genomes of *Buchnera aphidicola* (proteome ID: UP000001806), *Blochmannia floridanus* (proteome ID: UP000002192), *Onion yellows phytoplasma* (proteome ID: UP000002523), and *Wiggleworthia glossinidia brevipalpis* (proteome ID: UP000000562). For Buchnera the cytoplasmic fraction contained 190 proteins (compared to 572 proteins in the full genome), for Blochmannia 156 (583), for phytoplasma 87 (730), and Wigglesworthia 158 (617).

Calculation of protein charge

To calculate the charge of the GFP variants we counted the number of Asp, Glu, Lys and Arg residues, used the pK$_a$ values of all (de)protonatable residues and used the Henderson-Hasselbalch equation to calculate the net charge. Results of these calculations give net charge 1-2 higher than net charge calculated using the modified IPC and the IPC_protein pK$_a$ dataset (-31.1, -8.2, -1.3, +5.6, +9.5, +9.5, +13.6, +23.5 for -30, -7, 0, +7, +11a, +11b, +15 and +25 GFP respectively). In reality, ions can specifically bind to proteins and thereby change the base net charge (*i.e.* before any ionic screening effects occur) (56). This is especially true for anions (56), which could affect our (quantitative) interpretation. Two examples: bovine serum albumin (measured charge, -13.8; calculated charge, -18.3) (56) and hen egg white lysozyme (measured charge, +5.1; calculated charge, +11) (57); the actual values depend on the type(s) of ion(s) present(57). We also assumed that each residue of a particular type (e.g. all aspartates) have the same pK$_a$ independent of context. To take ion binding and context dependent pK$_a$ values into account would be a whole study in itself.

Derivation of equation 1

For diffusion in one dimension the probability density for the position of a particle after time, t, is given by:

$$p(x) = \frac{N}{\sqrt{4\pi Dt}} e^{-\frac{x^2}{4Dt}} \quad (7)$$

Here x is the position, $p(x)$ is the probability as a function of x, N is a normalization factor, D is the diffusion coefficient, and t is the time step. When the particle is free it moves with diffusion coefficient D_{free} and when bound with D_{bound}. The particle goes back and forth between free and bound states a number of times in a certain period of time, Δt. Because the motion in each time step is independent of the other time steps, we can sum all time steps and distances travelled for the free state and we can do the same for the bound state. We end up with two equations like equation 7 but in one we have $D = D_{free}$ and $t = f_{free}\Delta t$ and in the other $D = D_{bound}$ and $t = (1 - f_{free})\Delta t$, with f_{free} being the fraction of time that the particle is free. To get the probability density for the position of the

particle after time step, Δt, we convolute the two equations. A convolution of a Gaussian function leads to another Gaussian in the following way:

$$f(x) * g(x) = \frac{N}{\sqrt{2\pi}\sigma_f} e^{-\frac{x^2}{2\sigma_f{}^2}} * \frac{N}{\sqrt{2\pi}\sigma_g} e^{-\frac{x^2}{2\sigma_g{}^2}} = \frac{N}{\sqrt{2\pi(\sigma_f{}^2 + \sigma_g{}^2)}} e^{-\frac{x^2}{2(\sigma_f{}^2+\sigma_g{}^2)}} \quad (8)$$

Here, $*$, is the symbol for a convolution. By comparing equation 7 and 8 we can see that $\sigma_f{}^2 = 2D_{free}f_{free}\Delta t$ and $\sigma_g{}^2 = 2D_{bound}(1 - f_{free})\Delta t$. We can also define an effective diffusion coefficient, D_{eff}, such that $\sigma_f{}^2 + \sigma_g{}^2 = 2D_{eff}\Delta t$. Combining the last results and dividing by $2\Delta t$ we obtain equation 1.

References

(1) Klumpp S, Scott M, Pedersen S, Hwa T. Molecular crowding limits translation and cell growth. Proc Natl Acad Sci U S A 2013 OCT 15;110(42):16754-16759.

(2) Vijayakumar M, Wong K, Schreiber G, Fersht A, Szabo A, Zhou H. Electrostatic enhancement of diffusion-controlled protein-protein association: Comparison of theory and experiment on barnase and barstar. J Mol Biol 1998 MAY 22;278(5):1015-1024.

(3) van den Berg J, Boersma AJ, Poolman B. Microorganisms maintain crowding homeostasis. Nat Rev Micro 2017;15(5):309-318.

(4) McGuffee SR, Elcock AH. Diffusion, Crowding & Protein Stability in a Dynamic Molecular Model of the Bacterial Cytoplasm RID F-4799-2010. Plos Computational Biology 2010 MAR;6(3):e1000694.

(5) Crowley PB, Chow E, Papkovskaia T. Protein Interactions in the Escherichia coli Cytosol: An Impediment to In-Cell NMR Spectroscopy. Chembiochem 2011 MAY 2;12(7):1043-1048.

(6) Ye Y, Liu X, Zhang Z, Wu Q, Jiang B, Jiang L, Zhang X, Liu M, Pielak GJ, Li C. F-19 NMR Spectroscopy as a Probe of Cytoplasmic Viscosity and Weak Protein Interactions in Living Cells. Chemistry-a European Journal 2013 SEP 16;19(38):12705-12710.

(7) Latham MP, Kay LE. Probing non-specific interactions of Ca2+-calmodulin in E. coli lysate. J Biomol NMR 2013 MAR;55(3):239-247.

(8) Levy ED, De S, Teichmann SA. Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. Proc Natl Acad Sci U S A 2012 DEC 11;109(50):20461-20466.

(9) McConkey EH. Molecular evolution, intracellular organization, and the quinary structure of proteins. Proc Natl Acad Sci U S A 1982 May;79(10):3236-3240.

(10) Mika JT, Poolman B. Macromolecule diffusion and confinement in prokaryotic cells. Curr Opin Biotechnol 2011 FEB;22(1):117-126.

(11) Konopka MC, Sochacki KA, Bratton BP, Shkel IA, Record MT, Weisshaar JC. Cytoplasmic Protein Mobility in Osmotically Stressed Escherichia coli. J Bacteriol 2009 JAN;191(1):231-237.

(12) Mika JT, Schavemaker PE, Krasnikov V, Poolman B. Impact of osmotic stress on protein diffusion in Lactococcus lactis. Mol Microbiol 2014 NOV;94(4):857-870.

(13) Parry BR, Surovtsev IV, Cabeen MT, O'Hem CS, Dufresne ER, Jacobs-Wagner C. The Bacterial Cytoplasm Has Glass-like Properties and Is Fluidized by Metabolic Activity. Cell 2014 JAN 16;156(1-2):183-194.

(14) Joyner RP, Tang JH, Helenius J, Dultz E, Brune C, Holt LJ, Huet S, Mueller DJ, Weis K. A glucose-starvation response regulates the diffusion of macromolecules. Elife 2016 MAR 22;5:e09376.

(15) Munder MC, Midtvedt D, Franzmann T, Nueske E, Otto O, Herbig M, Ulbricht E, Mueller P, Taubenberger A, Maharana S, Malinovska L, Richter D, Guck J, Zaburdaev V, Alberti S. A pH-driven transition of the cytoplasm from a fluid- to a solid-like state promotes entry into dormancy. Elife 2016 MAR 22;5:e09347.

(16) Kumar M, Mommer MS, Sourjik V. Mobility of Cytoplasmic, Membrane, and DNA-Binding Proteins in Escherichia coli. Biophys J 2010 FEB 17 2010;98(4):552-559.

(17) Nenninger A, Mastroianni G, Mullineaux CW. Size Dependence of Protein Diffusion in the Cytoplasm of Escherichia coli. J Bacteriol 2010 SEP 2010;192(18):4535-4540.

(18) Mika JT, van den Bogaart G, Veenhoff L, Krasnikov V, Poolman B. Molecular sieving properties of the cytoplasm of Escherichia coli and consequences of osmotic stress. Mol Microbiol 2010 JUL;77(1):200-207.

(19) Konopka MC, Shkel IA, Cayley S, Record MT, Weisshaar JC. Crowding and confinement effects on protein diffusion in vivo. J Bacteriol 2006 SEP;188(17):6115-6123.

(20) van den Bogaart G, Hermans N, Krasnikov V, Poolman B. Protein mobility and diffusive barriers in Escherichia coli: consequences of osmotic stress. Mol Microbiol 2007 MAY;64(3):858-871.

(21) Spitzer J, Poolman B. The Role of Biomacromolecular Crowding, Ionic Strength, and Physicochemical Gradients in the Complexities of Life's Emergence. Microbiology and Molecular Biology Reviews 2009 JUN;73(2):371-+.

(22) Spitzer J, Poolman B. How crowded is the prokaryotic cytoplasm? FEBS Lett 2013 JUL 11;587(14):2094-2098.

(23) Shabala L, Bowman J, Brown J, Ross T, McMeekin T, Shabala S. Ion transport and osmotic adjustment in Escherichia coli in response to ionic and non-ionic osmotica. Environ Microbiol 2009 JAN;11(1):137-148.

(24) Poolman B, Hellingwerf KJ, Konings WN. Regulation of the glutamate-glutamine transport system by intracellular pH in Streptococcus lactis. J Bacteriol 1987 May;169(5):2272-2276.

(25) Pérez-Fillol M, Rodriguez-Vallera F. Potassium ion accumulation in cells of different halobacteria. Microbiologia 1986;2(2):73-80.

(26) Reyes-Lamothe R, Possoz C, Danilova O, Sherratt DJ. Independent positioning and action of Escherichia coli replisomes in live cells. Cell 2008 APR 4;133(1):90-102.

(27) Sanamrad A, Persson F, Lundius EG, Fange D, Gynna AH, Elf J. Single-particle tracking reveals that free ribosomal subunits are not excluded from the Escherichia coli nucleoid. Proc Natl Acad Sci U S A 2014 AUG 5;111(31):11413-11418.

(28) Bakshi S, Siryaporn A, Goulian M, Weisshaar JC. Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. Mol Microbiol 2012 2012-Jul;85(1):21-38.

(29) Bernstein J, Khodursky A, Lin P, Lin-Chao S, Cohen S. Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. Proc Natl Acad Sci U S A 2002 JUL 23;99(15):9697-9702.

(30) Milo R, Phillips R. What is the macromolecular composition of the cell? Cell biology by the numbers New York, NY, USA: Garland Science; 2015. p. 130.

(31) Vendeville A, Lariviere D, Fourmentin E. An inventory of the bacterial macromolecular components and their spatial organization. FEMS Microbiol Rev 2011 MAR;35(2):395-414.

(32) Schreiber G, Fersht AR. Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. Biochemistry 1993 May 18;32(19):5145-5150.

(33) Wallis R, Moore GR, James R, Kleanthous C. Protein-protein interactions in colicin E9 DNase-immunity protein complexes. 1. Diffusion-controlled association and femtomolar binding for the cognate complex. Biochemistry 1995 Oct 24;34(42):13743-13750.

(34) Stone SR, Dennis S, Hofsteenge J. Quantitative evaluation of the contribution of ionic interactions to the formation of the thrombin-hirudin complex. Biochemistry 1989 Aug 22;28(17):6857-6863.

(35) Slonczewski JL, Rosen BP, Alger JR, Macnab RM. pH homeostasis in Escherichia coli: measurement by 31P nuclear magnetic resonance of methylphosphonate and phosphate. Proc Natl Acad Sci U S A 1981 Oct;78(10):6271-6275.

(36) Zilberstein D, Agmon V, Schuldiner S, Padan E. The sodium/proton antiporter is part of the pH homeostasis mechanism in Escherichia coli. J Biol Chem 1982;257(7):3687-3691.

(37) Wilks JC, Slonczewski JL. pH of the cytoplasm and periplasm of Escherichia coli: Rapid measurement by green fluorescent protein fluorimetry. J Bacteriol 2007 AUG;189(15):5601-5607.

(38) Milo R, Jorgensen P, Moran U, Weber G, Springer M. BioNumbers-the database of key numbers in molecular and cell biology. Nucleic Acids Res 2010 JAN;38:D750-D753.

(39) Davis JH, Tan YZ, Carragher B, Potter CS, Lyumkis D, Williamson JR. Modular Assembly of the Bacterial Large Ribosomal Subunit. Cell 2016 12/1;167(6):1610-1622.e15.

(40) Blattner F, Plunkett G, Bloch C, Perna N, Burland V, Riley M, ColladoVides J, Glasner J, Rode C, Mayhew G, Gregor J, Davis N, Kirkpatrick H, Goeden M, Rose D, Mau B, Shao Y. The complete genome sequence of Escherichia coli K-12. Science 1997 SEP 5;277(5331):1453-&.

(41) McNaughton BR, Cronican JJ, Thompson DB, Liu DR. Mammalian cell penetration, siRNA transfection, and DNA transfection by supercharged proteins. Proc Natl Acad Sci U S A 2009 APR 14;106(15):6111-6116.

(42) Thompson DB, Villasenor R, Dorr BM, Zerial M, Liu DR. Cellular Uptake Mechanisms and Endosomal Trafficking of Supercharged Proteins. Chem Biol 2012 JUL 27;19(7):831-843.

(43) Linares DM, Kok J, Poolman B. Genome Sequences of Lactococcus lactis MG1363 (Revised) and NZ9000 and Comparative Physiological Studies. J Bacteriol 2010 NOV;192(21):5806-5812.

(44) Kuipers O, de Ruyter P, Kleerebezem M, de Vos W. Quorum sensing-controlled gene expression in lactic acid bacteria. J Biotechnol 1998 SEP 17;64(1):15-21.

(45) Strillinger E, Groetzinger SW, Allers T, Eppinger J, Weuster-Botz D. Production of halophilic proteins using Haloferax volcanii H1895 in a stirred-tank bioreactor. Appl Microbiol Biotechnol 2016 FEB;100(3):1183-1195.

(46) Brendel J, Stoll B, Lange SJ, Sharma K, Lenz C, Stachler A, Maier L, Richter H, Nickel L, Schmitz RA, Randau L, Allers T, Urlaub H, Backofen R, Marchfelder A. A Complex of Cas Proteins 5, 6, and 7 Is Required for the Biogenesis and Stability of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-derived RNAs (crRNAs) in Haloferax volcanii. J Biol Chem 2014 MAR 7;289(10):7164-7177.

(47) Holmes M, Kamekura M, Lam W, Nuttall S, Woods W, Jablonski P, Serrano J, Ngui K, Antón J, Allers T. The Halohandbook, Vol. 7. ; 2008.

(48) Neidhardt FC, Bloch PL, Smith DF. Culture medium for enterobacteria. J Bacteriol 1974 Sep;119(3):736-747.

(49) Elowitz MB, Surette MG, Wolf PE, Stock JB, Leibler S. Protein mobility in the cytoplasm of Escherichia coli. J Bacteriol 1999 JAN 1999;181(1):197-203.

(50) Taheri-Araghi S, Bradde S, Sauls JT, Hill NS, Levin PA, Paulsson J, Vergassola M, Jun S. Cell-size control and homeostasis in bacteria. Current Biology 2015;25(3):385-391.

(51) Maki Y, Yoshida H, Wada A. Two proteins, YfiA and YhbH, associated with resting ribosomes in stationary phase Escherichia coli. Genes to cells 2000;5(12):965-974.

(52) Puri P, Eckhardt TH, Franken LE, Fusetti F, Stuart MC, Boekema EJ, Kuipers OP, Kok J, Poolman B. Lactococcus lactis YfiA is necessary and sufficient for ribosome dimerization. Mol Microbiol 2014;91(2):394-407.

(53) Bateman A, Martin MJ, O'Donovan C, Magrane M, Apweiler R, Alpi E, Antunes R, Arganiska J, Bely B, Bingley M, Bonilla C, Britto R, Bursteinas B, Chavali G, Cibrian-Uhalte E, Da Silva A, De Giorgi M, Dogan T, Fazzini F, Gane P, Cas-tro LG, Garmiri P, Hatton-Ellis E, Hieta R, Huntley R, Legge D, Liu W, Luo J, MacDougall A, Mutowo P, Nightin-gale A, Orchard S, Pichler K, Poggioli D, Pundir S, Pureza L, Qi G, Rosanoff S, Saidi R, Sawford T, Shypitsyna A, Turner E, Volynkin V, Wardell T, Watkins X, Zellner H, Cowley A, Figueira L, Li W, McWilliam H, Lopez R, Xenarios I, Bougueleret L, Bridge A, Poux S, Redaschi N, Aimo L, Argoud-Puy G, Auchincloss A, Axelsen K, Bansal P, Baratin D, Blatter M, Boeckmann B, Bolleman J, Boutet E, Breuza L, Casal-Casas C, De Castro E, Coudert E, Cuche B, Doche M, Dornevil D, Duvaud S, Estreicher A, Famiglietti L, Feuermann M, Gasteiger E, Gehant S, Gerritsen V, Gos A, Gruaz-Gumowski N, Hinz U, Hulo C, Jungo F, Keller G, Lara V, Lemercier P, Lieberherr D, Lombardot T, Martin X, Masson P, Morgat A, Neto T, Nouspikel N, Paesano S, Pedruzzi I, Pilbout S, Pozzato M, Pruess M, Rivoire C, Roechert B, Schneider M, Sigrist C, Sonesson K, Staehli S, Stutz A, Sundaram S, Tognolli M, Verbregue L, Veuthey A, Wu CH, Arighi CN, Arminski L, Chen C, Chen Y, Garavelli JS, Huang H, Laiho K, McGarvey P, Natale DA, Suzek BE, Vinayaka CR, Wang Q, Wang Y, Yeh L, Yerramalla MS, Zhang J, UniProt Consortium. UniProt: a hub for protein information. Nucleic Acids Res 2015 JAN 28;43(D1):D204-D212.

(54) Kozlowski LP. IPC - Isoelectric Point Calculator. Biology Direct 2016 OCT 21;11:55.

(55) Schmidt A, Kochanowski K, Vedelaar S, Ahrne E, Volkmer B, Callipo L, Knoops K, Bauer M, Aebersold R, Heinemann M. The quantitative and condition-dependent Escherichia coli proteome. Nat Biotechnol 2016 JAN;34(1):104-110.

(56) Filoti DI, Shire SJ, Yadav S, Laue TM. Comparative Study of Analytical Techniques for Determining Protein Charge. J Pharm Sci 2015 JUL;104(7):2123-2131.

(57) Gokarn YR, Fesinmeyer RM, Saluja A, Razinkov V, Chase SF, Laue TM, Brems DN. Effective charge measurements reveal selective and preferential accumulation of anions, but not cations, at the protein surface in dilute salt solutions. Protein Science 2011 MAR;20(3):580-587.

Chapter 4: Introduction to membrane protein production

Descending into the abyss

A single cell finds its way into a somewhat isolated and uninhabited solution that is full of nutrients and not yet spoiled with waste. If you could monitor the molecular composition of this microscopic lagoon, you would notice that some types of molecules decrease in number and others increase. What's more, the rate of change of molecule numbers increases, and increases proportionally to the number of cells. What we are witnessing here is a transformation in the character of the solution. We would say, in our (justified) bias towards the cells, that some kind of production process is going on (i.e. in contrast to decay). A production process whose dynamics can be described simply by an exponential function that relates initial cell number, N_0, to later ones, N_t; and some parameter, k, to the rate of change (also see Figure 1):

$$N_t = N_0 e^{kt}$$

The nature of this function, by the way, is not in any way dependent on the molecular constitution of cells or the production processes within them. It is simply a logical consequence of the fact that the rate of production of the material is linearly proportional to the amount of material there already is, and as such is a fundamental property of the world. The value of the parameter k, on the other hand, is dependent on what the cells are made up of and possibly on the nature of the production process. Of course at some stage the nutrients will run out and the rate of multiplication will go down. This can be described by the following equation (Nowak, 2006), which has an s-shape (Figure 1):

$$N_t = \frac{K N_0 e^{kt}}{K + N_0(e^{kt} - 1)}$$

Here K is the maximum number of cells, the carrying capacity.

Let's consider that lonesome cell again. Once more it has found itself in a microscopic lagoon and is happily multiplying. All of a sudden its surroundings change, the microscopic lagoon fused with some other lagoon; or perhaps, in a somewhat slower change, the cells have reached the limit of the lagoon and the resources are getting scarce. Whatever the source of the disturbance, it has set in motion another transformation; this time of the cells themselves. Perhaps the cells become smaller (Milo *et al.*, 2016a), or filamentous (Young, 2006) or they develop the ability to swim around (Osterman *et al.*, 2015). Again we have to conclude that some form of production process may be at work here. What exactly is the role of the production process? Besides the obvious fact that the production process makes things, it almost certainly also plays a role in the rate of change and may also set boundaries on what kind of changes can happen.

Taking cells apart shows that there are various classes of constituents, that all need some kind of production process: proteins, DNA, RNA, lipids, sugar polymers, small molecules, ions, and others. Here I will focus on the protein production process because this is the introduction to chapter 5 which describes a study into protein production. It should however be kept in mind that protein production is a good place to start anyway as it is the linchpin of (pretty much) all production processes in the cell. The growth, multiplication, and transformations of cells described above are a function of, among other things, the production of proteins.
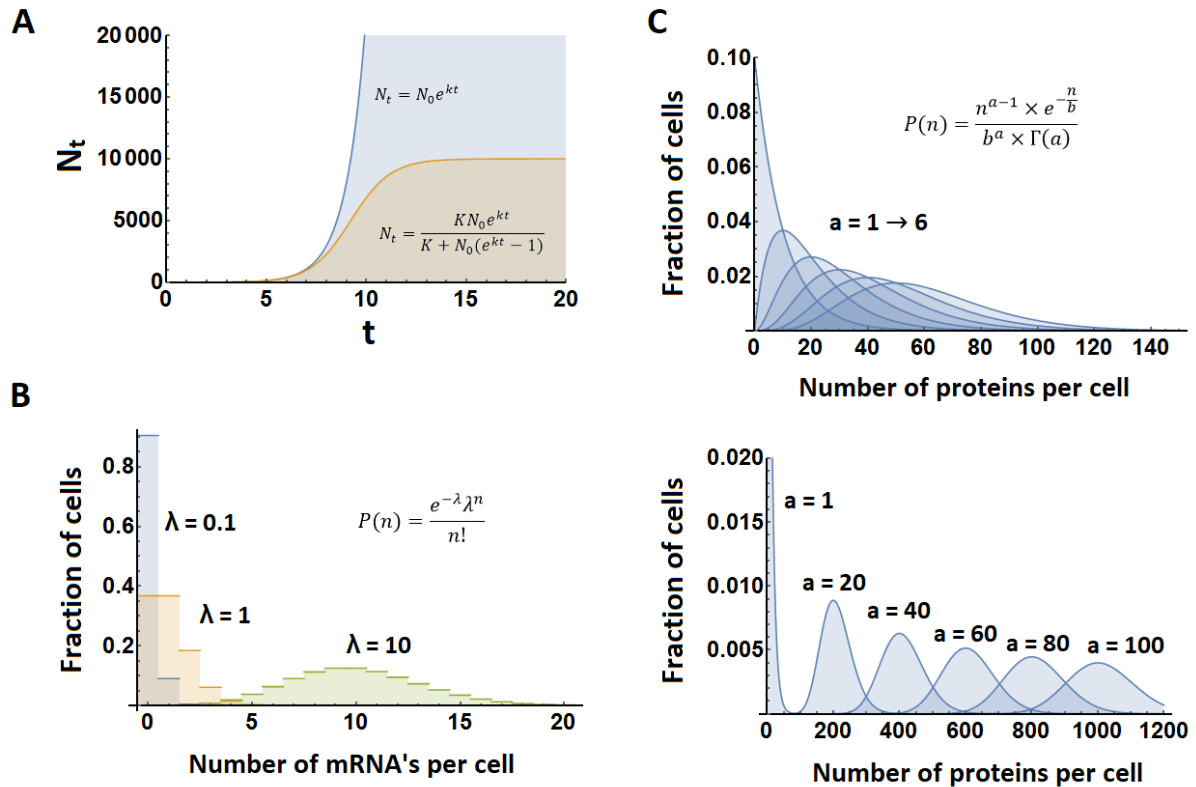
Figure 1: Illustration of equations. A) Exponential growth with and without carrying capacity. Carrying capacity, K, is 10 000, $N_0 = 1$, and $k = 1$ for both curves. The unit of time is arbitrary. B) Variation of mRNA number, n, in a population of cells, assuming that the production of mRNA's can be described by a Poisson process. λ is the mean number of mRNA's per cell. C) Variation of protein number, n, in a population of cells. Generated with the Gamma distribution. The number of mRNA's made in a cell cycle, a, is varied and the values are indicated in the plots. The mean number of proteins made per mRNA, b, is 10.

Monitoring the *total* amount of protein, e.g. by tallying the number of amino acids that are present in proteins, averaged over all cells and over a period of time that encompasses a cell cycle will yield the same dynamics as monitoring the cells as a whole. However, monitoring the total amount of protein at time scales smaller than that of the cell cycle, and doing so while synchronising the cell cycles of all cells in our mind's eye, may show ups and downs in the total protein content. Taking it one step further we can look at the protein content of individual cells and notice the differences between them, if they exist. These difference may be due to local differences of the environment the cells are in or may be a consequence of chance events at the molecular level. Global levels of protein in cells have a possible impact on many processes in the cell due to the phenomenon of excluded volume which can change equilibrium and rate constants of reactions, and diffusion coefficients of molecules (Dix *et al.*, 2008; Zhou *et al.*, 2008). Thus we have a coupling between the phenomenon of protein production and many other processes in the cell, which is independent of the production of the proteins involved directly in those processes. This is all in reference to cells that are simply multiplying, when cells transform this could of course also have an impact on total protein levels.

Arguably more interesting than the total amount of protein is the amount of individual protein types, as the goings on of the panoply of processes happening in the cell depend upon the nature of the individual protein types. Averaging the amounts of proteins over all cells and over the time of a single cell cycle will yield the same dynamics as the total protein content and cell growth. However, now we can compare the levels of different proteins. These differences in amounts immediately betray

something about the function of the proteins and about what it takes to make these proteins and maintain said functions (e.g. regulatory proteins don't need high concentrations). Zooming in again to the level of individual cells we see how differences in character of cells depends on differences in amounts of protein in these same cells (Rosenfeld *et al.*, 2005). Here also we can look at cells that simply multiply and cells that transform, the latter being the more interesting ones because transformations can happen in an almost endless number of ways.

Proteins, of course, are produced from a template of mRNA, and as such the rate of production depends on mRNA concentration. There is also removal of proteins due to cutting the proteins apart or diluting them out. Averaging over all cells in a population, and ignoring spatial heterogeneity within cells, the rate of change of protein concentration is given by:

$$\frac{d[protein]}{dt} = A[mRNA] - B[protein]$$

Here A depends on the strength of the ribosome binding site (among other things), and B is due to degradation, dilution, or a combination of the two. Setting the rate of change of protein concentration to 0 you can calculate the steady state concentration of protein:

$$[protein] = \frac{A[mRNA]}{B}$$

The givens in this equation, A, B, and $[mRNA]$, are by no means a given when dealing with real cells in real circumstances because their values depend on these circumstances, and number of circumstances is endless. Another way in which matters get more complicated is that cells are small, so small in fact that significant differences between cells in say mRNA concentration can occur by chance. Let's see what this means for protein copy numbers. We assume that mRNA's are produced in a Poissonian fashion, i.e. producing mRNA with a probability that is constant over time and as such mRNA production events do not influence one another. We also assume that the number of proteins that is produced from a single mRNA is exponentially distributed. Finally we assume that cells are growing exponentially. Then the distribution of protein copy numbers/concentrations is given by the gamma distribution (see Figure 1C) (Taniguchi *et al.*, 2010):

$$P(n) = \frac{n^{a-1} \times e^{-\frac{n}{b}}}{b^a \times \Gamma(a)}$$

Here $P(n)$ is the probability for a cell to have n proteins, a is the mean number of mRNA's produced per cell cycle, b is the mean number of proteins produced per mRNA, and $\Gamma(a)$ is the gamma function of a. The protein concentration can be obtained by dividing the protein number by the product of the volume of the cell and Avogadro's number. The gamma distribution was tested in *E. coli* cells whose protein numbers were determined after labelling the proteins with a fluorescent tag (Taniguchi *et al.*, 2010). For 1009 out of the 1018 genes tested the protein copy number distribution followed a gamma distribution. However, at larger mean copy numbers the relative variability of copy numbers over the population reached a lower limit at about 30 % (Taniguchi *et al.*, 2010).

In the previous discussion we have treated individual cells as if they are dots, without three dimensional shape or size. However, we know of a number of examples were the production of proteins is not homogeneous over space: there is no protein production in the eukaryotic nucleus, meter long neurons need to be produced and maintain their constitution from a single nucleus, the ribosomes in *E. coli* are excluded from the nucleoid under some conditions (Bakshi *et al.*, 2012), some RNA degradation proteins in *E. coli* are localised to the membrane (Moffitt *et al.*, 2016), genes in *E.*

coli have specific locations in the cell (Toro *et al.*, 2010), and different proteins that are part of the same complex can be produced from the same mRNA and thus start of in close proximity (Shieh *et al.*, 2015). In this last example there need not be spatial variation of the complex in the cell but producing the protein that make up the complex close together could reduce misfolding and aggregation. To take into account spatial variation you can include diffusion terms (Phillips *et al.*, 2009) into the differential equation shown above. A more detailed way is to treat every mRNA, protein, ribosome, and degradation machinery as a diffusing dot which can undergo reactions (Andrews *et al.*, 2010). Doing it this way also makes it easy to include barriers to diffusion such as a nucleus, organelles, or atypical cell shapes.

Monitor a single gene and every once in a while you will witness the arrival of an RNA polymerase. It will produce an mRNA that then will go on to be translated. The arrival of the polymerase and the subsequent production of mRNA may work like a Poisson process, with an exponential distribution in the time between mRNA production events and with the number of mRNA's being produced in a fixed time window following a Poisson distribution (see Figure 1B):

$$P(n) = \frac{e^{-\lambda}\lambda^n}{n!}$$

Here $P(n)$ is the probability that n mRNA's are produced in a fixed window of time, and λ is the mean number of mRNA's produced in this same time window. For this to hold the number of (free) RNA polymerases needs to be constant over time, which gives the constant transcription initiation probability. However, in at least some cases mRNA's are produced in bursts, for example due to the influence of DNA supercoiling on transcription (Chong *et al.*, 2014). If this is the case spread in number of mRNA's produced in a fixed time window will be broader than the Poisson distribution predicts (Li *et al.*, 2011).

Monitor a single mRNA and every once in a while you will witness the arrival of a ribosome. The ribosome will produce a single protein and leave again. If the lifetime of an mRNA was a single fixed time the distribution of proteins produced from a single mRNA would also follow a Poisson distribution. However, it turns out that the lifetime of mRNA's, at least in *E. coli*, is exponentially distributed (Chen *et al.*, 2015). This leads to an exponential distribution also for the number of proteins produced per mRNA (Li *et al.*, 2011). You expect the time between protein production events from the same mRNA to be distributed exponentially as well because there is a constant number of (free) ribosomes. Differences in the localisation of mRNA could influence how many proteins are produced from it. For example if an mRNA would remain, by chance, in the nucleoid region it will encounter less ribosomes, and be translated less often before finally being degraded. Additionally different copies of the same type of mRNA could harbour a different complement of RNA binding proteins, which could also affect the protein production rate. Studies in eukaryotes have shown that not all mRNA's behave in the same way (Yan *et al.*, 2016).

Monitor a single ribosome and you will witness the coming into being of a single protein. You will see the ribosome assembling the protein chain one amino acid at a time (Wen *et al.*, 2008). Occasionally, when encountering a particular grouping of codons, or mRNA structure, the ribosome will move more slowly (Wen *et al.*, 2008; Gardin *et al.*, 2014). On what are probably rare occasions there is a problem with the mRNA and translation has to be aborted with the tmRNA system (Janssen *et al.*, 2012). It will take the ribosome a certain amount of time to finish the single protein. The distribution of these times is likely to be Gaussian. The reason for this is that the protein is made in hundreds of steps each with some distribution associated with it. According to the central limit theorem it doesn't matter what the

distributions are for the steps because as long as you have many steps contributing to one final distribution this distribution will be Gaussian (see also (Li *et al.*, 2011)).

For every (membrane) protein chain that is produced and inserted into the membrane there is a series of coordinated movements of atoms. These atoms are organized in two main clusters, the ribosome and the Sec translocon. There are many additional factors involved but we'll ignore them in this discussion (but see the next section). Amino acids are brought into the ribosome covalently bound to tRNAs. Geometrical constraints in the ribosome create pockets that allow for initial binding of amino acid-tRNA's, checking whether the amino acid-tRNA matches the mRNA, peptide bond formation, and guiding of the amino acid chain out of the ribosome. A directed movement of the geometry is what makes everything happen, and, with the input of energy, makes it happen in one direction, towards amino acid chain formation (Moore, 2012). For (plasma) membrane proteins the insertion into the membrane happens in tandem with the synthesis of the amino acid chain. The Sec translocon, a channel which has a central pore through the membrane, can accommodate up to two transmembrane helices. These are released sideways by opening the Sec translocon at one of its sides. This can be done periodically for membrane proteins with many transmembrane helices (Driessen *et al.*, 2008). For both the ribosome and the Sec translocon their conformational rearrangements set limits to the rate of protein production.

The cast, the play and a note about their origins

Here I'll give a short overview of the different macromolecules that play a role in protein synthesis and membrane insertion. In a sense all elements of a cell are part of the production process because the reason for a cell's existence is to make more cells. However, I will focus on the synthesis and membrane insertion of proteins in *E. coli*. Let's start at the production of a mRNA. As soon as the ribosome binding site is made a ribosome can bind and start synthesizing a protein, this is before the mRNA has been fully synthesized. The ribosome binding site on the mRNA is bound by a complex of the small ribosomal subunit (30S), formylmethionine-tRNA, and initiation factors IF1 and IF3. (The initiation factor IF2 helps to deliver formylmethionine-tRNA to the small ribosomal subunit.) Next, the large ribosomal subunit (50S) binds (Marintchev *et al.*, 2004). Starting at the formylmethionine, amino acids are added to form an amino acid chain in cycles of conformational changes of the ribosome and assisted by the elongation factors EF-Tu and EF-G (Voorhees *et al.*, 2013). The specific amino acids come attached to specific tRNA's, to which they are attached by specialized amino acyl tRNA synthetases (Woese *et al.*, 2000). When the amino acid chain is done, signalled by a stop codon, the release and recycling factors, RF1, RF2, RF3, and RRF, come in to release the nascent chain and dissociate the small and large ribosomal subunits (Marintchev *et al.*, 2004).

For membrane proteins there is a part of the amino acid sequence that, as soon as it finds its way out of the ribosome, binds to a protein-RNA complex called the signal recognition particle (SRP). The signal recognition particle then guides the membrane protein-ribosome complex via the peripheral membrane protein FtsY to the Sec translocon. The Sec translocon, also called SecYEG, is composed of the proteins SecY, SecE, and SecG, and facilitates translocation of the membrane protein into the membrane. The part of the membrane protein that is first to emerge out of the ribosome can be inserted before the rest of the protein has been made (Driessen *et al.*, 2008).

The functioning of the ribosome, Sec translocon, and their associated factors depends on basic properties of molecules: diffusion, conformational changes, binding, and chemical reactions. This is true for all molecular phenomena that go on in the cell. Yet, we speak of a ribosome and Sec translocon, or translation and insertion, as if they are completely detached from the rest of the (macro)molecular phenomena. We do know similarities between the ribosome, and DNA- and RNA

polymerases (Bai *et al.*, 2006; van Oijen *et al.*, 2010), all three of them are template based polymerases. There are also clear similarities between the Sec translocon and the twin arginine translocase (Goosens *et al.*, 2014): both complexes transport proteins. I finally I note that the Sec translocon and DNA helicases (Patel *et al.*, 2000) are similar in that they both let polymers thread through them. Perhaps someday we can have a description of translation and insertion that makes these processes continuous with other now existing, once having existed, and perhaps someday to be created processes. Such a unified view would also shed light on the evolution of the phenomena of translation and insertion.

Protein fate maps and extension of cellular possibilities

In *E. coli* cells the processes of translation and insertion do not produce and insert proteins in isolation. There is a plethora of proteins (chaperones) that assures that proteins are correctly, and efficiently, folded. For example the proteins trigger factor, DnaJ, DnaK, GrpE, SurA and the GroEL/ES protein complex (Kim *et al.*, 2013). Trigger factor is associated with ribosomes and interacts with most proteins when they are being synthesized, preventing hydrophobic patches from interacting. DnaJ, DnaK, and GrpE facilitate protein folding in an ATP-dependent manner. SurA isomerizes prolines in proteins so that the proteins can fold properly. Finally, the GroEL/ES is a big complex that traps proteins inside of a big cavity and assists in their folding. Proteins also get degraded by proteins called proteases. For example in response to a change in environment that needs a change in the set of proteins in the cell. In *E. coli* there are the ClpAP, ClpXP, Lon, and HslUV proteases. These proteases specifically degrade a (broad) subset of proteins (Gur *et al.*, 2011).

For membrane proteins there are also various factors that help protein production along. In *E. coli* we have SecD, SecF, YajC, YidC, YidD, and FtsH. The SecDF(YajC) complex assists in the insertion of membrane proteins. YidC assists SecYEG but can also insert some membrane proteins by itself (Driessen *et al.*, 2008). YidD has been associated with the insertion of YidC dependent substrates (Yu *et al.*, 2011). FtsH degrades misassembled membrane proteins (Ito *et al.*, 2005).

To understand the life of a protein you have to take all the factors just mentioned into account. Doing this, you would end up with a protein fate map. This fate map contains everything that could happen to a protein and with what probability. And this is not limited to chaperoning or degradation. In filling in this fate map you would ask the following questions. What fraction of all amino acid chains whose synthesis is started, i.e. two amino acids coupled together, actually reach a full chain? What fraction of full amino acid chains fold up properly? What fraction of folded proteins carries out what number of actions (e.g. isomerisation reactions) during its lifetime? With what frequency does an amino acid chain (completely synthesized or not) encounter a chaperone? What fraction of those interactions is consequential? With what frequency does a protein (in folded or unfolded state) encounter each of the other proteins in the cell? With what frequency does a protein encounter the membrane? With what frequency does the protein unfold? How many cell cycles does a protein last? How many minutes does a protein last? What fraction of cell volume does the protein sample per unit time? Answers to these questions help in understanding what would happen to a protein when some of the cellular parameters change. For example if the concentration of the chaperone GroEL/ES changes, or if the diffusion rate becomes slower after an osmotic shock.

The key questions that occupies me when thinking about chaperones and proteases is this. What higher level functions of cells would not be possible without such systems? Let me warn you about interpreting this question too naively. I do not mean what systems will be crippled when you do a knockout of these systems in *E. coli* (or some other organism). Because in principle the cell could be arranged in a different manner which would avoid the need for chaperoning. For instance you could

adjust the amino acid sequence to make proteins better able to fold so they can do so without assistance. Or, you could use a molecular system that has the same effect but works by a different mechanism that is affected less by folding problems. Could we classify cellular functions by how vulnerable they are to failing when the chaperone and degradation functions are dialled down in efficacy?

Protein production numbers for *Escherichia coli*

Protein production is automatically associated with questions of how many? And, how fast? How many proteins are there in a single *E. coli* cell? For a cell with a volume of 1 µm^3 there are about 2.7 x 10^6 proteins (see Table 1). All these proteins need to be made by the complement of ribosomes present in the cell. There are about 30 000 ribosomes in a cell with a volume of 1 µm^3 (Vendeville *et al.*, 2011). Each cell cycle these ribosomes thus have to produce ~100 proteins each. All this is with a cell division time of 40 min (Vendeville *et al.*, 2011; Milo *et al.*, 2016b). How fast can a ribosome produce 100 proteins? The mean protein length is about 300 amino acids (Table 1). (This is based on the number of genes that code for a protein of certain length. It does not take into account abundance.) The translation rate is 10-20 amino acids/s. This gives a translation time of 15-30 s per protein. Multiplying by the total number of proteins gives a total time of 25-50 min. This matches, albeit coarsely, with the cell division time. Each of the numbers mentioned here, whether measured or calculated, can be used to test hypotheses. But there is more to the calculations. They validate the numbers by allowing cross-referencing.

Note that because the number of proteins that a ribosome can produce per unit time is limited, the number of ribosomes regulates the total number of proteins. This means that if there were a cellular crowding sensor embedded in the regulation of ribosome production this would allow for regulation of the total number of proteins per cell. A first step in regulating the numbers of individual protein types is by simply having ribosome binding sites of different strengths, whose competition for binding ribosomes determines the protein numbers.

Following their synthesis membrane proteins have to be inserted in the membrane. This provides another constraint on the number of proteins that are produced per cell cycle, because this depends on the number of SecYEG channels and the rate with which they translocate proteins. The PaxDb contains a collection of abundance data for *E. coli* and shows that there are great differences between the measurements of a single protein. For example the numbers for SecY range from 3-588 parts per million (ppm). A dataset that integrates all the data gives 97 ppm for SecY. This should be compared to 35 ppm for SecE and 183 for SecG, both of which reside with SecY in the same complex. Numbers for these and other players in membrane protein production are listed in Table 1, all based on the integrated dataset from PaxDb. Let's assume that the translocon numbers are the same as the SecY numbers reported in the PaxDb dataset, 100 ppm. For a division time of 40 min, when the cell volume is 1 µm^3, the number of translocons is 300. Let's assume that the time the translocon is occupied is directly related to the number of transmembrane helices present in a membrane protein. Typically there are about 3 transmembrane helices per membrane protein ((Linden *et al.*, 2012), data from *Acinetobacter baumannii*). Each transmembrane helix contains about 20 amino acids, and assuming that loops account for another 40 amino acids we end up with a total of 100 amino acids per membrane protein. Let's also assume that the insertion machinery works as fast as the ribosome, 10-20 amino acids/s. This means that it takes 5-10 s for a translocon to insert a single membrane protein. Each translocon can insert 240-480 proteins per 40 min cell cycle. The 300 translocons together can insert a total of 72000-144000 membrane proteins. Taking the area per membrane protein to be 4.5 nm^2 (Linden *et al.*, 2012), and the surface area of a spherocylinder with diameter 0.8 µm and length

2.2 μm (Vendeville *et al.*, 2011), the membrane fraction covered with membrane protein is 6-12 %. This is somewhat lower than estimates reported elsewhere on various types of membranes >20 % (Linden *et al.*, 2012). Given that the translocon is also responsible for transport of periplasmic, outer membrane, and extracellular proteins the number of 300 copies per cell seems too low. So, unlike the previous calculation, here we seem to have found a contradiction. This calculation therefore serves more as a call for more, and more accurate, numbers than as the final answer on the matter.

Table 1: Protein production numbers for *Escherichia coli*.

	Numbers	References and comments
Number of genes	4318, 4505	KEGG database (strain MG1655), (Keseler *et al.*, 2017) (strain MG1655)
Number of genes encoding proteins	4140	KEGG database (strain MG1655)
Number of genes encoding membrane proteins	~900	Helix bundle membrane proteins, (Krogh *et al.*, 2001)
Mean protein length	318 amino acids	Calculated from data from the KEGG database (strain MG1655), the mean is over genes and does not include protein copy numbers
Ribosome copy number per cell	6800-72000, 55000	Depends on cell growth rate (Vendeville *et al.*, 2011), (Bakshi *et al.*, 2012)
RNA polymerase copy number per cell	1500-11000, 4600	Depends on cell growth rate (Vendeville *et al.*, 2011), (Bakshi *et al.*, 2012)
Number of proteins per μm³ cell volume	2.7×10^6	(Milo, 2013)
Cell volume	0.4-2.5 μm³	(Milo *et al.*, 2016a)
Protein chain elongation rate	10-20 amino acids/s	(Milo *et al.*, 2016c)
RNA chain elongation rate (mean)	25 nt/s	(Chen *et al.*, 2015)
Average mRNA synthesis time	133 s	(Chen *et al.*, 2015)
Average mRNA lifetime	4.09 min	Different from value mentioned in the paper but calculated from the same data, (Chen *et al.*, 2015)
Average mRNA length	1286 nucleotides	(Chen *et al.*, 2015)
mRNA diffusion coefficient	0.04 μm²/s	When bound to ribosome, (Bakshi *et al.*, 2012)
Ribosome diffusion coefficient	0.04 μm²/s	(Bakshi *et al.*, 2012)
SecY copy number	97 ppm	PaxDb integrated, (Wang *et al.*, 2012)
SecE copy number	35 ppm	PaxDb integrated
SecG copy number	183 ppm	PaxDb integrated
YidC copy number	70 ppm	PaxDb integrated

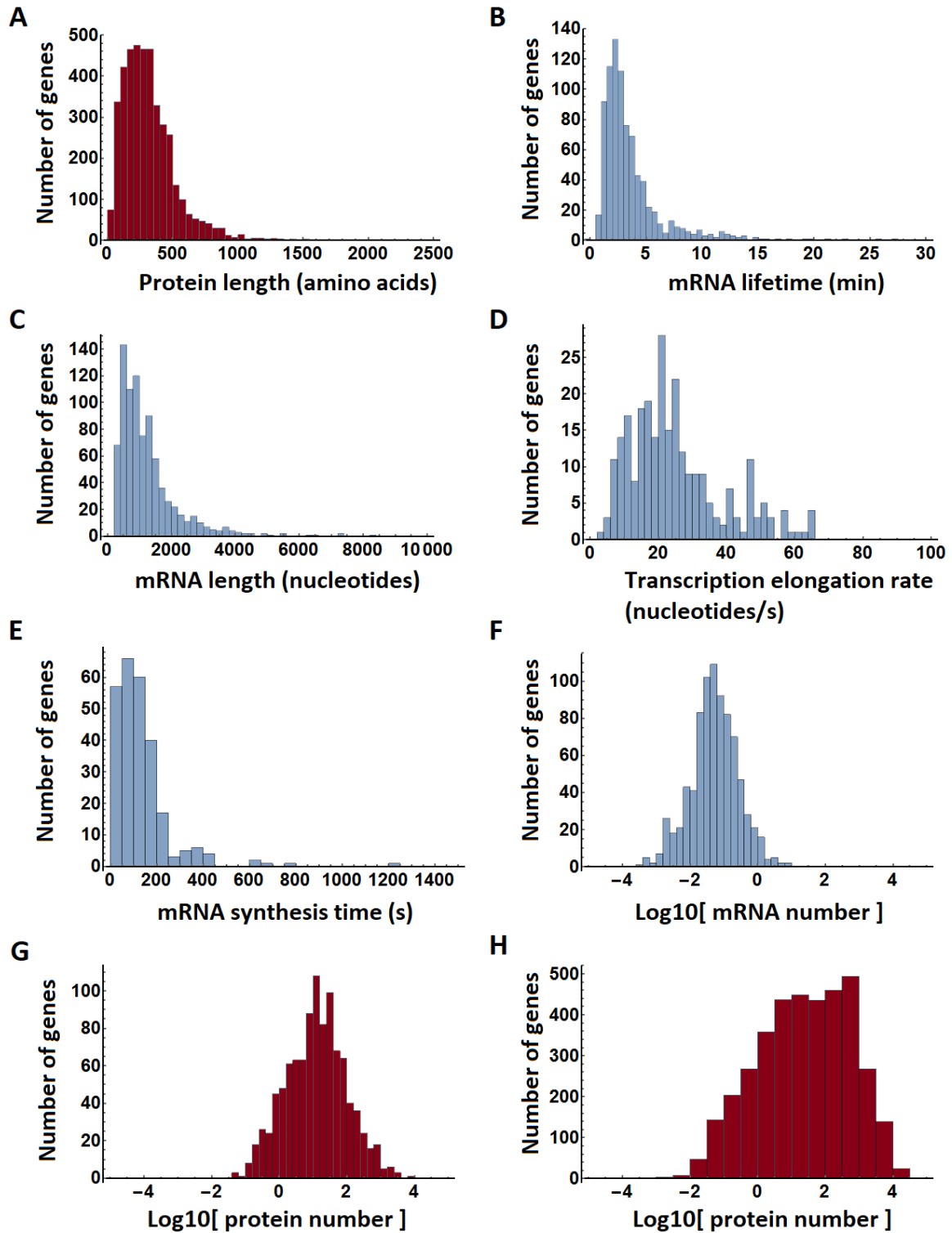| Ffh (signal recognition particle subunit) copy number	173 ppm	PaxDb integrated
FtsY copy number	215 ppm	PaxDb integrated
SecA copy number	255 ppm	PaxDb integrated

Figure 2: Genome wide parameters for protein production in *Escherichia coli*. Parameters for proteins in red, parameters for mRNA in blue. A) Protein length distribution. Taken from the KEGG database entry on *E. coli*

MG1655. B) mRNA lifetime distribution (Chen *et al.*, 2015). C) mRNA length distribution. (Chen *et al.*, 2015). D) Transcription elongation rate distribution (for mRNA) (Chen *et al.*, 2015). E) mRNA synthesis time distribution. Calculated from mRNA length and transcription elongation rate (Chen *et al.*, 2015). F) Distribution of mRNA copy number per cell (Taniguchi *et al.*, 2010). G) Distribution of protein copy number per cell, about 1000 protein are represented (Taniguchi *et al.*, 2010). H) Distribution of protein copy number per cell. Combination of datasets from PaxDb, with about 4000 proteins (Wang *et al.*, 2012). Note that the PaxDb numbers were multiplied by 3 because they were reported as parts per million and the number of *E. coli* proteins is about 3 million (see Table 1).

For the examples worked out above there was a heavy reliance on average numbers. There are however differences between genes, and their products mRNA and proteins. Numbers have been measured, or could be derived from measurements, for protein and mRNA length, mRNA lifetime, transcription elongation rate, mRNA synthesis time, and mRNA and protein copy number per cell. Histograms can be found in Figure 2. There are interesting observations in here that cry out for further consideration. For example in the distributions of mRNA lifetime, transcription elongation rate, and mRNA synthesis time there are some considerable outliers. It would be interesting to know how those facts tie into the production and functioning of these proteins.

In my consideration of numbers for protein production in *E. coli* I have not been entirely systematic. I have not scoured the literature for every number, nor have I considered for each number the exact condition and strain the number has been determined in. That would be a Herculean task. At some point in the future someone will have to do that, but here I sought to give you, and myself, a rough outline to serve as a platform from which to work. We seem to be getting to a point were numbers are known for most aspects of protein production. Now these numbers must be consolidated. To do this we need to continue improving experimental methods and consistency in growth conditions. We also need to link all numbers together by calculation. When you calculate (and measure) a number in a couple of different ways, and there is an inconsistency, then you have found your next problem to focus on. We report here only on numbers in *E. coli*. Many interesting things will likely emerge from a comparison of the *E. coli* numbers to those of other organisms.

Making decisions

At the start of this chapter I mentioned that cells transform. In many such transformations protein production (and destruction) play an important role. The purpose of this section is to make the distinction between two facets of protein production. One being the formation of the amino acid chain, its folding, and finally its breakdown. The second being the choice of what complement of proteins to have in what environments and what internal states. In *E. coli* the decisions about what protein to made when is a consequence of at least 200 transcription factors (Keseler *et al.*, 2017). There are also RNA based regulators that work after transcription (Nitzan *et al.*, 2017). There is a connection between these two facets of protein production. The rate of change from one cellular state to another is determined by the combination of by what mechanism the choice is made and how the proteins are produced. (Assuming that the states we are talking about are defined by what proteins are around, and not by conformational changes.) If a very fast change of state is required it should depend on changes in protein conformational state rather than changes in protein production. Think for example about the function of MscL upon osmotic shock, which prevents the cell from bursting by responding within a second (Bialecka-Fornal *et al.*, 2015). Knowledge of what type of regulation to use when thus depends in part on the rate of protein production.

Wait. What? There is more?

An outside onlooker into molecular biology may be forgiven for thinking that there is only one organism worth studying on our planet: *E. coli*. Yet for any of the examples of the workings of protein production given in this chapter you could ask: What if it were different? For example in Figure 1C we see the spread of protein copy numbers in different cells in a population. Say we imagine a protein that has an average of 1000 copy numbers in an *E. coli* cell. Its spread in copy number over the population is given in Figure 1C. If that same protein was needed at the same concentration in a *Mycoplasma* cell the relative spread in copy number would be much greater. The reason is that *Mycoplasma* cells have about 20 times less volume than *E. coli* cells (BNID108334, (Milo *et al.*, 2010)). This issue will arise for any protein we consider in *Mycoplasma*, and also for mRNA's. Note that *Mycoplasma* also has much fewer genes than *E. coli*, so if the crowding remained the same it could have a higher protein concentration per gene. Issues also arise when you make cells much bigger than *E. coli*. For example in the extreme case of some nerve cells where a single nucleus has to provide mRNA for the production of proteins that are needed 1 m away from the nucleus. This requires that protein production is appended by a directional transport system. Another example comes from big bacteria. *Epulopiscium fishelsoni* can be 600 μm in length and 80 μm in width (Schulz *et al.*, 2001). However, unlike the nerve cell *E. fishelsoni* has 10^4 copies of its genetic material (Angert, 2012) so it doesn't need the directional transport system. There are big differences in the number of protein partners a protein can encounter within a bacterial cell with the number of protein coding genes ranging from 500 (Fraser *et al.*, 1995) to >10000 (Dagan *et al.*, 2013; Han *et al.*, 2013). If would be interesting to know how this affects the necessity of chaperones. There are also big differences in mRNA lifetime, from a few minutes in *E. coli* (see Figure 2B), to tens of minutes in Yeast (Geisberg *et al.*, 2014), and hours in the Mouse (Schwanhaeusser *et al.*, 2011). The translation rate also varies between organisms with about 3 amino acids/s for Human cells (Yan *et al.*, 2016) and 10-20 amino acids/s for *E. coli* cells (see Table 1).

There are bound to be many more differences between organisms. To understand protein production we must also be able to understand why these differences exist. And also why some differences do not exist. It may be a good exercise to take a source that documents the diversity of organisms, for example the book *Kingdoms and Domains: An Illustrated Guide to the Phyla of Life on Earth* by Margulis and Chapman, and systematically go through it asking for each phenomenon: How is this affected by or how does this affect protein production?

Conclusion

Here I have provided a rough overview of protein production and its associated phenomena. These phenomena have been studied in much greater detail then what is treated in this introduction. I found it more interesting to present some of my own thoughts on the matter rather than providing you with a list of references. I have touched on topics that are somewhat distant from what will be discussed in the next chapter. My first reason for doing this is that it is not always clear at the outset whether two phenomena are connected and a more insightful reader may make a connection that I did not. The second reason is that it is useful to know which things are not related to my subject of study even though they go under the same name of protein production.

The next chapter will be about the development of a method to determine membrane protein production rates at a single molecule level. This will, hopefully, contribute to our knowledge of the numbers that govern protein production, some of which we already discussed earlier in this introduction. Some of the other phenomena discussed may help to interpret our (future) findings with this method. The interaction of chaperones, such as trigger factor, and insertion proteins, such as YidC,

may alter the timing of the events we want to measure. Regulation of protein expression at the level of mRNA may alter the timing between protein production events from that mRNA. We set out to study protein production in two different bacteria, *Lactococcus lactis* and *Escherichia coli*, which may reveal some interesting differences between organisms.

References

Andrews, S.S., Addy, N.J., Brent, R., and Arkin, A.P. (2010) Detailed Simulations of Cell Biology with Smoldyn 2.1. *PLoS Comput Biol* **6:** e1000705.

Angert, E.R. (2012) DNA Replication and Genomic Architecture of Very Large Bacteria. *Annu. Rev. Microbiol.* **66:** 197-212.

Bai, L., Santangelo, T.J., and Wang, M.D. (2006) Single-molecule analysis of RNA polymerase transcription. *Annu Rev Biophys Biomol Struct* **35:** 343-360.

Bakshi, S., Siryaporn, A., Goulian, M., and Weisshaar, J.C. (2012) Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. *Mol Microbiol* **85:** 21-38.

Bialecka-Fornal, M., Lee, H.J., and Phillips, R. (2015) The Rate of Osmotic Downshock Determines the Survival Probability of Bacterial Mechanosensitive Channel Mutants. *J Bacteriol* **197:** 231-237.

Chen, H., Shiroguchi, K., Ge, H., and Xie, X.S. (2015) Genome-wide study of mRNA degradation and transcript elongation in Escherichia coli. *Molecular Systems Biology* **11:** 781.

Chong, S., Chen, C., Ge, H., and Xie, X.S. (2014) Mechanism of Transcriptional Bursting in Bacteria. *Cell* **158:** 314-326.

Dagan, T., Roettger, M., Stucken, K., Landan, G., Koch, R., Major, P., *et al.* (2013) Genomes of Stigonematalean Cyanobacteria (Subsection V) and the Evolution of Oxygenic Photosynthesis from Prokaryotes to Plastids. *Genome Biol. Evol.* **5:** 31-44.

Dix, J.A., and Verkman, A.S. (2008) Crowding effects on diffusion in solutions and cells. *Annu. Rev. Biophys.* **37:** 247-263.

Driessen, A.J.M., and Nouwen, N. (2008) Protein translocation across the bacterial cytoplasmic membrane. *Annu Rev Biochem* **77:** 643-667.

Fraser, C., Gocanye, J., White, O., Adams, M., Clayton, R., Fleischmann, R., *et al.* (1995) The Minimal Gene Complement of Mycoplasma-Genitalium. *Science* **270:** 397-403.

Gardin, J., Yeasmin, R., Yurovsky, A., Cai, Y., Skiena, S., and Futcher, B. (2014) Measurement of average decoding rates of the 61 sense codons in vivo. *eLife* **3:** e03735.

Geisberg, J.V., Moqtaderi, Z., Fan, X., Ozsolak, F., and Struhl, K. (2014) Global Analysis of mRNA Isoform Half-Lives Reveals Stabilizing and Destabilizing Elements in Yeast. *Cell* **156:** 812-824.

Goosens, V.J., Monteferrante, C.G., and van Dijl, J.M. (2014) The Tat system of Gram-positive bacteria. *Biochim. Biophys. Acta* **1843:** 1698-1706.

Gur, E., Biran, D., and Ron, E.Z. (2011) Regulated proteolysis in Gram-negative bacteria - how and when? *Nature Rev. Microbiol.* **9:** 839-848.

Han, K., Li, Z., Peng, R., Zhu, L., Zhou, T., Wang, L., *et al.* (2013) Extraordinary expansion of a Sorangium cellulosum genome from an alkaline milieu. *Scientific Reports* **3:** 2101.

Ito, K., and Akiyama, Y. (2005) Cellular functions, mechanism of action, and regulation of FtsH protease. *Annu Rev Microbiol* **59:** 211-231.

Janssen, B.D., and Hayes, C.S. (2012) The tmRNA ribosome-rescue system. *Adv. in Prot. Chem. and Struc. Biol.* **86:** 151-191.

Keseler, I.M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides-Martinez, C., Caspi, R.*, et al.* (2017) The EcoCyc database: reflecting new knowledge about Escherichia coli K-12*. Nucleic Acids Res* **45:** D543-D550.

Kim, Y.E., Hipp, M.S., Bracher, A., Hayer-Hartl, M., and Hartl, F.U. (2013) Molecular Chaperone Functions in Protein Folding and Proteostasis. *Annu. Rev. Biochem.* **82:** 323-355.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J Mol Biol* **305:** 567-580.

Li, G., and Xie, X.S. (2011) Central dogma at the single-molecule level in living cells*. Nature* **475:** 308-315.

Linden, M., Sens, P., and Phillips, R. (2012) Entropic Tension in Crowded Membranes. *PLoS Comput Biol* **8:** e1002431.

Marintchev, A., and Wagner, G. (2004) Translation initiation: structures, mechanisms and evolution*. Q Rev Biophys* **37:** 197-284.

Milo, R., and Phillips, R. (2016a). In *Cell Biology by the Numbers*. New York, USA: Garland Science, pp. 11.

Milo, R., and Phillips, R. (2016b). In *Cell Biology by the Numbers*. New York, USA: Garland Science, pp. 10.

Milo, R., and Phillips, R. (2016c). In *Cell Biology by the Numbers*. New York, USA: Garland Science, pp. 232.

Milo, R. (2013) What is the total number of protein molecules per cell volume? A call to rethink some published values*. Bioessays* **35:** 1050-1055.

Milo, R., Jorgensen, P., Moran, U., Weber, G., and Springer, M. (2010) BioNumbers-the database of key numbers in molecular and cell biology. *Nucleic Acids Res* **38:** D750-D753.

Moffitt, J.R., Pandey, S., Boettiger, A.N., Wang, S., and Zhuang, X. (2016) Spatial organization shapes the turnover of a bacterial transcriptome. *eLife* **5:** e13065.

Moore, P.B. (2012) How Should We Think About the Ribosome? *Annu. Rev. Biophys.* **41:** 1-19.

Nitzan, M., Rehani, R., and Margalit, H. (2017) Integration of Bacterial Small RNAs in Regulatory Networks. *Annu. Rev. Biophys.* **46:** 131-148.

Nowak, M. (2006) In *Evolutionary Dynamics*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press, pp. 12.

Osterman, I.A., Dikhtyar, Y.Y., Bogdanov, A.A., Dontsova, O.A., and Sergiev, P.V. (2015) Regulation of flagellar gene expression in Bacteria. *Biochemistry (Moscow)* **80:** 1447-1456.

Patel, S., and Picha, K. (2000) Structure and function of hexameric helicases*. Annu Rev Biochem* **69:** 651-697.

Phillips, R., Kondev, J., and Theriot, J. (2009). In *Physical Biology of the Cell*. New York: Garland Science, pp. 481-512.

Rosenfeld, N., Young, J., Alon, U., Swain, P., and Elowitz, M. (2005) Gene regulation at the single-cell level*. Science* **307:** 1962-1965.

110

Schulz, H., and Jorgensen, B. (2001) Big bacteria. *Annu Rev Microbiol* **55:** 105-137.

Schwanhaeusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., *et al.* (2011) Global quantification of mammalian gene expression control. *Nature* **473:** 337-342.

Shieh, Y., Minguez, P., Bork, P., Auburger, J.J., Guilbride, D.L., Kramer, G., and Bukau, B. (2015) Operon structure and cotranslational subunit association direct protein assembly in bacteria. *Science* **350:** 678-680.

Taniguchi, Y., Choi, P.J., Li, G., Chen, H., Babu, M., Hearn, J., *et al.* (2010) Quantifying E-coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science* **329:** 533-538.

Toro, E., and Shapiro, L. (2010) Bacterial Chromosome Organization and Segregation. *Cold Spring Harb Perspect Biol* **2:** a000349.

van Oijen, A.M., and Loparo, J.J. (2010) Single-Molecule Studies of the Replisome. *Annu. Rev. Biophys.* **39:** 429-448.

Vendeville, A., Lariviere, D., and Fourmentin, E. (2011) An inventory of the bacterial macromolecular components and their spatial organization. *FEMS Microbiol Rev* **35:** 395-414.

Voorhees, R.M., and Ramakrishnan, V. (2013) Structural Basis of the Translational Elongation Cycle. *Annu. Rev. Biochem.* **82:** 203-236.

Wang, M., Weiss, M., Simonovic, M., Haertinger, G., Schrimpf, S.P., Hengartner, M.O., and von Mering, C. (2012) PaxDb, a Database of Protein Abundance Averages Across All Three Domains of Life. *Molecular & Cellular Proteomics* **11:** 492-500.

Wen, J., Lancaster, L., Hodges, C., Zeri, A., Yoshimura, S.H., Noller, H.F., *et al.* (2008) Following translation by single ribosomes one codon at a time. *Nature* **452:** 598-603.

Woese, C., Olsen, G., Ibba, M., and Soll, D. (2000) Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol. and Mol. Biol. Rev.* **64:** 202-236.

Yan, X., Hoek, T.A., Vale, R.D., and Tanenbaum, M.E. (2016) Dynamics of Translation of Single mRNA Molecules In Vivo. *Cell* **165:** 976-989.

Young, K.D. (2006) The selective value of bacterial shape. *Microbiol. and Mol. Biol. Rev.* **70:** 660-703.

Yu, Z., Laven, M., Klepsch, M., de Gier, J., Bitter, W., van Ulsen, P., and Luirink, J. (2011) Role for Escherichia coli YidD in Membrane Protein Insertion. *J Bacteriol* **193:** 5242-5251.

Zhou, H., Rivas, G., and Minton, A.P. (2008) Macromolecular crowding and confinement: Biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* **37:** 375-397.

Chapter 5: Determining membrane protein production rates from single mRNAs in *Lactococcus lactis* and *Escherichia coli*

The idea

We want to study the production of membrane proteins in single bacterial cells and want to do so in a manner that allows us to discriminate between individual proteins. We have arrived at an idea that will allow us to this. Here, we describe this idea and our attempts at making it work in the lab.

The production of membrane proteins starts, as is the case for all proteins, at the DNA. First, a DNA sequence is transcribed into an mRNA sequence, which is then translated into a string of amino acids, i.e. a protein. For membrane proteins there are some unique complications to this scheme. The mRNA, with associated ribosome and protein nascent chain, needs to travel to the membrane before the proteins are produced, and released from the ribosome. This is resolved by the fact that as soon as the N-terminus (i.e. the first part) of membrane protein comes poking out of the ribosome, it is recognized by the signal recognition particle (SRP), which shuttles the mRNA-ribosome-nascent chain complex to the membrane. At the cytoplasmic membrane of bacteria the membrane protein is inserted by the protein complex SecYEG. Insertion by SecYEG and synthesis by the ribosome of the membrane protein happens concurrently for a single protein, so that when the first part of the protein is inserted the rest is still being synthesized (Driessen *et al.*, 2008). Furthermore, folding can also happen as soon as part of the protein is inserted into the lipid bilayer. All in all this means that the production of individual membrane proteins include these three processes, synthesis, insertion, and folding, at the same time.

Broadly speaking we are interested in the rate of the production of membrane proteins. Later on we will dive into what parts of the dynamics of this process we want to determine the rates for. By now the standard technique for measuring a rate of a process in a cell is by labeling the protein of interest with a fluorescent protein and monitoring the development of fluorescence. The gene for the membrane protein of interest and the gene for a fluorescent protein are fused so that a fluorescent chimera is produced as a single polypeptide chain. Cells producing this fusion protein can then be observed with a fluorescence microscope. This has indeed been done to follow the production of membrane proteins at the level of individual molecules. Yu *et al*. (Yu *et al.*, 2006) made a fusion of the membrane-localizing chemotaxis protein Tsr and the fluorescent protein Venus, and put this behind a *lac* promoter on the *Escherichia coli* chromosome. They monitored the expression of Tsr-Venus under repressing conditions so that only low levels of mRNA and protein were produced. Cells were illuminated for 1200 ms every three minutes. The first 100 ms were collected for image analysis, the rest was discarded. The long illumination time ensures that the cells are fully photo-bleached so that in each frame only the fluorophores that formed in the previous three minutes were observed. From the time series of fluorescence images Yu et al. extracted the time between transcription events, on average 1.2 events per cell cycle, and the number of proteins produced per mRNA, on average 4.2 proteins per mRNA.

However, for our purpose the method of Yu *et al* is inadequate, as it is limited to a time resolution of minutes. There are two reasons for this: (1) The long laser illumination times cause damage to the cells necessitating a few minutes of "cool down" time between subsequent frames. (2) The fluorescent proteins need on the order of 10 minutes, or more, to become fluorescent (Yu *et al.*, 2006; Shaner *et al.*, 2013; Chu *et al.*, 2014), and do so in a stochastic manner so that the signal is not only slow but also scrambled in time. The times we are interested in are shorter than the resolution provided by this method: the translation of an average size protein in *E. coli* takes about 20 s (BNID 100059, Milo *et al.*, 2010) and the lifetime of mRNA in *E. coli* is typically a couple of minutes (Bernstein *et al.*, 2002).

We are interested in two rates (Figure 1): (i) the rate of production of single protein chains (number 3 in Figure 1), and (ii) the rate of production of multiple protein molecules from a single mRNA (number

1 and 2 in Figure 1). We decided to start with determining rate number 2, the rate of appearance of C-termini. For the rest of this chapter we deal with the rate of appearance of the C-termini of the membrane protein chimera, but we note that this includes insertion and folding as well.
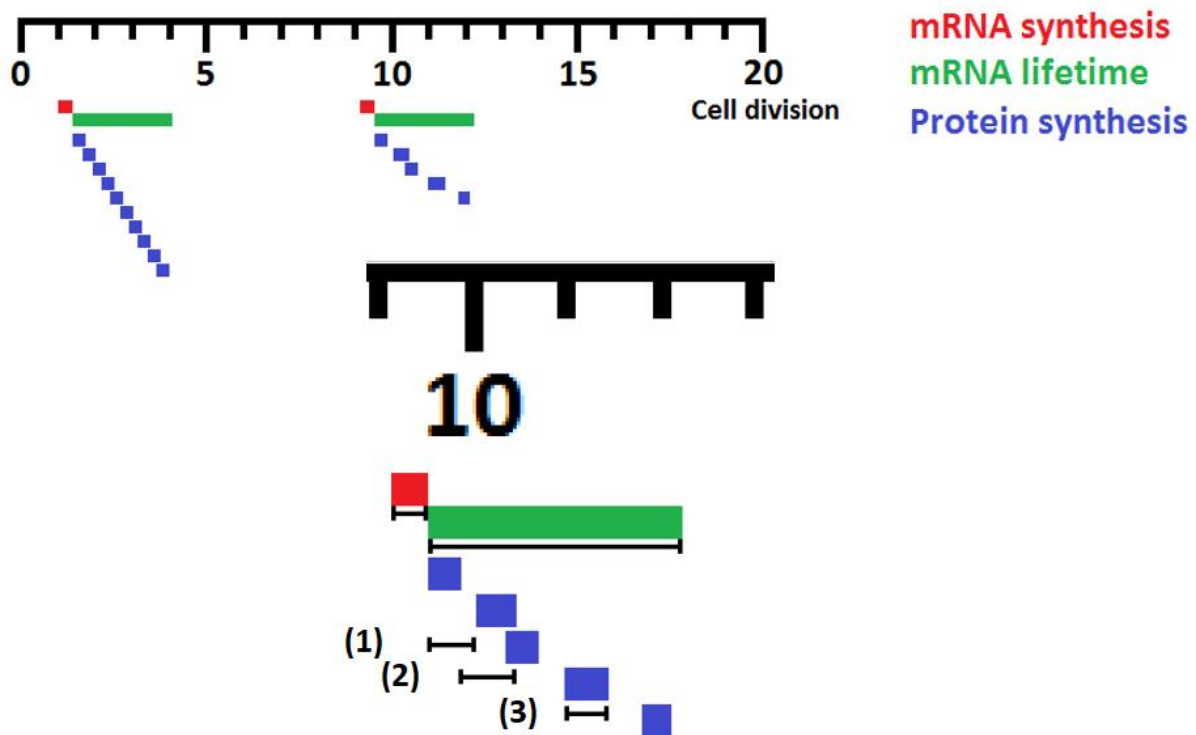

Figure 1: Example cell cycle with mRNA and protein production events. All data used in this figure are based on *E. coli* (BNID 100059 and 104900, and Bernstein *et al.*, 2002). Note that the protein synthesis times do not include insertion and folding.

As mentioned, we have a time resolution problem when fusing the fluorescent protein to the membrane protein directly. Instead we will make use of a method we call localisation labelling (Li *et al.*, 2011). To perform this method we need the fluorescent protein to be present in the cytoplasm and in the fluorescent state before the membrane protein is made. As soon as the membrane protein is produced the fluorescent protein should bind it. We can the detect the moment of binding because of the difference in diffusion rate between cytoplasmic and membrane proteins, which will allow us to pick an exposure time in the fluorescence microscope that causes the cytoplasmic fluorescence to spread out and the membrane protein fluorescence to be localised (Figure 2A and B). The diffusion rates (D) of membrane and cytoplasmic proteins have been determined for both *E. coli* and *L. lactis*. For membrane proteins with about 12 transmembrane helices $D = 0.03$ μm^2/s, and for cytoplasmic proteins of about 200 AA in length $D = 10$ μm^2/s (Kumar *et al.*, 2010; Mika *et al.*, 2014).

Rates of protein production can be determined by following the appearance of fluorescent spots over time. We start by observing a fluorescent cell that contains no spots and we wait for a spot to appear (in practice many cells are observed at the same time in all manner of states). The first spot contains no information on timing because there is no point of reference. Then a second spot appears. Now we can measure the time between the appearances of the first and second spot. This can be continued for a few more spots until the cell gets too crowded to see new spots appear. We cannot use increases in fluorescence intensity to observe an appearance, because we use localisation to label membrane proteins and the non-localised FP's contribute to the total fluorescence intensity. By measuring the

time between the appearances of spots in many cells we obtain a distribution of times. This time distribution can then be related to the underlying membrane protein production process.
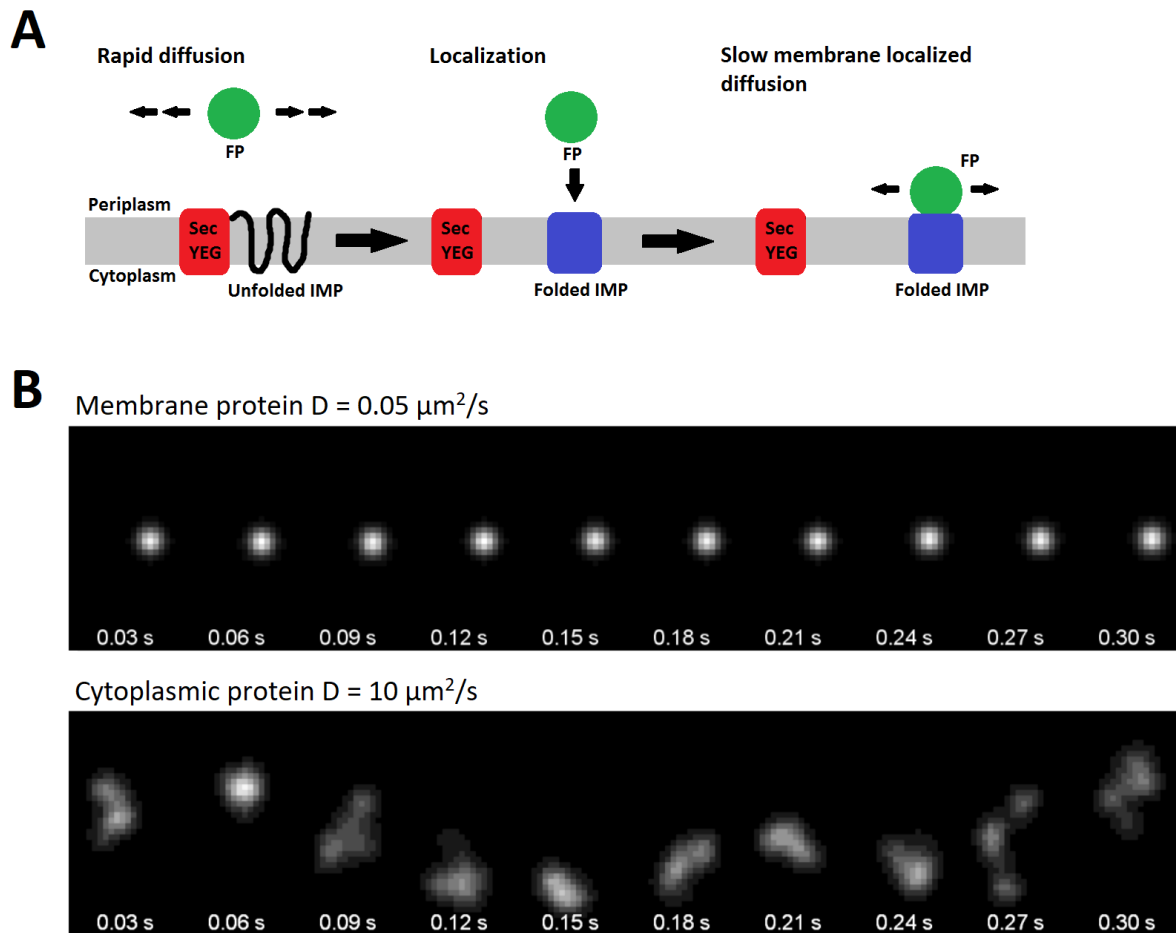

Figure 2: Illustration of localisation labelling for visualisation of membrane protein production. (A) Cartoon illustrating the concept of localisation labelling to detect when a membrane protein is being formed. FP = fluorescent protein, IMP = integral membrane protein. (B) Simulated fluorescence microscopy images of a bacterial cell that contains either one membrane protein (top, discrete fluorescent spot) or one cytoplasmic protein (bottom, diffuse signal from rapidly moving protein). Diffusion coefficients and time stamps are shown in the figure. The cell was modelled as a spherocylinder with a length of 3 μm and a width of 1 μm.

Before turning to the experimental section of this chapter we present some ideas to keep in mind while reading the rest:

- The method of localisation labelling, as proposed here, can be carried out with fluorophores present in the cytoplasm, but in principle also with fluorophores present in the periplasm or the extracellular space.
- When you combine the protein labelling system with an mRNA labelling system, it is possible to get the time between mRNA synthesis and the synthesis of the first protein.
- By labelling both the N-terminus and the C-terminus we can obtain the time it takes to synthesize a single membrane protein. This time includes translation, insertion, and folding. This can be done with two fluorescent proteins of different color to distinguish N- and C-

terminus. However, it can also be done with a single color by comparison to a membrane protein with only one interaction domain.

- To tease apart synthesis, insertion, and folding, and to study the effect of various elements in the membrane protein production pathway, the localization labelling technique should be carried out in mutant strains and in presence of various kinds of antibiotics. We can also vary ribosome binding site strength, and the length and nature of the membrane protein.

- In a slightly modified form the localisation labelling method can also be used to determine rates of assembly of membrane protein complexes. Take as an example ABC transporters, in which you could label with a fluorophore both the transmembrane domains, by localisation labelling, and the nucleotide binding domains, as a genetic fusion, to see with what rate the nucleotide binding domains bind to the transmembrane part. The same could be done for substrate binding domains on the outside of the cell.

- Association times are exponentially distributed which could influence the time between the appearances of spots independent of translation, insertion and folding. This hurdle can be (largely) overcome by looking at the distribution of times between membrane protein production events. The effect of binding times depends on the magnitude of the binding time and the magnitude of the time difference between the appearances of spots.

- The localisation labelling method is limited to proteins expressed at low levels. At high expression levels the background number of membrane proteins would be too high and/or the spots would represent proteins from multiple mRNA's.

- Even with our method maturation is still a problem as it causes part of the pool of fluorophores to be non-fluorescent. The reason is that when cells are constantly dividing you need to replenish the fluorescent proteins that are lost due to dilution. With cell division times of <1 h there will always be a number of freshly made fluorescent proteins that have not yet become fluorescent. Also not all fluorescent proteins necessarily become fluorescent (Chu *et al.*, 2014).

Getting the idea to work

Choice of organisms

All work was performed on two organisms: the bacteria *Lactococcus lactis* and *Escherichia coli*. *L. lactis* was used because of its utility as alternative host for the overexpression of membrane proteins; *E. coli* is the paradigm for research on bacterial physiology. Improved knowledge of the membrane protein production process may improve our ability to obtain large quantities of membrane proteins for biochemical and structural studies.

Genetic setup

For our measurements we needed to introduce two proteins into the cells of interest: (1) an integral membrane protein that is fused to a protein-protein interaction domain; and (2) a fluorescent protein that is fused to a complementary protein-protein interaction domain. For all our localisation labelling setup we use two compatible plasmids to carry the genes for the expression of the two proteins. Over the course of the project we swapped the coding sequences for membrane proteins, fluorescent proteins, interaction domains, and promoters; and also tested different hosts to optimise the method.

We started our work in *L. lactis*, with LacSΔIIA-SYNZIP2 as our membrane protein and YPet-SYNZIP1 as our fluorescent protein in the cytoplasm. We chose LacSΔIIA as our membrane protein because we had worked with it previously and seen that when expressed from a nisA promoter on a pNZ8048 plasmid it was present at a low enough level for identification of single molecules (Mika *et al.*, 2014). LacSΔIIA has both its N- and C-terminus pointed into the cytoplasm. We used YPet as fluorescent protein because we had used it previously to visualise single molecules in *L. lactis* cells (Mika *et al.*, 2014). SYNZIP1 and SYNZIP2 are small, leucine zipper based, complementary protein-protein interaction domains (Thompson *et al.*, 2012). They have high affinity <10 nM, are small, and should fold quickly due to their simple structure. There are many more SYNZIP pairs that do not cross-react with other SYNZIPs, which will be useful for determining the production time of a single membrane protein. To express YPet-SYNZIP1 we used the low copy number pIL252 vector, which was modified to include a transcription termination sequence behind the YPet-SYNZIP1 gene. We used a series of constitutive promoters of varying strength (CP1, CP2, etc.) to tune the expression of YPet-SYNZIP1 (Jensen *et al.*, 1998).

Microscope setup

The imaging was all performed on single-molecule sensitive fluorescence microscopes, like the one depicted in Figure 3A. These microscopes are equipped with lasers for excitation of the fluorophores and an EM-CCD camera for detection of the emitted fluorescence. Where appropriate samples were observed in a temperature controlled chamber on the microscope (Figure 3B). For many of our measurements we used a flow cell to measure for extended periods of time (hours) while keeping the cells growing (Figure 3C).
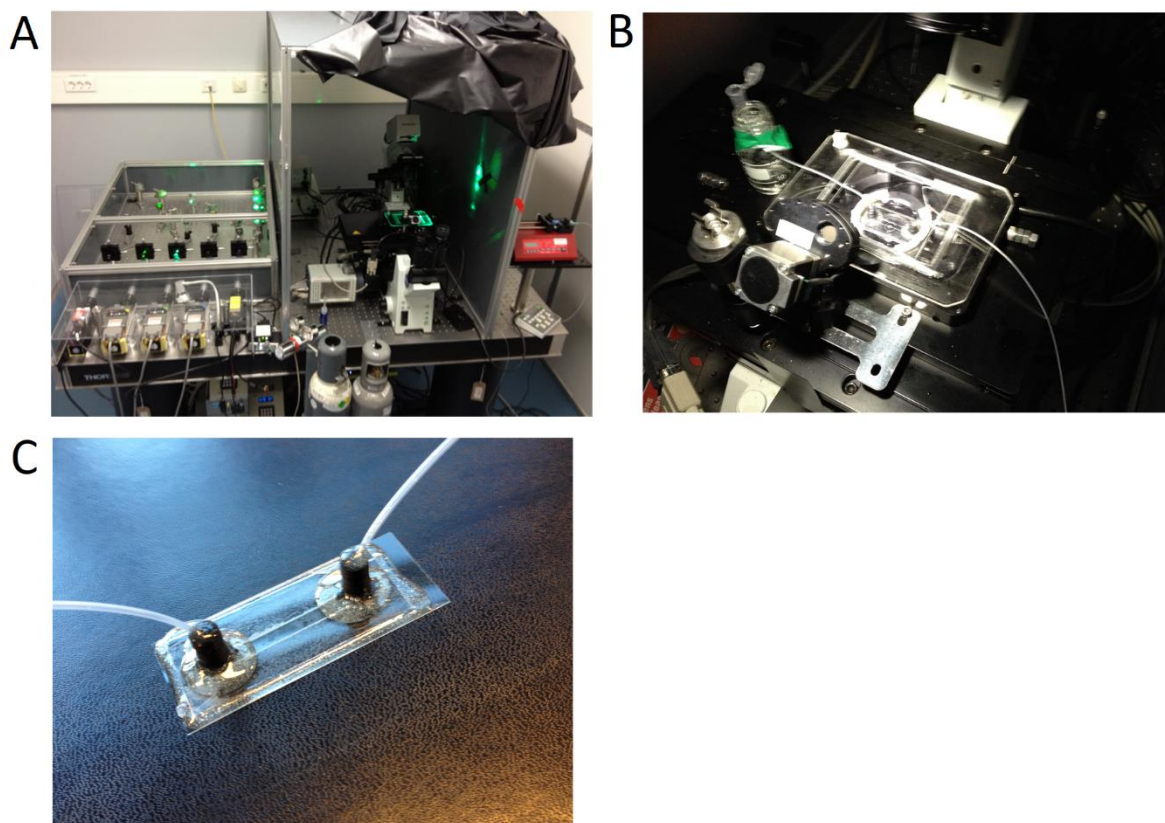
Figure 3: Microscope setup for the detection of single fluorescent molecules in cells. (A) The entire microscope setup. *Bottom left:* five lasers of different wavelengths, *middle left:* optics for bringing the laser beams into the microscope body, *right* microscope body with the sample holding chamber and EM-CCD camera, *far right* the syringe pump (in red) for controlling the flow of medium through the flow chamber. (B) Temperature controlled sample holding chamber. (C) Flow cell.

The dependence of time resolution on k_{on} and fluorophore copy number

The time resolution of localisation labelling depends on the association time between fluorophore and membrane protein. This in turn depends on the fluorophore copy number, the membrane protein copy number, and the on rate constant, k_{on}, of their interaction. The membrane protein copy number should be as low as possible, and here we consider it to be one per cell. We need a time resolution on the order of seconds because the estimates for synthesis time of an average sized protein is 20 s and the lifetime of mRNA is a couple of minutes (see Figure 1). We performed a series of simulations to determine what k_{on}'s and fluorophore copy numbers would allow for such a time resolution. For each simulation we started with a random distribution of a specified number of fluorophores in the cytoplasm and a random position of the membrane protein on the membrane. Then, we recorded the time until the membrane protein had interacted with a fluorophore. We used a range of interaction rate constants that is biologically plausible, *i.e.* below the diffusion limit (Alsallaq *et al.*, 2008). The results of the simulations are shown in Figure 4A and B, and Table 1. The average time from synthesis of the membrane protein to its visualisation is 5 s when the interaction has a k_{on} of 10^7 M^{-1}s^{-1} and when only 10 copies of the fluorescent protein are present in the cytoplasm. These are parameters that we can achieve (see below); under favourable conditions we could go to even better time resolution.
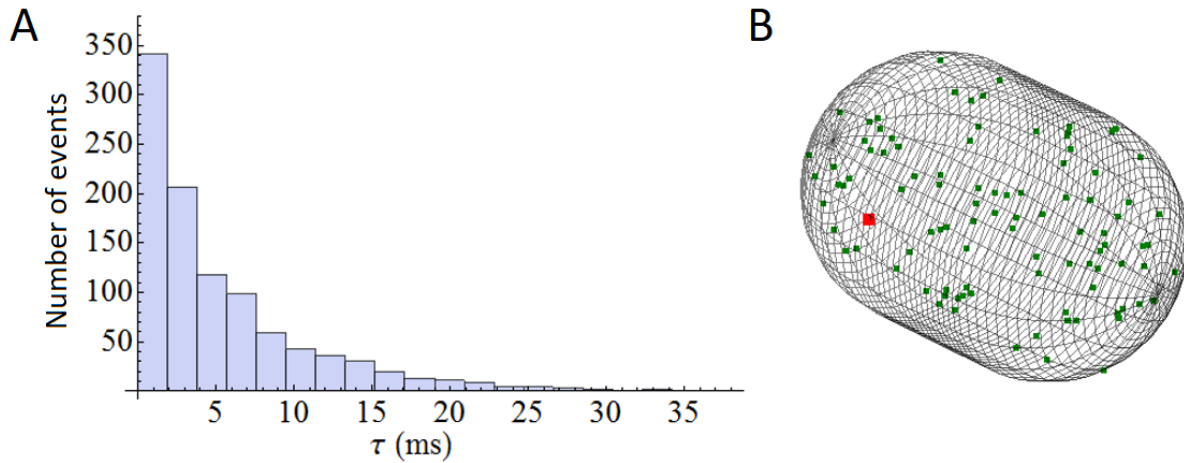
Figure 4: Simulation of association times in a model *L. lactis* cell. A) Histogram of association times for 1000 binding events. With the number of fluorescent proteins = 100 and k_{on} = 10^9 M^{-1}s^{-1}. B) Visualisation of the simulated system at time = 0. The cell is spherocylindrical with length 1.5 µm and width 1 µm. A hundred cytoplasm-localized fluorescent proteins are shown in green and the membrane protein is in red.

Table 1: Association times as a function of fluorophore copy number and k$_{on}$.

Number of fluorophores	k_{on} = 10^5 M^{-1}s^{-1}	k_{on} = 10^7 M^{-1}s^{-1}	k_{on} = 10^9 M^{-1}s^{-1}
10	500 s	5 s	0.05 s
100	50 s	**0.52 s**	**0.0052 s**
1000	5 s	0.05 s	**0.00052 s**

Note: the numbers in bold are the simulation results (means). The other numbers are extrapolations. When we did these simulations we did not have a clear sense of how much molecules we could have in the background. Therefore we focussed on the case with a 100 FP's even though we would need to be much closer to 10 fluorescent proteins.

The bleaching biogenesis experiment

Earlier we described the experiments that Yu et al. performed on Tsr-Venus in *E. coli*. We repeated these experiments on LacSΔIIA-YPet in *L. lactis* (see Figure 5 for an example). From these experiments we obtained the number of mRNA produced per unit time and the number of proteins produced per mRNA. Both of which are important for the interpretation of our localization labelling data. These measurements also allow us to compare the time resolution of this method to the time resolution of localisation labelling.

We performed three separate experiments with 1, 3 and 5 min between time points, respectively. At each time point we recorded a bright field image, three fluorescence images with a short exposure and one image with a long exposure. The fluorescence images with a short exposure time are used for data analysis. The long exposure is done to bleach all fluorophores so that only newly formed fluorophores are seen at each time point. For the experiment with 1 min between time points we imaged a total of 121 cells and find that the cells stop growing (from the bright field images) and that the total number of spots that appear in each frame drop after a couple of time points (Figure 6A). Subsequent determination of cell growth rates in the presence and absence of laser illumination showed that the long exposures make the cells stop growing. A similar issue was reported for *E. coli* (Yu *et al.*, 2006).
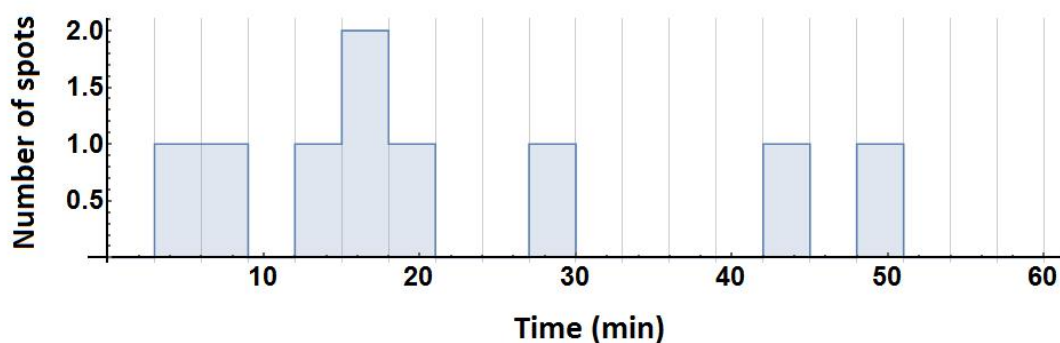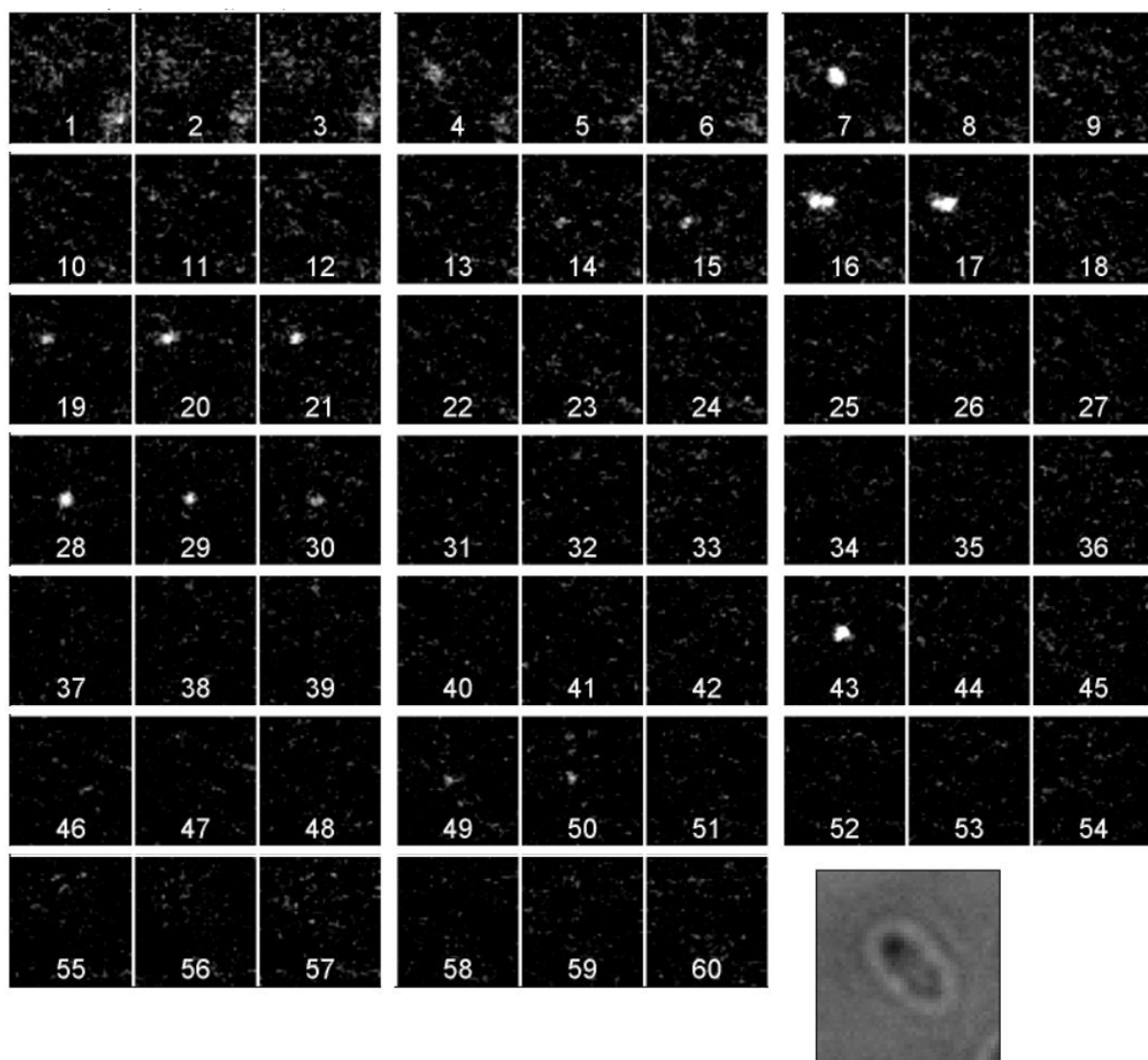
Figure 5: Bleaching biogenesis experiment on LacSΔIIA-YPet in *L. lactis*. The array of images is a movie recorded in the 514 channel. It shows the fluorescence from one cell. Each set of three images represents one time point (the numbers indicate the picture index, not time). These three images were recorded immediately after each other with an exposure time of 100 ms. The time points (each set of three images) are three minutes apart. A transmission (bright field) image is shown to the bottom right to show where the cell is. Note that the position of the cell shifted during the experiment. The histogram at the bottom indicates the number of spots over time in this single cell. The vertical lines indicate times at which the sets of three images are recorded.
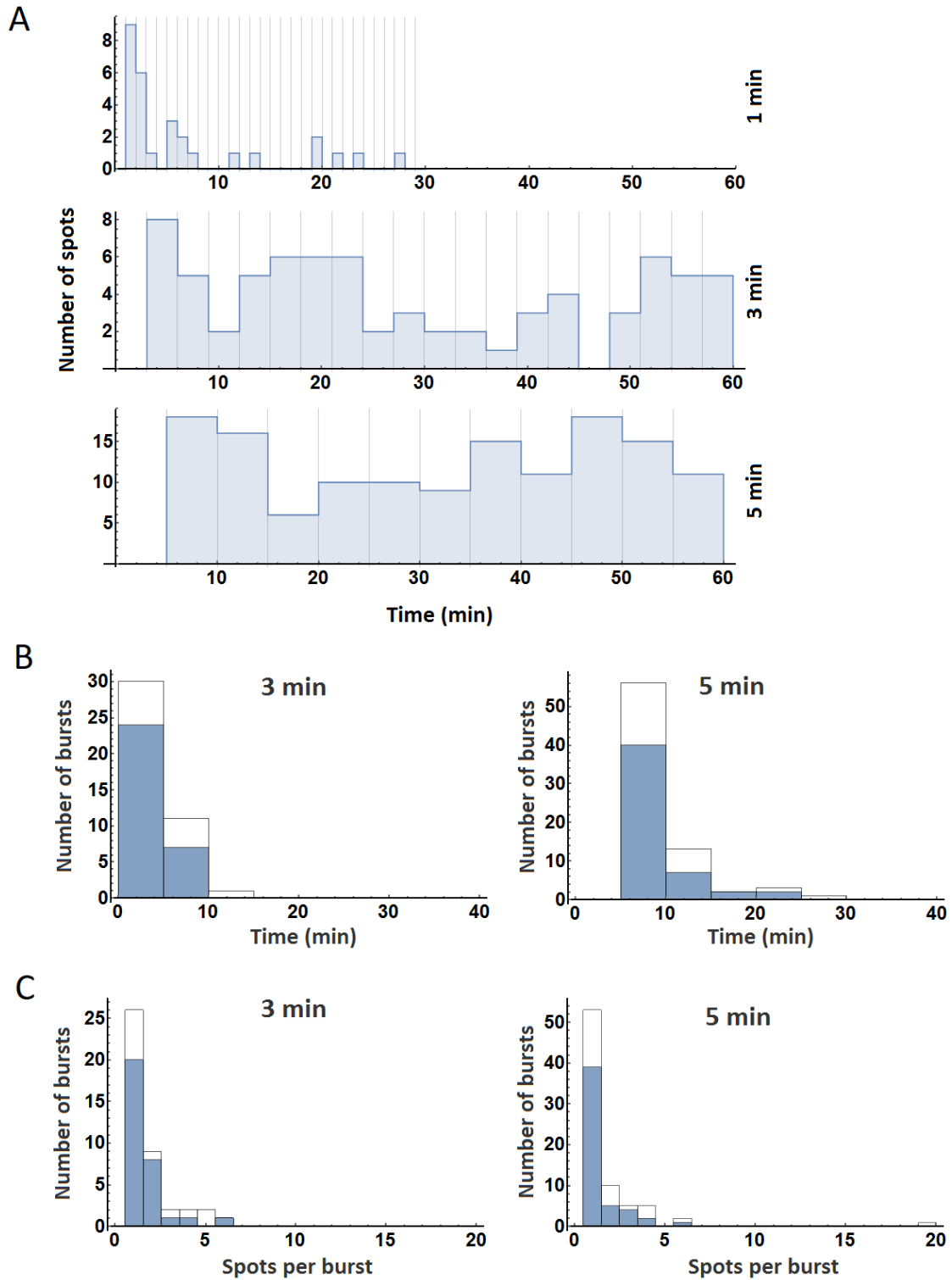
Figure 6: Results of the LacSΔIIA-YPet bleaching biogenesis experiment in *L. lactis*. (A) Total number of spots in all cells over time, for a time step of 1, 3 and 5 min, which shows whether cells have a constant (steady state) production of membrane proteins. Vertical lines indicate when images are recorded. (B) Histograms of the duration of a burst of LacSΔIIA-YPet production for a 3 and 5 min time step. The blue bars show the histogram when excluding bursts that begin or end in the first or last frame, respectively. Blue plus white bars show the histogram when including those bursts. (C) Number of spots per burst for a 3 and 5 min time step. Blue and white bars have the same meaning as in (B).

We proceeded with 3 and 5 min between time points to reduce the adverse effects of laser illumination. From Figure 6A it can be seen that the total number of foci per time point is stable through time for both the 3 min and 5 min time step experiments. This shows that the time resolution is limited to >1 min (this does not take into account formation of the fluorophore). For the experiment with a 3 min time step we analysed 37 cells of which 18 did not produce any LacSΔIIA-YPet for the entire experiment. In total we observed 74 spots, which is 2 spots per cell on average. For the experiment with a 5 min time step we analysed 60 cells of which 15 were empty. In total we observed 139 spots, which is 2.3 spots on average. We note the possibility of overlap of the signals from two proteins, but we assume this occurs too rarely to affect the numbers.

In Figure 6B and C we show histograms of the time each burst lasts and how many spots are found per burst. For both there is a distribution that looks like an exponential. Average burst times and number of spots per burst are shown in Table 2. The total number of bursts are 42 and 76 for the 3 min and 5 min time step experiments, respectively. If we take into account the number of cells analysed for each experiment and the time of observation, 60 min, we can calculate the average time between a burst, i.e. the production of an mRNA. This gives an average time of 53 min and 47 min for the 3 and 5 min time step experiments, respectively. Here, we assumed that all cells statistically behave in the same way so that all differences we have seen between cells are because of chance differences in "grabbing" cells from a distribution. We found that only a small fraction of bursts produced more than one protein, which is of importance for the localisation labelling experiments shown later on.

Table 2: Summary LacSΔIIA-YPet biogenesis experiments.

	3 min time step	5 min time step
Number of cells	37	60
Number of empty cells	18	15
Total number of spots	74	139
Average number of spots per cell	2	2.3
Total number of bursts[1]	31 (42)	51 (76)
Average time between bursts[1]	72 (53) min	71 (47) min
Burst time[2, 3]	3.9 ± 1.8 min	6.7 ± 3.7 min
Average number of spots per burst[2]	1.6 ± 1.1	1.5 ± 1.0

1) The numbers outside of the brackets are excluding bursts that start or end in the first or last frame, the number within brackets includes all bursts.
2) Bursts that started or ended in the first or final frame were not included.
3) These numbers are influenced by the size of the time step.

To extract the number of proteins that are produced from a single mRNA in these experiments, it is important to know that a burst is produced by only one mRNA. The same holds for determining timing between protein production events from a single mRNA. Various studies have reported bursts in the production of mRNA's (Golding *et al.*, 2005; Le *et al.*, 2005; Li *et al.*, 2011; Chong *et al.*, 2014). One way to determine the number of mRNA's per burst is to label mRNA's with fluorophores and count their numbers under a fluorescence microscope (Bertrand *et al.*, 1998; Golding *et al.*, 2005). We can also determine the number of mRNA's from the distribution of the number of proteins produced per burst. When there is only one mRNA per burst, we should see an exponential distribution of the number of proteins per burst (Li *et al.*, 2011). When there are two, three, four, etc. mRNA's per burst we see the distribution shifting from an exponential to a Gaussian (central limit theorem). This behaviour can be described by a gamma distribution. The reason for this is as follows. In the case of two mRNAs the number of proteins produced in a particular burst is the sum of the number of proteins

from the one mRNA and the number of proteins from the second mRNA. For each mRNA separately the number of proteins is exponentially distributed. For each protein number on the one mRNA you can add an amount of proteins from the other mRNA with an exponentially distributed probability; this operation is called a convolution. It turns out that the convolution of two or more exponentials follows a gamma distribution (Blitzstein *et al.*, 2014):

$$f(x) = \frac{1}{\Gamma(a)} (\lambda x)^a e^{-\lambda x} \frac{1}{x} \qquad (1)$$

Here $\Gamma(a)$ is the gamma function, a is the number of exponentials (or mRNAs), λ is the decay constant of the exponential decay (mRNA life time), and x is the number of proteins. By putting in the appropriate exponential decay constant and varying the number of mRNAs we get the graphs presented in Figure 7. By comparing the experimentally determined distribution of number of proteins per burst to a gamma distribution that represents a particular number of mRNAs, we can determine the number of mRNAs per burst. Our data seems to support a single mRNA per burst (compare Figures 6C and 7). A final note on exponential and gamma distributions. Both these distributions are continuous, whereas the number of proteins follows a discrete distribution. However, for the interpretation we simply have to make the conversion between the two in our heads, for the numbers this doesn't matter.

In the experiments presented here we have a rather limited number of cells and therefore we can't make very strong statements about the exact values of the number of bursts per unit time, the number of proteins per mRNA, and the number of mRNA's per burst. Therefore these experiments should be regarded is exploratory and not the final word. However, the two separate experiments do get roughly the same values for number of mRNA's per unit time and number of proteins per mRNA, so we decided to use this data in modelling the number of useful events in a localisation labelling experiment (see below). There is a difference in the burst time between the two experiments but that is probably more a reflection of the difference in the size of the time step than an actual difference between the populations of cells. Another issue with these measurements is that some long bursts of protein production may actually be the result from >1 mRNA, and, conversely, two bursts of protein production separated by one empty frame may in fact be the result of one mRNA. Bunching bursts that are 1 frame apart for the 5 min data, and 1 or 2 frames apart for the 3 min data, yields distributions similar in shape but slightly broader than those shown in Figure 6C.
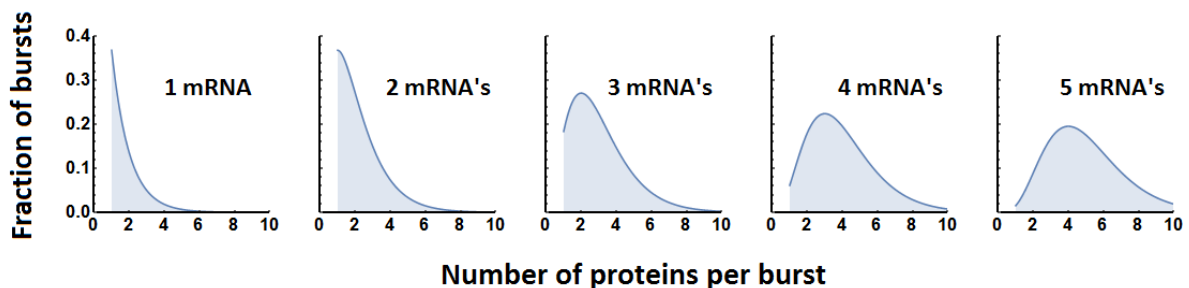

Figure 7: Distributions of the number of proteins per burst as a function of the number of mRNA's per burst. These probability distributions were calculated using the probability density function of the gamma distribution (equation 1).

First iteration of localisation labelling: LacSΔIIA-SZ2 and YPet-SZ1

To keep things simple we started off trying to determine the time between successive protein production events from a single mRNA. We used *Lactococcus lactis* strain NZ9000 and the pNZ8048 vector to express LacSΔIIA-SZ2 from the nisin inducible promoter, and the pIL252 vector to express YPet-SZ1 from the CP1 or CP23 promoter. In Figure 8 we show fluorescent images for the strains with promoters CP1 and CP23, and in the presence and absence of nisin for the induction of LacSΔIIA-SZ2. We compare this to non-induced expression of LacSΔIIA-YPet to get an idea of the expression level of LacSΔIIA-SZ2. From the data it is clear that (1) YPet-SZ1 is produced from both CP1 and CP23 promoters; (2) CP23 yields a higher level of expression, as was shown before (Jensen *et al.*, 1998); (3) the number of spots visible for the localisation labelling is much lower than for LacSΔIIA-YPet; and (4) upon induction with nisin the fluorescence of YPet-SZ1 moves to the membrane in both the CP1 and CP23 strains. From the membrane localisation after induction we can conclude that the interaction between SZ1 and SZ2 occurs. However, the localisation labelling didn't work due to the low number of spots, which may have a variety of causes:

(i) YPet is an unstable fluorophore, which limits the amount of photons we can capture per image. More photons means less noise and thus, possibly, more visible foci. Note that this doesn't influence "noise" that derives from the inhomogeneity in the distribution of background fluorescent molecules.

(ii) The binding between SZ1 and SZ2 is too weak. The K_d is reported to be <10 nM (Thompson *et al.*, 2012). If we assume that $K_d = 10$ nM then with 20 molecules in the background one third of the binding sites would be unoccupied. There is also uncertainty about the applicability of K_d's determined in dilute solution to the cytoplasm.

(iii) The background expression level of YPet-SZ1 is too high. This is clearly the case for CP23 where no spots are seen. The reason for this is that even though the signal to noise of the background decreases with more background fluorescence the absolute level of noise increases. It is this absolute level that we are concerned about when we need to detect spots against this background. The sparsity of spots in CP1 could also be caused by a too high background.

We have attempted to address each of these three points. Point (i) by trying different fluorescent proteins. Point (ii) by using protein-protein interaction domains with higher affinities. We also didn't know the k_{on} of the SYNZIP's and initially assumed it would be fast enough. After more detailed study of the literature on protein-protein interactions we decided to use protein interactions for which the k_{on} values are known and high enough (many protein pairs interact slowly with $k_{on} = 10^5$-10^6 M^{-1}s^{-1} (Alsallaq *et al.*, 2008)). Finally, point (iii) by having more than one fluorophore per membrane protein.
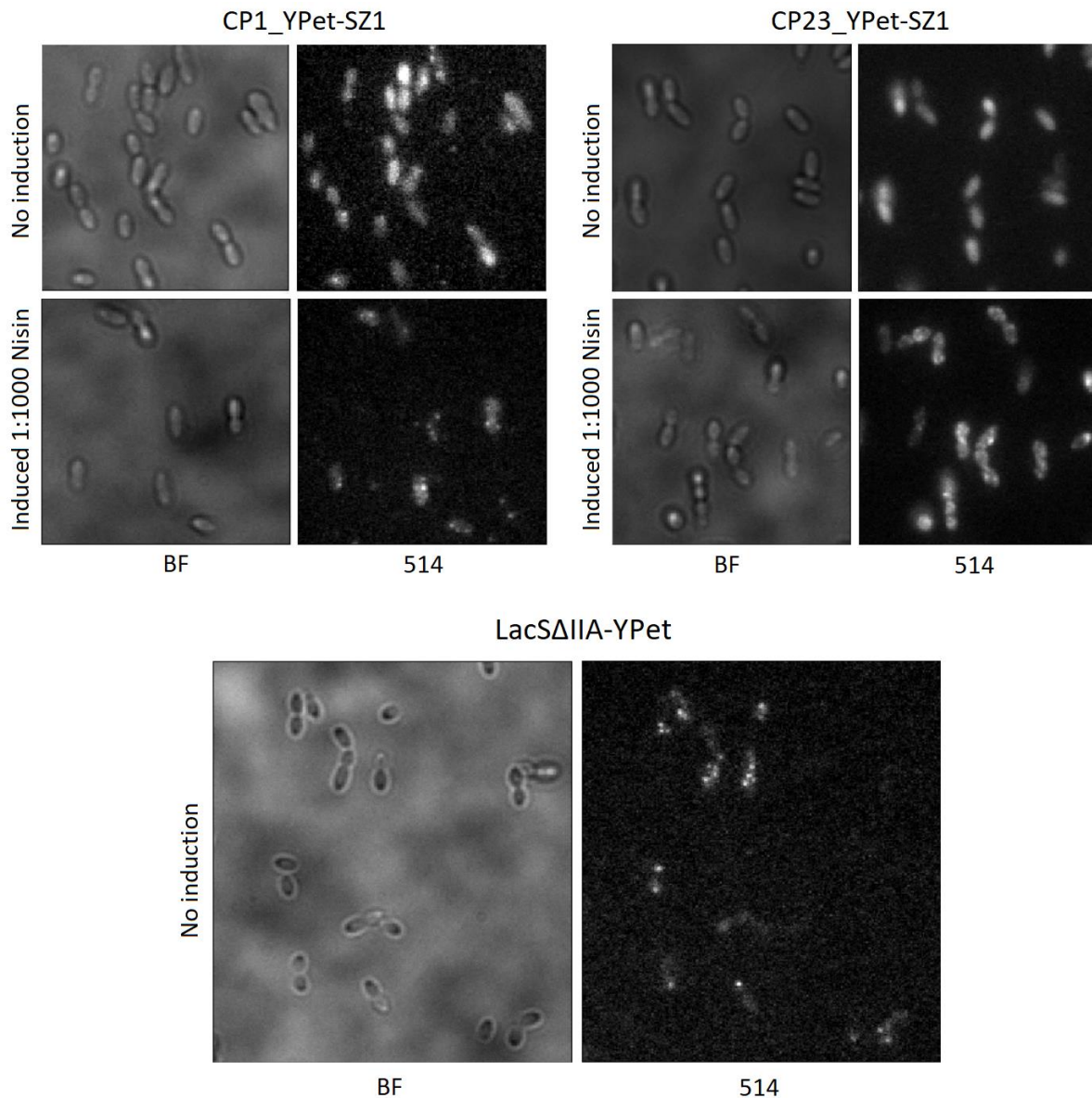
Figure 8: Localisation labelling using LacSΔIIA-SZ2 and YPet-SZ1 in *L. lactis*. Results shown for expression of YPet-SZ1 with the CP1 and CP23 promoters, with and without the induction of LacSΔIIA-SZ2. LacSΔIIA-YPet is shown to indicate the number of LacSΔIIA-SZ2 we expect when we do not induce. Note that the localisation labelling data shown here was recorded with an exposure time of 500 ms and the LacSΔIIA-YPet data with an exposure time of 50 ms. Localisation labelling data for the CP1 strain shows the same results at an exposure time of 50 ms as it does at 500 ms.

Faster protein-protein interactions and more stable fluorescent proteins

For the next round of localisation labelling we introduced different interaction domains to increase the k_{on} and used a more stable fluorescent protein. We switched our protein-protein interaction domain pair to Barnase-Barstar or ColicinE9-Im9. For these proteins we know that the k_{on}'s are very high: 10^9 M^{-1}s^{-1} at low salt concentrations and about 10^7 M^{-1}s^{-1} at the salt concentration we expect for *L. lactis* (Poolman *et al.*, 1987; Alsallaq *et al.*, 2008). Each of these proteins is roughly 100 amino acids in length and as such their synthesis and folding shouldn't much affect the measured times.

We switched to mNeonGreen as our fluorescent protein as it is reported to have a similar brightness (= extinction coefficient x quantum yield) as YPet but be 3 times as photo stable (Shaner *et al.*, 2013).

We also tried mCardinal which is about 20 times as photo stable as YPet but suffers from low brightness (Shaner *et al.*, 2013; Chu *et al.*, 2014). At higher levels of expression of LacSΔIIA-mCardinal we could detect mCardinal signal but we were not able to detect single molecules.

We ended up with a *L. lactis* strain that contained the plasmids pNZ8048, which produced LacSΔIIA-ColicinE9 from a nisin inducible promoter; and pIL252, which produced mNeonGreen-Im9 from either a CP1 or CP23 promoter. In Figure 9 we compare fluorescence images of these two strains, in uninduced and induced state, to images of LacSΔIIA-mNeonGreen. When mNeonGreen-Im9 is produced from the CP23 promoter its level of fluorescence is too high for the detection of spots. The level of cytoplasmic fluorescence from CP1 is low enough for spots to be visible, so we decided to perform a proper localisation labelling experiment with the CP1 strain. After induction of LacSΔIIA-ColicinE9, in both the CP1 and CP23 strains, the fluorescence seems to be present predominantly in spots, showing that ColicinE9 and Im9 interact.

Next we carried out a localisation labelling experiment on the *L. lactis* strain containing LacSΔIIA-ColicinE9 and CP1_mNeonGreen-Im9 (Figure 10). We took a bright field image to localise the cells and to check whether they are in focus. Then we recorded 10 fluorescence images with 100 ms exposure time and a 5 s time step between images. Finally, we took another bright field image to see if any of the cells moved during the recording of fluorescent images. We performed the same experiment on a strain that contained only CP1_mNeonGreen-Im9 to see how much localisations occur in the absence of LacSΔIIA-ColicinE9. Before estimating the time between protein synthesis events from the appearance times of foci, we checked how many spots appear and disappear over time. This data is summarised in Table 3.

For the strain containing both LacSΔIIA-ColicinE9 and mNeonGreen-Im9 there are 49 increases in the number of spots and 59 decreases. This is for a total of 55 cells and 45 s of observation time per cell. The numbers of increases and decreases do not bode well for the extraction of the time between protein synthesis events. First we consider the number of increases. Comparing with the bleaching biogenesis experiment above we expect a burst of synthesis to happen only once per ~50 min per cell, and since each burst has only 1.5 spots on average this means that we expect to see only 0.02 increases per cell in the measured time window. This means that the signal is overwhelmed by the noise. The noise is probably a combination of (i) spots moving in and out of focus; (ii) blinking of the fluorophore; (iii) fluctuations in the positions of cytoplasmic fluorophores; and (iv) camera noise. Now going to the number of decreases. We expect the number of decreases to be approximately zero over the measurement time window because proteins typically have lifetimes of several hours (BNID 109921, for *E. coli*), and dilution of the proteins by cell growth also happens on the timescale of an hour. By examining the *L. lactis* strain that produces only mNeonGreen-Im9 we can determine the number of spurious interactions mNeonGreen-Im9 makes with other components in the cell. As indicated in Table 3, the number of spurious binding events is about 15 % as the number of binding events seen in the presence of LacSΔIIA-ColicinE9. This means that the signal, i.e. localisation of mNeonGreen-Im9 to LacSΔIIA-ColicinE9, is overwhelmed by noise.

The conclusion is that the system as it stands is overwhelmed by noise; and localisation labelling, as performed here, cannot be used to extract the time between synthesis events.

Table 3: Experiment parameters of localisation labelling for *L. lactis* expressing LacSΔIIA-ColicinE9 and CP1_mNeonGreen-Im9.

	LacSΔIIA-ColicinE9 and CP1_mNeonGreen-Im9	CP1_mNeonGreen-Im9
Number of cells	55	48
Number of fluorescent cells	39	34
Number of cells with spots	25	14
Average number of spots in cells with fluorescence	4.33	0.65
Number of *increases* in the number of spots	49	-
Number of *decreases* in the number of spots	59	-

Extracting information from data with only two time points

Better fluorescent proteins would allow us to record more frames, and to maximise the number of fluorescent proteins that stay fluorescent throughout the measurement. However, it is possible to extract a mean time between appearances of spots with only two frames. This method relies on a number of assumptions that may or may not hold in actual measurements. Under conditions in which the time between the two measurements is significantly smaller than the average time between the appearances of spots, the average time between the appearances of spots is given by:

$$t_{average} = \frac{N_{1\to2} + N_{1\to1}}{N_{1\to2}} \times t_{interval} \qquad (2)$$

Here $t_{interval}$ is the time between measurements, $t_{average}$ the time between the appearances of spots, and $N_{1\to2}$ the number of cells in which the first frame has a single spot and the second frame has two spots. The reasoning for this equation is as follows. First we assume that all times between the appearances of spots are exactly the same (when this is not the case the average time that we get out is slightly different from the actual average time). We also assume that at any one moment there is only one mRNA producing proteins. The production of the proteins is followed by the appearance of fluorescent spots. Every cell that we image under the fluorescent microscope can by classified by the number of spots there are in the first and the second frame. The probability for a particular cell to produce the mRNA, and subsequently the proteins, is the same for each moment in time. This means that, given an appropriately large number of cells, we can use cell counts as a measure of time. So $N_{1\to2}$ is associated with an amount of time. The time that is represented by $N_{1\to2}$ is equal to $t_{interval}$. The time that is represented by $N_{1\to2} + N_{1\to1}$ is equal to $t_{average}$. The times represented by the cell numbers are calibrated by $t_{interval}$ so that $t_{average}$ is given by equation 2.

Note that we have also assumed that all cells produce more than one protein, which we know to be incorrect. However, performing a bleaching biogenesis experiment as shown above will give you the fraction of cells producing only one protein and allow you to correct $N_{1\to1}$. Another problem is the fact that there are proteins left over from previous bursts, which will also change $N_{1\to1}$. This can be solved by measuring in cells that grow and divide and produce a mRNA only every couple of generations. In that way proteins from previous bursts get diluted out.

LacSΔIIA-mNeonGreen

BF   514

LacSΔIIA-ColicinE9 and **CP1**_mNeonGreen-Im9

Uninduced   Induced

LacSΔIIA-ColicinE9 and **CP23**_mNeonGreen-Im9
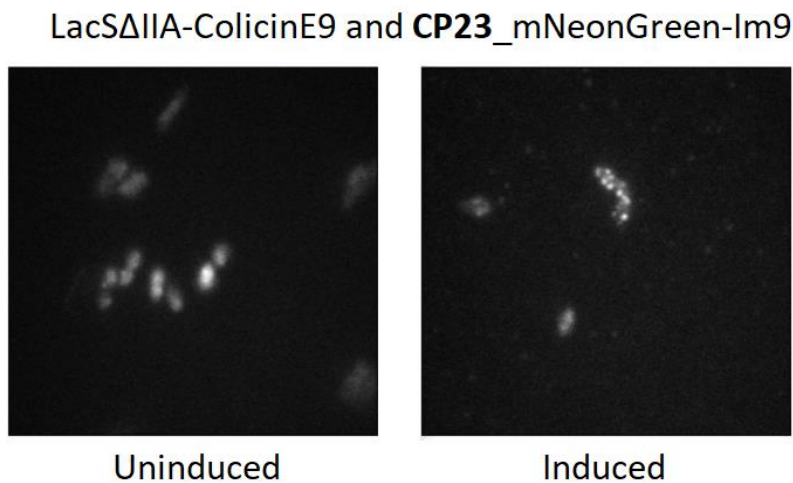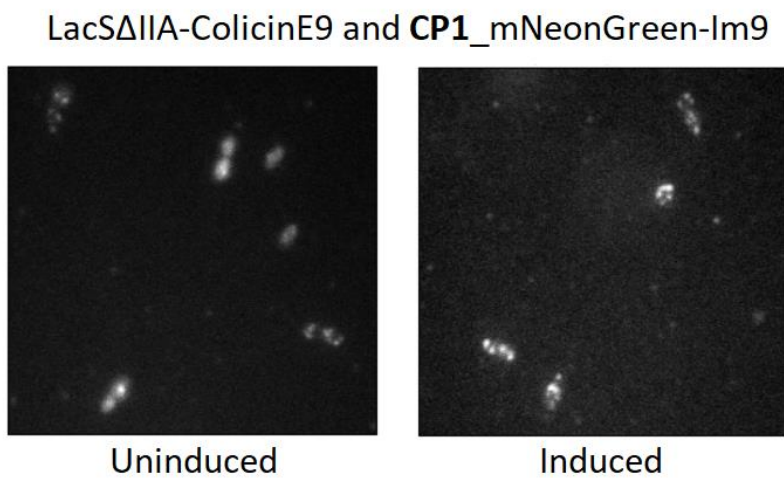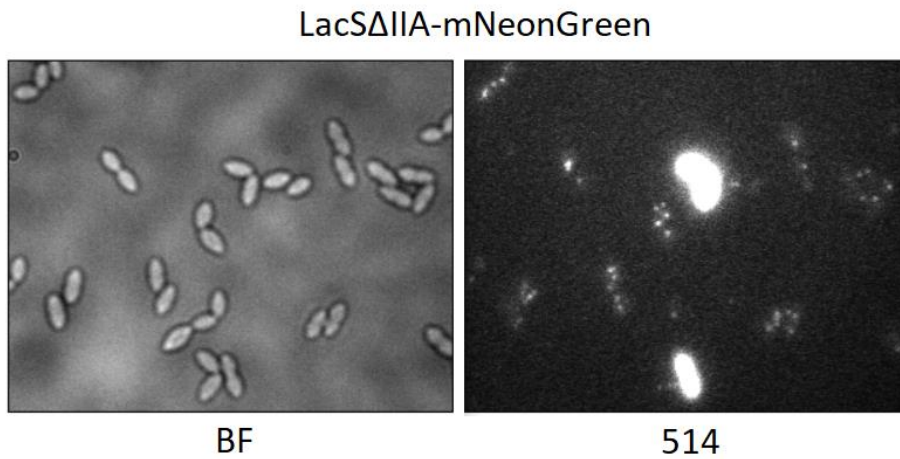
Uninduced   Induced

Figure 9: Testing of mNeonGreen/ColicinE9-Im9 strains for localisation labelling in *L. lactis*. The LacSΔIIA-mNeonGreen strain is imaged to see how many membrane proteins are expected. The presence of spots in uninduced LacSΔIIA-ColicinE9 and mNeonGreen-Im9 strains indicates that localisation of mNeonGreen-Im9 occurs, and that the cytoplasmic level of fluorescence is low enough. The induced case for the strains with LacSΔIIA-ColicinE9 and mNeonGreen-Im9 is a control to test for the interaction between the ColicinE9 and Im9. All fluorescence images were recorded in the 514 channel, with 100 ms exposure times.
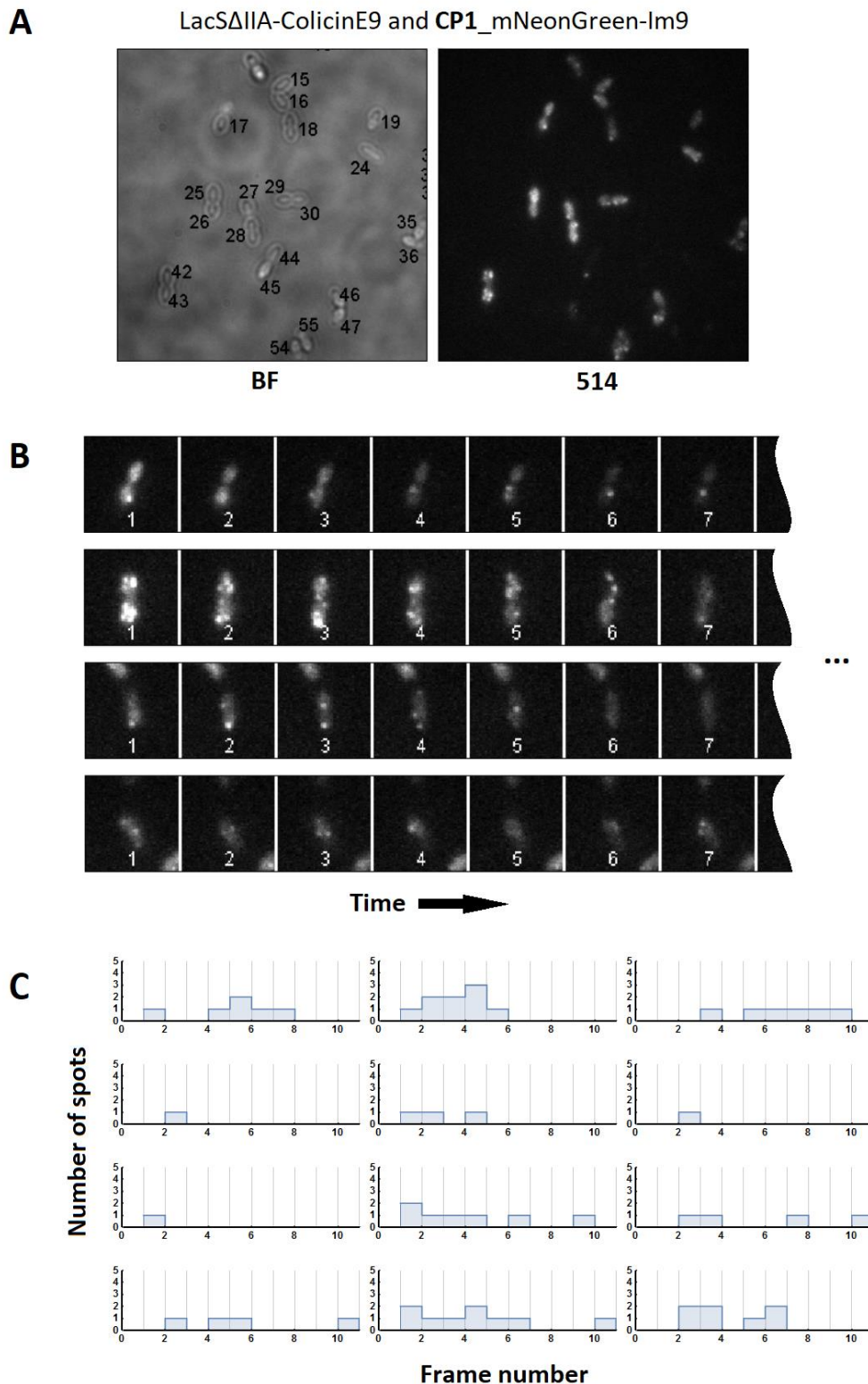
Figure 10: Localisation labelling in *L. lactis* with LacSΔIIA-ColicinE9 and CP1_mNeonGreen-Im9. (A) Overview image of a field of cells expressing LacSΔIIA-ColicinE9 and mNeonGreen-Im9. Fluorescent spots indicate the presence of LacSΔIIA-ColicinE9. (B) Time series of fluorescent images of four different cells. The time between frames is 5 s, the exposure time is 100 ms. The frame numbers are indicated in the images. (C) Number of spots over time extracted from the time series for 12 different cells. All fluorescence images in this figure were recorded in the 514 channel.

130

Localisation labelling in *Escherichia coli*

We also tested the localisation labelling method in *Escherichia coli* MG1655, using strains: (1) containing plasmids pACYC_rhaBAD_Barstar-mNeonGreen and pBAD_araBAD_LacY-Barnase, and (2) containing plasmids pACYC_rhaBAD_Im9-mNeonGreen and pBAD_araBAD_LacY-ColicinE9. Working in *E. coli* has two potential advantages over *L. lactis*: (1) we have two inducible promoters which allows for much more rapid testing of expression levels; and (2) *E. coli* has a much lower cytoplasmic ion concentration than *L. lactis* (Poolman *et al.*, 1987; Konopka *et al.*, 2009), which means that the interactions between Barnase and Barstar, and ColicinE9 and Im9 will be faster. However, fluorescence microscopy of the strain containing Barstar-mNeonGreen and LacY-Barnase showed that in the absence of an inducer no spots are observed, which may be due to a high background level of fluorescence. Upon induction of the expression of LacY-ColicinE9 or LacY-Barnase fluorescent spots are observed but there is still a lot of background fluorescence, suggesting that there is something amiss with the interaction between Barnase and Barstar, and ColicinE9 and Im9. Alternatively, the expression of LacY-Barnase and LacY-ColicinE9 is not high enough to pull all fluorophores to the membrane. The conclusion is that localisation labelling cannot be performed in these strains under the tested conditions.

ACE expression system for *Lactococcus lactis*

We also tested a second inducible promoter for the *L. lactis* localisation labelling system, the ACE promoter (Linares *et al.*, 2015). A second inducible promoter would allow us to test expression levels more quickly. We performed the test by putting the ACE promoter in front of LacSΔIIA-YPet and imaging the resulting strain, without induction, under the fluorescent microscope. We also performed the same experiment in the absence of lysine and arginine, because these amino acids are structurally similar to the inducer of the ACE promoter (agmatine). This could lead to induction in the absence of agmatine. Under all these conditions the cytoplasmic expression level was too high for localisation labelling to work, so we abandoned further work on this promoter system.

Labelling from the outside

In parallel to trying to do localisation labelling with fluorophores from the cytoplasm we made an attempt to do the same with fluorophores from the extracellular medium. This is possible in *L. lactis* as its plasma membrane is in direct contact with the extracellular medium. Labelling from the outside gives us more options for the choice of a fluorophore. It also gives us the opportunity to use Trolox to stabilise the fluorophore against photo-bleaching (van der Velde *et al.*, 2016). Another possible benefit is the ability to switch the outside medium in a pulse like manner with high concentration peaks of fluorophore and with the measurements happening in the troughs with low levels of fluorophore. Control over the outside medium also gives us the option of reducing the ion concentration to increase the k_{on} of electrostatically steered interactions. A disadvantage of labelling from the outside is that there is a cell wall surrounding the plasma membrane which could block or slow down an interaction between the membrane protein and the fluorophore. We can remove the cell wall, e.g. with mutanolysin (Kondo *et al.*, 1982), but this might disrupt membrane protein production. Another disadvantage is the large fluorescence background that arises from having the flow cell compartment being fluorescent. This could be minimized by having very thin flow cells.

We chose OpuABC as the membrane protein because its C-terminus sticks out of the cell (Biemans-Oldehinkel *et al.*, 2006). We fused Barnase to the C-terminus of OpuABC and expressed this fusion in *L. lactis*. Barstar was produced in *E. coli* and labelled with Sulfocyanine-5 NHS ester. We reached a labelling efficiency of 90 % (ignoring multiple labels on the same protein). However, when we added

the labelled barstar to the cells the entire glass surface was covered with fluorescent spots. This happened independent of how we treated the glass: KOH cleaned glass, polylysine coated, dichlorodimethyl silane coated, or (3-aminopropyl) triethoxysilane. The background thus obscured any possible signal that would have come from barstar binding to *L. lactis*. Going into the future there are two other options that we have not yet tried: (1) coating the entire glass surface with PEG, and (2) adding the protein BSA, to block the binding sites on the glass.

A model for predicting the number of cells needed for localization labelling

From the bleaching biogenesis experiment on LacSΔIIA-YPet (described above) we were able to estimate the number of appearances of membrane proteins per unit time. In this section we will derive a model that describes the appearance of membrane proteins, and our ability to see them, more systematically. This is useful for optimizing the efficiency of data collection. We assume that the probability of mRNA production is constant through time (steady state). The model gives the fraction of the cells observed that show a pair of membrane protein production events that arise from one mRNA and that we can observe. It has the following form:

$$Fraction = \frac{t_{obs}}{t_{mRNA}} \times \frac{\int_2^\infty e^{-\frac{z \times t_{prot}}{t_{mRNAlife}}} dz}{\int_0^\infty e^{-\frac{z \times t_{prot}}{t_{mRNAlife}}} dz} \times \frac{\int_0^1 \frac{n^{a-1} \times e^{-\frac{n}{b}}}{b^a \times \Gamma(a)} dn}{\int_0^\infty \frac{n^{a-1} \times e^{-\frac{n}{b}}}{b^a \times \Gamma(a)} dn} \qquad (3)$$

with,

$$a = \frac{t_{cellcycle}}{t_{mRNA}} \qquad (3a)$$

$$b = \frac{\int_0^\infty z \times e^{-\frac{z \times t_{prot}}{t_{mRNAlife}}} dz}{\int_0^\infty e^{-\frac{z \times t_{prot}}{t_{mRNAlife}}} dz} \qquad (3b)$$

Here, t_{obs} is the time we observe the cells, t_{mRNA} is the average time between the appearances of subsequent mRNA's, $t_{protein}$ is the average time between the appearances of subsequent proteins from a single mRNA, $t_{mRNAlife}$ is the time constant for the exponential decay of the mRNA, $t_{cellcycle}$ is the cell cycle time, and $\Gamma(a)$ is the gamma function of a.

Equation 3 is derived as follows. The fraction of cells in which we can observe a pair of subsequent protein production events is the product of three other fractions. (1) The fraction of cells that produces a mRNA in the observation time window, F_{mRNA}; (2) The fraction of mRNA's that produces two or more proteins, $F_{protein \geq 2}$; and (3) The fraction of cells that have a low background (0-3 copies) of the membrane protein of interest already present, $F_{background}$. This gives:

$$Fraction = F_{mRNA} \times F_{protein \geq 2} \times F_{background} \qquad (4)$$

The number of mRNA's produced per unit of time per cell is given by $\frac{1}{t_{mRNA}}$, the number of mRNA's produced per cell in the observation time window is $\frac{t_{obs}}{t_{mRNA}}$. So

$$F_{mRNA} = \frac{t_{obs}}{t_{mRNA}} \qquad (5)$$

For $F_{protein \geq 2}$ we take into consideration the fact that the mRNA lifetime is exponentially distributed and is given by $e^{-\frac{t}{t_{mRNAlife}}}$. We can rewrite this function into terms of numbers of protein instead of time by using the average time between the production of proteins, t_{prot}, which yields: $e^{-\frac{z \times t_{prot}}{t_{mRNAlife}}}$, with z being the number of proteins. We want to calculate the fraction of mRNA's that produces two or more proteins. The collection of all mRNA's is represented by $\int_0^\infty e^{-\frac{z \times t_{prot}}{t_{mRNAlife}}} dz$, and the collection of all mRNA's that produce two or more proteins is represented by $\int_2^\infty e^{-\frac{z \times t_{prot}}{t_{mRNAlife}}} dz$. The fraction of mRNA's that produces two or more proteins is then given by:

$$F_{protein \geq 2} = \frac{\int_2^\infty e^{-\frac{z \times t_{prot}}{t_{mRNAlife}}} dz}{\int_0^\infty e^{-\frac{z \times t_{prot}}{t_{mRNAlife}}} dz} \qquad (6)$$

Finally, we need to calculate the fraction of cells that has the correct background level of membrane proteins, meaning between 0-3. In an extensive experimental study it was shown that for most proteins in *E. coli* the copy number distribution over a population of cells follows a gamma distribution (Taniguchi *et al.*, 2010):

$$P(n) = \frac{n^{a-1} \times e^{-\frac{n}{b}}}{b^a \times \Gamma(a)} \qquad (7)$$

Here $P(n)$ is the probability as a function of n, n is the number of proteins, and for low protein copy numbers a is the number of bursts per cell cycle and b the mean number of proteins per burst. $\Gamma(a)$ is the gamma function of a. (Note that the gamma distribution we used before, equation 1, has a somewhat different form but represents the exact same distribution.) To get the fraction of cells that under steady state production has the correct background level of membrane proteins, we divide the number of cells that have this particular background level by the total number of cells. The total number of cells is represented by $\int_0^\infty \frac{n^{a-1} \times e^{-\frac{n}{b}}}{b^a \times \Gamma(a)} dn$, and the number of cells with a low background (i.e. 0 membrane proteins) is represented by $\int_0^1 \frac{n^{a-1} \times e^{-\frac{n}{b}}}{b^a \times \Gamma(a)} dn$. The ratio of these gives the fraction of cells with the correct background:

$$F_{background} = \frac{\int_0^1 \frac{n^{a-1} \times e^{-\frac{n}{b}}}{b^a \times \Gamma(a)} dn}{\int_0^\infty \frac{n^{a-1} \times e^{-\frac{n}{b}}}{b^a \times \Gamma(a)} dn} \qquad (8)$$

The number of bursts per cell cycle, a, is given simply by the number of mRNA's produced per unit time, $\frac{1}{t_{mRNA}}$, multiplied by the cell cycle time, $t_{cellcycle}$ (equation 3a). To calculate the mean number of proteins produced per burst, b, we multiply the number of proteins, z, by the probability of obtaining that particular copy number $e^{-\frac{z \times t_{prot}}{t_{mRNAlife}}}$. Doing this for all possible number of proteins we get $\int_0^\infty z \times e^{-\frac{z \times t_{prot}}{t_{mRNAlife}}} dz$. After normalisation, so that the probabilities add up to 1, we obtain equation 3b.

133

Plugging in the numbers

We have five parameters in the model: t_{obs}, $t_{cellcycle}$, $t_{mRNAlife}$, t_{prot}, and t_{mRNA}. The observation time t_{obs} is set by how we perform the experiment. When we have only a small number of frames we can observe (<10), and when we need a time resolution of a couple of seconds, we put t_{obs} = 10 s. The cell cycle time and the mRNA life time are also known, so $t_{cellcycle}$ = 3600 s and $t_{mRNAlife}$ = 180 s (rough estimate from *E. coli*) (Bernstein *et al.*, 2002). Note that we don't know $t_{mRNAlife}$ precisely for our mRNA in *L. lactis*, which is not critical because we deal only with the ratio $\frac{t_{prot}}{t_{mRNAlife}}$, which we can determine experimentally with the bleaching biogenesis experiment described above. We are left with t_{prot} and t_{mRNA}, which we would like to vary to minimise the number of cells we need to observe. The fewer cells we need to observe the more data can be obtained in a fixed amount of time. Also with more actual production events per cell the noise, e.g. of spurious binding events, plays a smaller role. Plugging in the values described above and taking ranges of t_{prot} = 10-180 s and t_{mRNA} = 600-10800 s we obtain Figure 11A. We find a maximum of 0.00045 at t_{prot} = 50 s and t_{mRNA} = 5400 s. In the equation derived above we have assumed that we need 0 membrane proteins in the background. This is probably too strict. When we take a maximum of 3 membrane proteins in the background we get Figure 11B, with a maximum of 0.0012 at t_{prot} = 90 s and t_{mRNA} = 1800 s. Taking rough estimates for t_{prot} and t_{mRNA} from the bleaching biogenesis experiments, 180 s and 3000 s, we find that we are far from the optimum fraction of useful cells.

The useful fraction of cells increases linearly with t_{obs}. Increasing t_{obs} of course also reduces the number of times we can measure a field of cells. The number of spurious binding events also increases linearly with t_{obs}. However, it takes some time to find a good field of view and do the focussing. Increasing t_{obs} would reduce that time and thus allow us to collect more data per unit time. We can increase t_{obs} by stabilising a fluorophore with Trolox (van der Velde *et al.*, 2016), when labelling from the outside. We can also increase t_{obs} by having more than one binding site on the target membrane protein.
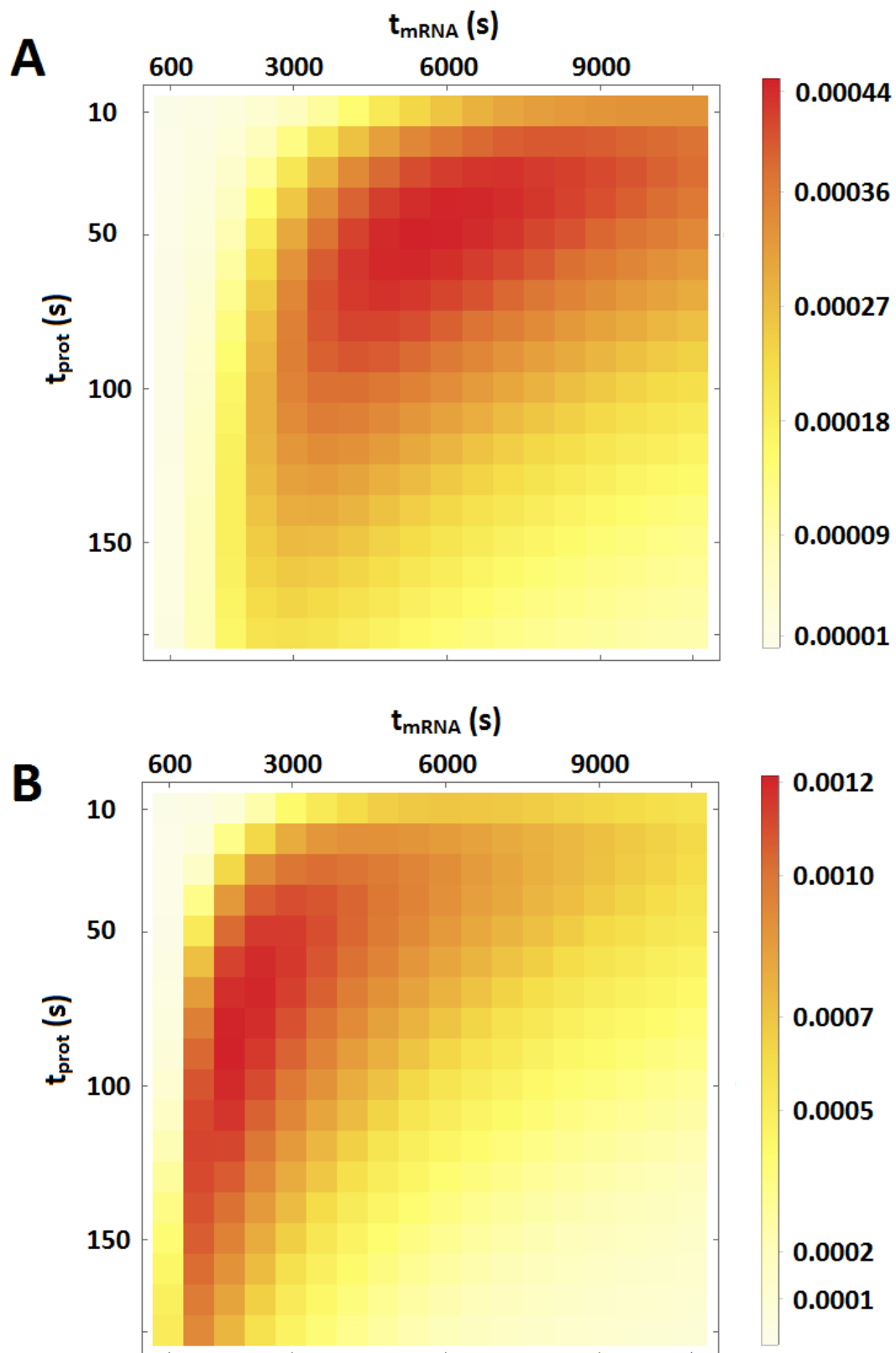
Figure 11: Prediction of the useful fraction of cells for localisation labelling depending on t_{mRNA} and t_{prot}. (A) The useful fraction of cells when we discard measurements on cells that start with a non-zero number of the target membrane protein. (B) The useful fraction of cells when we discard measurements on cells that start with >3 copies of the target membrane protein.

Having many labels

When performing a localisation labelling experiment on mNeonGreen containing cells we found that the signal was overwhelmed by noise. This noise is probably a combination of spots moving in and out of focus, blinking and bleaching of the fluorophore, fluctuations in the positions of cytoplasmic fluorophores and camera noise. Then there is also the issue that there will always be a fraction of fluorophores that is not fluorescent. This is caused by the stochastic nature of fluorophore maturation, long maturation times, and the need for replenishing fluorescent proteins because of dilution due to cell growth. All of the above mentioned issues are less of a problem when you use multiple binding sites on the membrane protein rather than one.

Recently multiple binding sites were used to visualise the production of cytoplasmic proteins in human cells (Yan *et al.*, 2016). In this study the target protein, Kif18b, has 24 SunTag peptides fused to its N-terminus (SunTag$_{24x}$-Kif18b). Each of the SunTag peptides can be bound by the protein fusion scFv-GFP. Additionally the mRNA used to produce SunTag$_{24x}$-Kif18b has 24 recognition sites for the protein PP7-mCherry$_{3x}$. Expression of these three protein fusions in one cell allows for the detection of the production of proteins at mRNAs. The introduction of a prenylation (CAAX) sequence to PP7-mCherry$_{3x}$ caused the mRNA's to localise to the plasma membrane so that individual mRNA's could be followed for > 1 h. By making use of translation targeting inhibitors and applying a mathematical model to spot intensities, the authors were able to determine translation initiation rate, translation elongation rate, and the number of ribosomes on the mRNA.

There are many differences between these measurements in human cells and the measurements we want to do in bacteria. For example: the mRNA lifetime in eukaryotes is much longer, there is more space in the human cells for the visualisation of spots, and they used a soluble protein whereas we want to study membrane proteins. Nonetheless they do demonstrate that protein production can be visualised on a single molecule level in cells when you use many labels on the same target protein, and we can adapt their method to work in our situation.

We decided to try localisation labelling in *E. coli* with the following pairs of constructs: (1) scFv-mNeonGreen and LacY-Suntag$_{20x}$, and (2) mNeonGreen-SYNZIP1 and LacY-SYNZIP2$_{10x}$. In both pairs the fluorescent protein is expressed from an L-rhamnose inducible promoter and the membrane protein is expressed from an L-arabinose inducible promoter. We collected fluorescence images both in the absence and in the presence of L-arabinose. For *L. lactis* we used LacSΔIIA-SunTag$_{20x}$, expressed from a Nisin inducible promoter, and scFv-mNeonGreen, expressed from the CP1 promoter. For *L. lactis* we also recorded fluorescence images in the presence and absence of inducer. These strains were imaged with a fluorescence microscope. For all strains, and conditions, tested we saw fluorescence in the cytoplasm. However, we did not see the expected bright spots. We did not investigate the reason for this lack of spots.

Conclusion

We have attempted to measure the time between the production events of individual membrane proteins that arise from the same mRNA. In the effort we have tried different fluorescent proteins (YPet, mCardinal, mNeonGreen), different interaction domains (SYNZIP1-SYNZIP2, Barnase-Barstar, ColcinE9-Im9), different organisms (*L. lactis*, *E. coli*), different promoters (ACE, CP1, etc.), and different numbers of binding sites on the membrane protein. We also attempted to follow the production of the membrane protein OpuABC by labelling from the outside. In the end we were unsuccessful in our attempts. The reasons were excessive noise in the case of single binding sites, and fluorophore binding

to the cover slide for labelling from the outside. For the multi-binding site experiments we did not see localisation, for, as of yet, unknown reasons. A great bottleneck in getting the method to work was our inability to quickly change expression levels. This was in part due to the unavailability of suitable promoter systems that were tight enough to work at single molecule levels of protein. There was also a failure of prudence on our side. Perhaps this is coloured by hindsight but some of the problems that we discovered experimentally later on in the project, could have been detected at the start by doing the appropriate simulations. Simulations that we had already planned to do, and that we had the necessary information for. This would have led us to go immediately to fast interacting proteins and to using multiple binding sites per membrane protein rather than one.

Methods

The dependence of time resolution on k_{on} and fluorophore copy number

All simulations were performed in Smoldyn (Andrews *et al.*, 2010) and analysis was done with Mathematica. We set up a spherocylinder with dimensions that approximate a *L. lactis* cell size, i.e. a length of 1.5 µm and a width of 1 µm. A single membrane protein is allowed to diffuse over the membrane with D = 0.03 µm^2/s. Cytoplasmic proteins diffuse with a D of 10 µm^2/s. We varied two parameters: (1) the number of cytoplasmic proteins, and (2) the association rate constant, k_{on}, with which the cytoplasmic molecules bind the membrane protein. We performed simulations with three parameter sets: (i) with 100 molecules and k_{on} is 10^7 M^{-1}s^{-1}, (ii) with 100 molecules and k_{on} is 10^9 M^{-1}s^{-1}, and (iii) with 1000 molecules and k_{on} is 10^9 M^{-1}s^{-1}. We started the simulations with all molecules randomly positioned in their designated compartment (i.e. membrane or cytoplasm). Then the molecules were allowed to diffuse and interact. The simulation time step was 10^{-5} s. When the interaction occurred the simulation was stopped and the time was recorded. For parameter set (ii) we performed 1000 simulations, and for parameter sets (i) and (iii) we performed 100 simulations.

The bleaching biogenesis experiment

All measurements were done on *Lactococcus lactis* strain NZ9000 (Linares *et al.*, 2010) containing a pNZ8048 plasmid with LacSΔIIA-YPet (Mika *et al.*, 2014) under the control of the nisA promoter (Kuipers *et al.*, 1998).

For the experiment with 1 min time between frames we used a chemically defined medium (CDM) described in (Mika *et al.*, 2014) where it is referred to as CDM^{RP}. We add glucose (1 % w/v) as a carbon source and chloramphenicol (5 µg/mL) to maintain the plasmid. A reaction tube with 4 mL CDM was inoculated with *L. lactis* cells from a -80 °C glycerol stock. This culture was incubated at 30 °C overnight without shaking. The next morning the overnight culture was used to inoculate a new culture to an OD$_{600}$ of 0.1. It was incubated at 30 °C without shaking until reaching an OD$_{600}$ of about 0.4.

To prepare the sample for microscopy a couple of µL of culture was deposited on a cover slide. Before putting another cover slide on top, two pieces of cover slide were placed in between to increase the distance between the two cover slides. Before use the cover slides were treated as follows. First we cleaned them by sonicating for 1 h in 5 M KOH, rinsing 10 times with MilliQ and blowing off the remaining MilliQ with pressurized nitrogen. We then deposited the slides in acetone that contained 2 % v/v (3-aminopropyl)triethoxysilane. We incubated for 5 min at room temperature, removed the acetone and APTES and rinsed the slides 10 times with MilliQ. Again, the remaining MilliQ was blown off with pressurized nitrogen.

The sample was then deposited on the microscope stage, which was preheated to 30 °C. At every time point we first recorded a bright field image, then three images with 34 ms exposures in the 514 channel, and finally a 5 s long exposure in the 514 nm channel. This sequence is repeated every minute for a total of 30 times. We determined the number of fluorescent spots per cell and in each frame by eye.

For the experiments with 3 min and 5 min time between frames we used a chemically defined medium (CDM) described in (Mika *et al.*, 2014) where it is referred to as CDM^{RP}, with one difference, here we also added L-proline (0.68 g/L, final concentration). We also add glucose (1 % w/v) as a carbon source and chloramphenicol (5 µg/mL) to maintain the plasmid. A reaction tube with 4 mL CDM was inoculated with *L. lactis* cells from a -80 °C glycerol stock. This culture was incubated at 30 °C overnight without shaking. The next morning the overnight culture was used to inoculate a new culture to an

OD$_{600}$ of 0.1. It was incubated at 30 °C without shaking until reaching an OD$_{600}$ of about 0.4. Cells were then deposited in a flow cell. The flow cell was made as follows. Cover slides were cleaned by sonicating 60 min in 5 M KOH. They were rinsed ~10 times with MQ and dried by blowing of the water with pressurized nitrogen. The cover slides were attached to a quartz slide (containing two holes for passing through growth medium) with two sided sticky tape. The cover slide quartz piece combo was incubated in the oven at 110 °C for 60 min while clamping it between two object glasses. Afterwards tubes (PE-60) were attached to the holes and the chamber was sealed with epoxy glue.

Cells were loaded into the flow cells by pulling the culture through with a syringe. Fresh CDM (without chloramphenicol) was supplied throughout the experiment by a syringe pump. At every time point we first recorded a bright field image, then three images with 100 ms exposures in the 514 nm channel, and finally a 3 s long exposure in the 514 nm channel. The time between frames is 3 min (20 time steps) and 5 min (12 time steps), with a total observation time of 60 min. We determined the number of fluorescent spots per cell and in each frame by eye.

First iteration of localisation labelling: YPet-SZ1 and LacSΔIIA-SZ2

All measurements were done on *Lactococcus lactis* strain NZ9000 (Linares *et al.*, 2010). It contained either the plasmids pNZ8048_pNISA_LacSΔIIA-SZ2 and pIL252_**CP1**_Ypet-SZ1, or the plasmids pNZ8048_pNISA_LacSΔIIA-SZ2 and pIL252_**CP23**_Ypet-SZ1. As control we used the same strain described above for the bleaching method (expressing LacSΔIIA-YPet). Between YPet and SZ1, and LacSΔIIA and SZ2 there is linker with sequence TRESGSIGS. The source for YPet is (Nguyen *et al.*, 2005). The pIL252 vector that we used (in all experiments described in this chapter) has a transcription terminator added behind the gene of interest.

We used a chemically defined medium (CDM) described in (Mika *et al.*, 2014) where it is referred to as CDM^{RP}, with one difference, here we also added L-proline (0.68 g/L, final concentration). We also add glucose (1 % w/v) as a carbon source. To maintain the plasmids we used chloramphenicol (5 μg/mL) and erythromycin (5 μg/mL) for the double plasmid strains and only chloramphenicol (5 μg/mL) for the LacSΔIIA-YPet strain. Cultures were inoculated by taking some cells from a -80 °C glycerol stock and putting them in growth medium. The cultures were incubated overnight at 30 °C without shaking. The next morning the cultures were used to inoculate a new culture to an OD$_{600}$ of 0.1. The cultures were grown (with dilutions when necessary for timing) to an OD$_{600}$ of ~0.4 and then used for microscopy. Induction was done by adding 1:1000 nisin solution at an OD$_{600}$ of 0.2, which is roughly an hour before microscopy. The nisin solution is the filtered culture supernatant of *Lactococcus lactis* strain NZ9700 which produces the nisin.

After an OD$_{600}$ of 0.4 is reached, cells were loaded into a flow cell. Throughout the measurements fresh CDM was supplied with a syringe pump. For the double plasmid strains antibiotics are present in the CDM throughout the measurements. For LacSΔIIA-YPet the experiment was done with or without antibiotics, showing the same results. The flow cells were prepared as described in "The bleaching biogenesis experiment", except that the cover slide was treated differently. Cover slides were cleaned by sonicating 30 min in 5 M KOH. They were rinsed 10 times with MQ and dried by blowing of the water with pressurized air. The cleaned slides were incubated in acetone containing 2 % dichlorodimethylsilane (≥99.5 %) for 5 min with gentle shaking. Slides were then again rinsed 10 times with MQ and dried by blowing of the water with pressurized air.

For all strains and conditions we first recorded bright field image to determine the position and focus of the cells. Immediately after we recorded a series of images in the 514 nm channel to determine the position and dynamics of the YPet molecules. For the CP23 strain the 514 images were recorded with

an exposure time of 500 ms. For the CP1 strain the 514 images were recorded with an exposure time of 50 or 500 ms. For the LacSΔIIA-YPet strain the 514 images the exposure time was 50 ms.

Faster protein-protein interactions and more stable fluorescent proteins

All measurements were done on *Lactococcus lactis* strain NZ9000 (Linares *et al.*, 2010). This strain carried one of the following complements of plasmids: (1) pNZ8048_pNISA_LacSΔIIA-ColicinE9 and pIL252_**CP1**_mNeonGreen-Im9, (2) pNZ8048_pNISA_LacSΔIIA-ColicinE9 and pIL252_**CP23**_mNeonGreen-Im9, or (3) only pIL252_**CP1**_mNeonGreen-Im9. Again we have the linker with sequence TRESGSIGS between all protein domains. ColicinE9 (we use only the DNase domain) and Im9 are proteins from *E. coli* that form a very tight complex and do so with a high k_{on} (Wallis *et al.*, 1995).

Growth of all three strains was done as described above for localisation labelling with YPet-SZ1 and LacSΔIIA-SZ2. (With the appropriate changes in the use of antibiotics.) We induced with 1:1000 nisin.

Cells are used when an OD_{600} of 0.4 is reached. For the localisation labelling experiment with the plasmids pNZ8048_pNISA_LacSΔIIA-ColicinE9 and pIL252_**CP1**_mNeonGreen-Im9 we loaded the cells into a flow cell (as above for YPet-SZ1 and LacSΔIIA-SZ2). For the other experiments we put the cells on a cover slide (in some cases after concentrating the cells by centrifugation) and put an object slide on top for stability. Cover slides and object slides were cleaned, before use, by sonicating 1-2 h in 5 M KOH. They were rinsed 10 times with MQ and dried by blowing of the water with pressurized air. In contrast to the experiments with YPet-SZ1 and LacSΔIIA-SZ2 we did not treat the slides further. Cells stick well on cleaned slides and without the dichlorodimethylsilane there is less background fluorescence.

To test our ability to visualise spots and to test the binding between ColicinE9 and Im9 (with the induction of LacSΔIIA-ColicinE9) for both CP1 and CP23 we first recorded a bright field image to localise the cells and to check the focus. Then we recorded a fluorescence image with an exposure time of 100 ms in the 514 nm channel. For the localisation labelling experiment with pNZ8048_pNISA_LacSΔIIA-ColicinE9 and pIL252_**CP1**_mNeonGreen-Im9 and the background localisation control, pIL252_**CP1**_mNeonGreen-Im9, we first took a bright field image. Then 10 fluorescence images 5 s apart with an exposure time of 100 ms. Finally, we recorded another bright field image, to check whether cells had moved. The fluorescence images were recorded in the 514 nm channel.

To test whether mCardinal can be used for localisation labelling we fused it to the C-terminus of LacSΔIIA. This construct was put on a pNZ8048 plasmid and expressed with a nisin inducible promoter (nisA). The fluorescence microscopy was carried out in the same way as for mNeonGreen.

Localisation labelling in *Escherichia coli*

For the experiment without induction we used *Escherichia coli* MG1655 containing the plasmids pACYC_rhaBAD_Barstar-mNeonGreen and pBAD_araBAD_LacY-Barnase; the *rhaBAD* promoter is inducible with L-rhamnose and the *araBAD* promoter with L-arabinose. The experiments in which the expression of membrane protein (LacY) was induced were carried out on *E. coli* MG1655 containing (1) the plasmids pACYC_rhaBAD_Barstar-mNeonGreen and pBAD_araBAD_LacY-Barnase, or (2) the plasmids pACYC_rhaBAD_Im9-mNeonGreen and pBAD_araBAD_LacY-ColicinE9. Barnase and Barstar are proteins from *Bacillus amyloliquefaciens* that form a tight interaction with a high k_{on} (Schreiber *et al.*, 1993). We used the same linker between domains as for the other constructs (described above). LacY is a lactose transporter from *E. coli*.

Glycerol stocks from a -80 freezer were used to inoculate EZ medium containing 5-30 µg/mL chloramphenicol and 100 µg/mL ampicillin to maintain the plasmids. For the experiment *without* induction we used glucose (0.2 % w/v) as a carbon and energy source, for the experiments *with* induction we used glycerol (0.2 % v/v). The cultures were incubated overnight at 37 °C with 200 rpm shaking. The next day 4-16 µL overnight culture was used to inoculate a new culture (same medium composition). In case of the experiments with induction the cultures were incubated till an OD_{600} of 0.2 and then induced with 0.1 % (w/v) L-arabinose. The cultures were incubated further till an OD_{600} of 0.4-0.6, after which they were used for microscopy.

A few microliters of culture was deposited on a cover slide and an object slide was put on top for stability. Before use both the cover and object slides were treated with 5 M KOH for 1.5 h in a sonication bath. After which the KOH was rinsed off with milliQ and the slides were dried by blowing off the milliQ with pressurized air. The cover slides were further treated with a solution of 2 % (3-aminopropyl) triethoxy silane in acetone for 5 min which was rinsed off with milliQ and the slides were dried with pressurized air. Microscopy was performed at room temperature (~20 °C). First a bright field image was taken and immediately after that we recorded a series of fluorescence images in the 514 nm channel with a 100 ms exposure time.

ACE expression system for *Lactococcus lactis*

We used a *Lactococcus lactis* NZ9000 strain containing the pNZ8048 plasmid with LacSΔIIA-YPet under the control of the ACE promoter (Linares *et al.*, 2015). Culturing and microscopy were performed as mentioned above for the *L. lactis* mNeonGreen strains. With one difference, in one of the experiments with the ACE promoter we removed lysine and arginine from the growth medium.

Labelling from the outside

The strain used is *Lactococcus lactis* NZ9000 containing a pNZ8048 vector with OpuABC-Barnase under the control of a nisin inducible promoter. The complementary interaction protein barstar has the TwinStrepII peptide fused to its C-terminus. It is produced from a pBAD24 plasmid in Escherichia coli MC1061.

The production of barstar was done as follows. On day one we inoculated 10 mL LB (with 100 µg/mL ampicillin) with the *E. coli* producing barstar-TwinstrepII, from a -80 °C glycerol stock. This culture was incubated overnight at 37 °C with 200 rpm shaking. The next day the 10 mL culture was used to inoculate 1 L of LB (with 100 µg/mL ampicillin). This culture was incubated at 37 °C with 200 rpm shaking until an OD600 of 0.5 was reached. At this point we added 100 µL L-arabinose (20 % (w/v) stock solution in milliQ). After incubating for another 3 h we centrifuged the culture to obtain the cells (Avanti J-20 XP, 9.1000 rotor, 4 °C, 6000 RPM, 15 min). The supernatant was discarded and the pellet was resuspended in buffer (50 mM sodium phosphate, 300 mM NaCl, 10% glycerol, pH 8.5). The resuspended cells were frozen in liquid nitrogen and stored at -80 °C.

To purify barstar, the cell suspension was thawed and 0.1 mg/mL DNase and 1 mM PMSF were added. The cells were broken by sonication and the cell debris was removed by centrifugation (Optima MarE ultracentrifuge, MLA-80 rotor, 4 °C, 80000 g, 30 min). The supernatant (25 mL) was mixed with 1 mL streptavidin column material (0.5 mL bed volume). This suspension was incubated at 4 °C while mixing constantly for 1.5 h. The column was washed with buffer (same as above) and then the barstar was eluted from the column in fractions of 0.5 mL with buffer containing 2.5 mM desthiobiotin. We determined protein concentrations in a Nanodrop spectrophotometer to determine which fraction contained the most protein. This fraction was mixed with 0.5 mL buffer containing 0.2 mg sulfocyanine-5 NHS ester. The mixture was incubated at 4 °C with constant mixing for 2 days. After

the incubation the mixture was put on a NAP-10 column (Illustra NAP-10 column from GE Healthcare) to remove free label. The labelled barstar was eluted with an ammonium acetate buffer (0.2 M ammonium acetate, 1 mM EDTA, 2 mM 2-mercaptoethanol, 10 % glycerol, pH 8.0). Again we determined the most concentrated elution fraction using the Nanodrop spectrophotometer and loaded this fraction on a gel filtration column (Superdex 200 Increase 10/300 GL). We eluted with the same ammonium acetate buffer as was used above. The fraction bound was determined from the absorptions at 280 and 637 nm. We stored aliquots of the labelled barstar on ice (to be used the next day) or at -80 °C after freezing in liquid nitrogen.

We put solution with labelled barstar under the microscope on: cleaned cover slides, cover slides incubated for 15 min in polylysine solution, and cover slides treated with (3-aminopropyl) triethoxy silane or dichlorodimethyl silane (prepared as explained above). We did these experiments in the presence or absence of *L. lactis* cells expressing OpuABC-Barnase, and at different concentrations of labelled barstar. All measurements were performed in the 637 nm channel.

Having many labels

For the experiments with *E. coli* we used strain MG1655 containing either (1) pACYC_(rhaBAD)_scFv-GCN4-mNeonGreen-GB1 (called scFv-mNeonGreen) and pBAD_(araBAD)_LacY-Suntag$_{20x}$ (called LacY-Suntag$_{20x}$), or (2) pACYC_(rhaBAD)_mNeonGreen-SYNZIP1 (called scFv-SYNZIP1) and pBAD_(araBAD)_LacY-SYNZIP2$_{10x}$ (called LacY-SYNZIP2$_{10x}$).

Cell growth was done as described in "Localisation labelling in *Escherichia coli*". We used glucose as a carbon source for both strains, and for the experiments with and without L-arabinose induction. We also observed cells in the presence and absence of L-arabinose for strain number 1 in which we used glycerol instead of glucose. L-arabinose induction was done 1-2 h before microscopy.

Microscope slide preparation and the acquisition of images was also done as described in "Localisation labelling in *Escherichia coli*".

For *L. lactis* we used strain NZ9000 containing pNZ_(CP1)_scFv-GCN4-mng-GB1 (called scFv-mNeonGreen) and pIL252_(pNisA)_LacSΔIIA-SunTag (called LacSΔIIA-SunTag).

Cells were grown as described in "First iteration of localisation labelling". The slides were prepared as described in "Faster protein-protein interactions and more stable fluorescent proteins", with the exception that we did not use flow cells but just added some cell suspension to a cover slide. Induction with Nisin was done 1.5 h before microscopy. Microscopy was done essentially as described in "Faster protein-protein interactions and more stable fluorescent proteins".

References

Alsallaq, R., and Zhou, H. (2008) Electrostatic rate enhancement and transient complex of protein-protein association. *Proteins* **71:** 320-335.

Andrews, S.S., Addy, N.J., Brent, R., and Arkin, A.P. (2010) Detailed Simulations of Cell Biology with Smoldyn 2.1. *PLoS Comput Biol* **6:** e1000705.

Bernstein, J., Khodursky, A., Lin, P., Lin-Chao, S., and Cohen, S. (2002) Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proc Natl Acad Sci U S A* **99:** 9697-9702.

Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S., Singer, R., and Long, R. (1998) Localization of ASH1 mRNA particles in living yeast. *Mol Cell* **2:** 437-445.

Biemans-Oldehinkel, E., Doeven, M., and Poolman, B. (2006) ABC transporter architecture and regulatory roles of accessory domains. *FEBS Lett* **580:** 1023-1035.

Blitzstein, J.K., and Hwang, J. (2014). In *Introduction to probability*. Boca Raton, United States: CRC Press, pp. 356-360.

Chong, S., Chen, C., Ge, H., and Xie, X.S. (2014) Mechanism of Transcriptional Bursting in Bacteria. *Cell* **158:** 314-326.

Chu, J., Haynes, R.D., Corbel, S.Y., Li, P., Gonzalez-Gonzalez, E., Burg, J.S., *et al.* (2014) Non-invasive intravital imaging of cellular differentiation with a bright red-excitable fluorescent protein. *Nature Methods* **11:** 572-578.

Driessen, A.J.M., and Nouwen, N. (2008) Protein translocation across the bacterial cytoplasmic membrane. *Annu Rev Biochem* **77:** 643-667.

Golding, I., Paulsson, J., Zawilski, S., and Cox, E. (2005) Real-time kinetics of gene activity in individual bacteria. *Cell* **123:** 1025-1036.

Jensen, P., and Hammer, K. (1998) The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. *Appl Environ Microbiol* **64:** 82-87.

Kondo, J., and Mckay, L. (1982) Mutanolysin for Improved Lysis and Rapid Protoplast Formation in Dairy Streptococci. *J Dairy Sci* **65:** 1428-1431.

Konopka, M.C., Sochacki, K.A., Bratton, B.P., Shkel, I.A., Record, M.T., and Weisshaar, J.C. (2009) Cytoplasmic Protein Mobility in Osmotically Stressed Escherichia coli. *J Bacteriol* **191:** 231-237.

Kuipers, O., de Ruyter, P., Kleerebezem, M., and de Vos, W. (1998) Quorum sensing-controlled gene expression in lactic acid bacteria. *J Biotechnol* **64:** 15-21.

Kumar, M., Mommer, M.S., and Sourjik, V. (2010) Mobility of Cytoplasmic, Membrane, and DNA-Binding Proteins in Escherichia coli. *Biophys J* **98:** 552-559.

Le, T., Harlepp, S., Guet, C., Dittmar, K., Emonet, T., Pan, T., and Cluzel, P. (2005) Real-time RNA profiling within a single bacterium. *Proc Natl Acad Sci U S A* **102:** 9160-9164.

Li, G., and Xie, X.S. (2011) Central dogma at the single-molecule level in living cells. *Nature* **475:** 308-315.

Linares, D.M., Alvarez-Sieiro, P., del Rio, B., Ladero, V., Redruello, B., Cruz Martin, M., *et al.* (2015) Implementation of the agmatine-controlled expression system for inducible gene expression in Lactococcus lactis. *Microbial Cell Factories* **14:** 208.

Linares, D.M., Kok, J., and Poolman, B. (2010) Genome Sequences of Lactococcus lactis MG1363 (Revised) and NZ9000 and Comparative Physiological Studies. *J Bacteriol* **192:** 5806-5812.

Mika, J.T., Schavemaker, P.E., Krasnikov, V., and Poolman, B. (2014) Impact of osmotic stress on protein diffusion in Lactococcus lactis. *Mol Microbiol* **94:** 857-870.

Milo, R., Jorgensen, P., Moran, U., Weber, G., and Springer, M. (2010) BioNumbers-the database of key numbers in molecular and cell biology. *Nucleic Acids Res* **38:** D750-D753.

Nguyen, A., and Daugherty, P. (2005) Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nat Biotechnol* **23:** 355-360.

Poolman, B., Hellingwerf, K.J., and Konings, W.N. (1987) Regulation of the glutamate-glutamine transport system by intracellular pH in Streptococcus lactis. *J Bacteriol* **169:** 2272-2276.

Schreiber, G., and Fersht, A.R. (1993) Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. *Biochemistry* **32:** 5145-5150.

Shaner, N.C., Lambert, G.G., Chammas, A., Ni, Y., Cranfill, P.J., Baird, M.A.*, et al.* (2013) A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. *Nature Methods* **10:** 407-409.

Taniguchi, Y., Choi, P.J., Li, G., Chen, H., Babu, M., Hearn, J.*, et al.* (2010) Quantifying E-coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science* **329:** 533-538.

Thompson, K.E., Bashor, C.J., Lim, W.A., and Keating, A.E. (2012) SYNZIP Protein Interaction Toolbox: in Vitro and in Vivo Specifications of Heterospecific Coiled-Coil Interaction Domains. *ACS Synth. Biol.* **1:** 118-129.

van der Velde, J.H.M., Oelerich, J., Huang, J., Smit, J.H., Jazi, A.A., Galiani, S.*, et al.* (2016) A simple and versatile design concept for fluorophore derivatives with intramolecular photostabilization. *Nature Communications* **7:** 10144.

Wallis, R., Moore, G.R., James, R., and Kleanthous, C. (1995) Protein-protein interactions in colicin E9 DNase-immunity protein complexes. 1. Diffusion-controlled association and femtomolar binding for the cognate complex. *Biochemistry* **34:** 13743-13750.

Yan, X., Hoek, T.A., Vale, R.D., and Tanenbaum, M.E. (2016) Dynamics of Translation of Single mRNA Molecules In Vivo. *Cell* **165:** 976-989.

Yu, J., Xiao, J., Ren, X., Lao, K., and Xie, X. (2006) Probing gene expression in live cells, one protein molecule at a time. *Science* **311:** 1600-1603.

Chapter 6: The custodians of life's meaning

"The significance of our lives and our fragile planet is then determined only by our own wisdom and courage. We are the custodians of life's meaning. We long for a Parent to care for us, to forgive us our errors, to save us from our childish mistakes. But knowledge is preferable to ignorance. Better by far to embrace the hard truth than a reassuring fable." –Carl Sagan, in Pale Blue Dot, p. 54.

Some ideas come to me not as a crisp finished product, needing only the slightest tweaks before they can be implemented practically, but more as a tendency to ask certain questions. For years now I have been in the thrall of such a will-o'-the-wisp of the mind. Its subject can be stated roughly as diversity and possibility. I'll work this out in the second part of this chapter. First I'll have to lay some foundation about what I think science is and how it relates to society.

Science

Let me start off with a strong, though in my eyes truthful, claim. The greatest discovery of science is naturalism! Which essentially means that there is no magic; everything is governed by natural laws. And I really do mean everything; not only is it true for the natural sciences but also for the social sciences, history, ethics and esthetics. The world is one. We need to strive to unify our knowledge of it, we need, to use E.O. Wilson's term, consilience (Wilson, 1998). And I would also add mathematics and philosophy to the mix. Truth in all of these subjects is found only through logic and/or observation. There seems to be an attitude among people, not just lay people but scientists too, to think that science is a tool you can pick up or leave be; a coat that you wear at work but put down when going home. In reality the laws of nature do not switch off, or stop to be relevant when moving into a different subject. The game is rigged, you are playing anyway. The only difference between people (in this context) is in how well they play this game. And you can only play it well if you accept naturalism.

I find my engagement with science to be deeply emotionally moving. My main interest in knowing about the world is esthetic. The world, or nature, when examined in detail is extremely beautiful. And, in that there always seems to be more to find out, for both individual and humanity at large, it also supplies purpose. I believe that this is not just true of me alone, or of a small group of scientifically minded people, but that it rings true for most of humanity. Science is a cultural pursuit as well as a quest for better technology. I'm certainly not the first, or only person, to hold this view. Richard Feynman stated (Feynman *et al.*, 2006):

"Finally, may I add that the main purpose of my teaching has not been to prepare you for some examination – it was not even to prepare you to serve industry or the military. I wanted most to give you some appreciation of the wonderful world and the physicist's way of looking at it, which, I believe, is a major part of the true culture of modern times."

And Carl Woese said it as follows (Woese, 2005):

"What was formally recognized in physics needs now to be recognized in biology: science serves a dual function. On the one hand it is society's servant, attacking the applied problems posed by society. On the other hand, it functions as society's teacher, helping the latter to understand its world and itself. It is the latter function that is effectively missing today."

Something that I'll come back to is the generation, by science, of interesting thoughts and experiences. And you should ask yourself, is it really true that you experience your maximum wellbeing when just sitting about chatting with your friends about the weather, the sandwich you're eating at that moment, or any other superficial thing? Or is it perhaps the case that you're culturally naïve due to a lack of knowledge and understanding of the world. That you perhaps need to spend a good deal of

your free time suffering through difficult topics to truly flourish. People can be wrong about what they think they like.

Science derives its worth not from what you do in the here and now. In fact the daily goings on of scientists is nigh pointless. But when you connect science in the now to its known past and uncertain future it becomes a true marvel. Only in the long run do the incremental advances of science add up to meaningful technological change and appreciation of nature. If you do science for the right reasons you have to study its history, and anticipate its future, to find your motivation. There is a peculiarity to the history of science that is not so obvious in the history of other things. It has an arrow! We get better in our dealings with nature. This is another source of meaning and purpose. And another reason for doing science even if it didn't grant you any technological progress.

We are faced with a single reality, which means that everything we find out needs to fit together. So either we have to show that different phenomena can be treated separately, or we have to show that they are in fact aspects of one phenomenon. We should be mindful of this when dabbling in biochemistry, cell biology or any other subject. These subjects connect up and down, and how they do so should be of at least some concern to you. There is a nice illustration of this up and downness in the book "Simulating the Physical World" (Berendsen, 2007). The author starts off at the bottom with relativistic quantum dynamics and moves upward via, among other things, molecular dynamics and Brownian dynamics all the way up to steady-flow fluid dynamics. At each step it is indicated how one level relates to the next. Yet, up and down doesn't cover every possibility. There are properties that seem to come out of the blue and that may couple very disparate topics. Take for example natural selection. Where does it appear in the hierarchy of things? Or take a random walk. A concept that shows up in diffusion but also in the conformation of polymers, or in navigating a fitness landscape. There is also a more fundamental reason for minding the context of a discipline. Understand something is to make it general. So a deep understanding by necessity crosses disciplines.

The notion most scientists seem to have of discovery seems to me at least somewhat mystical. As if there is a list, of things to be discovered, whose items you can somehow hit when you are pipetting in the lab. I think the situation is more complex. First off, there is no list with separate entries. You can discover something halfway. And many discoveries overlap to some extent. Second, discoveries are not just out there but also have something to do with us. A discovery is a representation of the world, and some representations are more effective than others. Finding a different representation of the same phenomenon should still count as a discovery. Third, there is a difference between having a discovery represented in a paper and having it represented in an individual. In a paper you could store loads of facts and ungainly descriptions. For individuals, that need to navigate through the world as quickly as possible, you would want the representations to be sleeker and more intuitive.

My scientific interests are theoretical and strongly bound to the individual. Theoretical approaches such as molecular dynamics or bioinformatics are useless to me because they don't deal with the structure of biology per se. I want to put the world, or as much of it as I possibly can, into my head. I'm not interested in anything in particular, only in everything. It is a mistake to think that you can just go out into the world and see everything. Understanding allows you to see so much more. It allows you to ask questions previously inaccessible to you, understanding causes spurious whatever's to be being transformed into deep mysteries. Having obtained a new understanding you walk around and the old world is no more, transformed into something more comprehensible, yet with a greater reach. Here again we can see the connection with culture because a greater reach includes a greater set of experiences. In effect your understanding is the ability to navigate the world. This problem of navigation is interesting also in relation to another curious fact: you don't know what you know! Your knowledge has to be pulled out of you either by a previous thought or by an occurrence in your

environment. Like the translation apparatus that your life depends upon, you are a tape reader and writer. Good theory, which deals with concepts, gives you the ability to pull better and more things out of your head, it is a generator of possibilities, an algorithm for novelty.

Years ago I lend someone my copy of *The selfish gene* by Richard Dawkins. The book deals with concepts in evolutionary biology. After reading the book the person I lend it to mentioned that he/she had liked it but wasn't sure whether it was true, i.e. whether the theories discussed in the book were verified by observation. I think this is a common response from experimental scientists or, at least, molecular biologists. And I imagine that most people regard this as a perfectly normal response. I do not. In fact, it strikes me as utterly bizarre. When you encounter ideas that you have never encountered before, the proper way to deal with them is to keep them in mind. To look at the world anew. It is preposterous to worry about their truthfulness at that very instant.

What I have done in this section is discuss how a scientific field, and the practice thereof, is embedded in society. When busy with our detailed projects we need to be mindful of other scientific fields, culture, history, and whatever else is out there. And to do that we need to consider the navigational ability/understanding of the individual. We also need to allow ourselves to develop the ability to navigate the world, by not interjecting too soon with requests for observations. In my eyes many scientists, and certainly most other people, are philistines in that they don't put themselves and their thoughts and actions into this broader context. Science in large part should be about not being a philistine, about being elite. Modesty is overrated, some people are better than others. This should be made explicit lest we take away one of the great gifts nature has bestowed upon us: our ability to improve and our realization of its necessity, both as individuals and as a society.

Biology

My aim with the second part of this chapter is best illustrated with a quote about the 19th century explorer and naturalist Alexander von Humboldt (Wulf, 2015):

"Towards the end of his life, Humboldt often talked about understanding nature from 'a higher point of view' from which those connections could be seen; the moment when he realized this was here, on Chimborazo. With 'a single glance', he saw the whole of nature laid out before him."

The Chimborazo is a tall mountain in the Andes, and "those connections" are the similarity between the gradation of plant types towards the peak of the mountain with the gradation of plant types from equator to the poles. The key point for us is the last sentence. It is my goal to capture all of life's diversity, all the possibilities of form and function, in a glance. Obviously this is metaphor and some parts of life may be so algorithmically complex that they evade simplification altogether. Yet it is my conviction that much more can be done to make biology smaller, so that it's most important parts can be fit into the head of a single human being.

Life does not exist. At least not as a unitary phenomenon, a spark that is there when you mix the right molecules together to form some sort of individual. Most people have a folk theory of life (Machery, 2012), which is a semi-coherent view that involves concepts such as reproduction, heredity, evolution, homeostasis, and metabolism. For every precise definition of life there seems to be the twilight case that doesn't quite fit in. Let's run through a few examples. Most viruses replicate yet have no metabolism of their own (Koonin, 2011). Some viruses, e.g. *Pandoravirus salinus*, have a large complement of genes (>2000) of which some encode enzymes suspected to interact with the host metabolism (Philippe *et al.*, 2013). Thus there appears to be somewhat of a gradient from viruses towards cells. You could almost see these bigger viruses as diffuse cytoplasm's that travel in capsids from cell to cell. There is also a gradient spanning DNA-bearing-organelles such as mitochondria (Lynch

et al., 2006) and what are considered symbiotic bacteria (Nakabachi *et al.*, 2006; Perez-Brocal *et al.*, 2006; McCutcheon *et al.*, 2011; Van Leuven *et al.*, 2014), and you can ask yourself: when do these endosymbionts stop being alive and start becoming an organelle? Then we have the fact that many individual organisms, that are moving about healthily, can't reproduce. There are many healthy individuals formed out of a hybridization between species that can't reproduce. One example is the mule (Smith, 1993a), which is cross between a male donkey and a female horse. Another example is a cross between *Triturus cristatus* and *Triturus marmoratus*, both of which are species of Newt (Smith, 1993b). An example of infertility much closer to home is the menopause, after which women lose the ability to reproduce. Should we take away from these women the rank of living? Then we have many examples of artifacts left by organisms. These objects, or phenomena, do not have homeostasis or the ability to reproduce, yet they cry out for explanation. Among other things we have the O_2 atmosphere; beaver lakes; termite mounds; deposits of oil, coccolithophores and diatoms; and remains of animals from carcasses to fossilized skeletons.

To attribute all of this to something singular strikes me as a logical mistake akin to thinking there is something north of the north pole. Instead I would argue that this phenomenon we call life is in fact a chimaera. A confluence of physical phenomena such as reproduction, heredity, homeostasis, metabolism, natural selection, and others. A more helpful view of life emanates from Carl Woese's work, and is described by the following (Goldenfeld *et al.*, 2011):

"The second consequence of a lack of fundamental understanding in biology is the failure to recognize that biology is a manifestation of evolution—not the other way around."

To put this somewhat differently; there is a process, called evolution, which makes artifacts. The evolutionary process is a pattern through time, not just in space. What we see around us, e.g. cells, viruses, termite mounts, and diatom deposits, is a slice out of the evolutionary process. And there is no particular reason why the elements, i.e. individuals, that make up this slice should be able to be grouped together neatly in some definition of life. And I submit to you that the insistence to do so is caused by a deep misunderstanding of what biology is, and is in fact a vestige of vitalism.

There is a general perception that molecular biology is fundamental to the rest of biology. That if only you would get enough crystal structures and molecular mechanisms solved, the rest of biology would follow from that immediately. In one sense I think molecular biology is fundamental; every state that an organism can be in is represented by a state at the molecular level. Despite this I believe the fundamental importance of molecular biology is greatly exaggerated. Let me give two examples from this thesis. (1) In chapter 1 I discussed how diffusion limitation of an association can be negated by other influences so that a higher level process that depends on this interaction is not diffusion limited in its rate. (2) In chapter 4 I mentioned that the exponential growth of a culture of cells can be described without any reference to the molecules that constitute those cells. I see another issue with molecular biology that is related to what I just discussed. Molecular biologists seem to come along only after the interesting stuff has been done. (Done by nature that is.) They typically just ask what a particular cell, or a particular protein within that cell, is like. Moreover, they ask this in a very shallow sense. They ask, what does it do? I would argue that there are deeper questions that complement this shallow one. (1) Really asking what a protein, for instance, is like, deals not just with what it is but also with what it could be. What makes a glycolysis enzyme different from the ribosome, or from a flagellin? (2) Asking why a cell or protein is the way it is, i.e. the evolutionary question.

Venture out into nature and you'll find all kinds of things, or phenomena. These things, at least at first, appear distinct; each having its own essence. Let me bring a whole set of these things before your eyes to activate your brain: *protein, replisome, genome, bolas spider, flying, protein production,*

transmembrane transport, spirochete, Bacillus subtilis, gradient, predator, energy, entropy, ribosome, gas vesicle, endoplasmic reticulum, nucleus. This list is not meant to be representative of what is present in the world; because, frankly, I wouldn't know how to do that. And that strikes me as an important problem that requires a solution. Continuing: *multifork replication, horizontal gene transfer, pangenome, diploid, Boltzmann distribution, natural selection, biogeography, cell wall, pellicle, swarming motility, genetic code, enzyme, folding, diffusion equation, reaction rate, territory, symbiont, virus.* Represented here are objects such as a gas vesicle but also principles like the Boltzmann distribution. All of these things are elements of the biological world. There is more: *crowding, morphology, neuron, halophile, hyper flagellation, mitochondrion, chemotaxis, meiosis, exponential decay, scales, holdfast, conjugation, contractile vacuole, organism, gene, plasmodium, mucus, immune system, niche, cristae, cell size, karyokinesis, soil, myoneme, sorocarp.* This world of biology is gigantic, which in my eyes is both wonderful and terrible. The final stretch: *mouth, diatom, eyespot, bloom, carboxysome, toxin, gliding motility, biofilm, pattern, dinoflagellate, helicase, sec translocase, genome size, posttranslational modification, metabolism, food storage, life cycle, colony, signal, gene expression, development, growth, nutrient, thermodynamic activity, dormancy, barnase, nuclear pore complex, processivity, conservation, shedding, developmental constraint, behavior, pilus, euglena, progenote.* I don't know about you but this variety of things annoys the bejeezus out of me. What are all these things? That is the prime question that rises in my mind when stepping through such a list. What are the relations between these phenomena? Could we predict the existence of these phenomena from first principles? The very fact that I can name them and that you can picture them suggests that these things have, in some sense, an existence of their own; independent of the substrate of elementary particles and forces that they owe their physical existence to.

At this stage I think it is useful to recall an element of my schooling at university that has profoundly influenced my thinking. When studying protein crystallography I encountered the concept of the space group. There are 230 space groups, and they describe the possible crystal morphologies (note that for proteins the number of space groups is 65 due to their chirality) (Van Holde et al., 2006). What struck me is that we know that there are 230, and not 229 or 231. Since then I have always wondered to what extent this could be applied to biology. Are, for example, plants, animals, and fungi, in some sense space groups of biology? What about primary producer, herbivore, and predator? Or, parasitism, commensalism, and mutualism? Or, ribosome, replisome, Sec translocase, and pyruvate dehydrogenase? There are two principles to consider in the 'catching' of the diversity of phenomena. (1) To what extent do we have in our biological literature, or in our brain, completed the list of phenomena? How do we generate them? (2) How are these phenomena related? Can we fill up, as it were, the space between these, at first sight, completely different things; so that our description of those parts of the world becomes unified? My interest is not just in what *is* in the world of biology, but also what could have been and what can be. In the same way that a coordinate is pointless without knowing about the axes, you can't understand life as it *is* if you don't know what it could have been. And if molecular biology is serious about its fundamentalness it has to tell us something about these axes in biology.

Luckily the challenge to capture life's diversity has not been neglected entirely. Rob Phillips and colleagues describe a whole set of physical principles underlying, and unifying to some extent, biology in the books *Physical Biology of the Cell* and *Cell Biology by the Numbers*. Brian Goodwin has described an interesting approach to understanding the development of organisms in *How the Leopard Changed its Spots*. Then we have the classic by D'Arcy Thompson, *On Growth and Form*, in which he describes physical principles underlying cell and tissue shape, the shape of horns and tusks, and the forms of skeletons of both microscopic cells and macroscopic vertebrates. Eugene Koonin discusses the possible set of replication-expression strategies that can be used by viruses in his book *The Logic of*

Chance. And, finally Richard Dawkins who summarizes life by looking at it from the perspective of the genes in *The Selfish Gene* and *The Extended Phenotype*. There is also an interesting story about the use of logical constraint on biological systems from the heyday of molecular biology, when all the elements of the central dogma where still in the process of being elucidated. The story comes from an interview with Sydney Brenner in which he describes the "don't worry hypothesis" (https://www.webofstories.com/play/sydney.brenner/57). The point is that if the system you study has a certain logic to it, you should not worry too much about the molecular details. The example given by Brenner is that of the replication of DNA. That it ought to happen and how it would work roughly was obvious from the structure of DNA and known principles of heredity. Yet it was not yet put in molecular terms how the two wound up chains could separate. Later on, enzymes were found to catalyze this separation: helicase and topoisomerase. The key point of this last example is that the central dogma has a logic to it that is somewhat independent of the molecular details.

Where a view of diversity and possibility gets most interesting is where core biological phenomena are concerned. So I will discuss in some detail the work of Carl Woese (Woese, 2002) on the evolution of translation, which *ipso facto* is the coming into existence of the genotype-phenotype connection and the modern cell itself. Translation as it exists today is a highly complex process consisting of on the order of 100 components many of which are highly dependent on each other; these components are the ribosome, aminoacyl tRNA synthetases, initiation factors, elongation factors, and termination factors. The ability of translation to produce proteins accurately, likely depends on this complexity. Let's start at the modern translation apparatus and walk back in time. As the translation apparatus becomes less complex it becomes less able to construct big and accurate protein sequences. Many of the activities in the modern cell can only be performed by complex proteins. As such, when walking back in time, these activities will be diminished and ultimately they disappear. The translation apparatus itself is also made, for a large part, out of proteins, which means that going back in time the ability of the translation apparatus to build more copies of itself also diminishes. Another implication of lack of precision in translation is that cellular componentry back in the day could not be as integrated as it is now. This lack of integration greatly increases the power of horizontal gene transfer to exchange genes, effectively leading to the absence of lineage and (stable) species. This state of life has been dubbed the progenote (Woese, 2002). Here a diversity view of life comes into play. I think the progenote state is best seen, as is the case for viruses, as different facets of the phenomenon of life; alien life forms if you wish. And our search for what life is should be focused on finding, and defining, more of such life forms; rather than searching for something monolithic. I do not view viruses and the progenote state as entirely separate from modern cellular life, but in many ways continuous with it; systematically related but described by different effective theories. Woese's view of the evolution of translation also poses a question for people studying proteins: what set of protein functions is possible at each level of translation precision? The origin of life makes no sense if you do not have a theory of diversity, an idea of what possibilities lie out there for the upstarts; indeed a theory of what those upstarts are. This view carries through to the rest of life. It makes no sense to study life in the absence of its possibilities. The view that a universal theory of biology should be about properties that are present in every individual organism or species strikes me as highly peculiar, because it rules out that diversity has anything to do with biology.

In the light of the previous I want to place some critical notes about two projects in molecular biology: the whole cell simulation and the synthesis of a cell. I do think that there is merit in both of these projects, but that this is overstated. Some time ago a whole cell model was published for the bacterium *Mycoplasma genitalium* (Karr *et al.*, 2012). This model integrates 28 different cellular processes, including translation, replication, metabolism, FtsZ assembly, and cytokinesis. The models of the cellular processes make use of all annotated gene products. The whole cell model simulates a

whole cell cycle and allows you to look under the hood while this is happening, so that you can for example follow, over time, the number of DnaA copies at the origin of replication, the chromosome copy number, and the dNTP concentration. I consider the effort to construct whole cell models as crucial for ironing out the quantitative details of how a particular cell functions. And in some sense it is the culmination of the whole of molecular biology. But here is the problem, we are simulating a *particular* cell. I think there is the distinct danger of over interpreting the value of such models for the understanding of biology. As I mentioned previously understanding what something is entails that you also know what it is not. The detailed whole cell model can obviously be varied, but making such a model for every type of cell is not possible in practice. Besides, who is going to tell you what every type of cell is? You have no theory for "enumerating" the possible cell types, or the components the cells are made from.

Now for the second project: the synthesis of a cell. The main benefit of the synthesis of (simple) cells is, in my eyes, the completion of the array of increasingly complex experimental methods to study biological cells; from the isolated protein to a whole (evolved) cell, say *E. coli*. There is a catchphrase that goes something like this: if I can build something I understand it. This is mistaken. You can by trial and error, and by looking at nature, arrive at a recipe to synthesize a cell. But this provides no understanding of why the elements of the cell exist or why they are organized the way they are. Another problem with this statement arises if you take the reverse to also be true, that the lack of the ability to build a cell from scratch means that we do not understand a cell. Not being able to synthesize a cell could simply mean that you don't know how to synthesize a cell. After all, a cell in nature never synthesizes itself. You can also wonder what you could learn about the biology of a cell by synthesizing it if you can already make a *working* whole cell model. As was the case for the whole cell model I think we can overestimate the importance of synthesizing a cell for the understanding of biology. Again we can ask: what bloody cell are we talking about? We must also harken back to the kerfuffle about the definition of life. Life, in my submission, is not to be found in single cells or individuals. To think that synthesizing a cell is going to be the last nail in the coffin of vitalism (Ridley, 2008), is itself a vitalist position.

To capture life's diversity we need ways to generate the possibilities. This is a very difficult problem whose nature is well capture by the following example from (Kauffman, 2014).

"I begin in an odd way: 'Here is a screwdriver. Tell me all the uses of a screwdriver.'. Try it: screw in a screw, open a can of paint, stab an assailant, scrape putty off a window, tie to a stick and spear a fish, rent the spear to locals and make 5% of the catch... Do we agree that: (i) the number of uses of a screwdriver is indefinite. Next, the integers, 1, 2, 3, are orderable. Are the uses of the screwdriver I just listed orderable in any natural way? No, they are just "names" of different uses, a nominal scale only. But if we accept these two premises, indefinite and unorderable, then no algorithm, or "effective procedure", can list all the uses of a screwdriver or find the next use. I have just, I claim, shown you that uses of a screwdriver are not to be found algorithmically."

I suspect the prospect expressed in the last sentence is too glum. First off, the fact that Kauffman could list some of the uses of a screwdriver already shows that he has mastered some sort of algorithm. It may be incomplete, but it is by no means absent. Second, if you asked hundred people to list the uses of a screwdriver some of them will come up with more uses than others; and, some of them will find more interesting, i.e. less symmetrical, options than others. Meaning that some algorithms work better than others. Which, in turn, means that improvement is possible. What should also be clear from a list of uses of the screwdriver, is that the uses depend on the context. And I would claim that if we were to construct a theory of function it would have to be a theory of context. We do know of a lot of contexts in biology already and it seems to me that a strategy for discovery could be to bring

elements, e.g. some type of protein or cell, and contexts, e.g. other proteins or cells, together to generate novel concepts. Note also that novelty-generating algorithms are exactly what we use when having interesting internal experiences, converse with others, contribute to culture, or pose questions for further research.

When I think about the matters discussed in this chapter I often ask myself: have I gone off the beaten path or off the rails? However, I believe that you should have the courage of your own convictions; as such, I have stated my thoughts without apology. And I urge you to withhold nothing in your attacks on these ideas. Because, in sharp contrast to most of what I presented in the earlier chapters, the problems examined here haunt me and I genuinely care about their resolution. My hope is that one day our grasp of biology will be such that we stop being mere onlookers and start being painters of life.

References

Berendsen, H. (2007). In *Simulating the Physical World*. Cambridge, UK: Cambridge University Press, pp. 9-12.

Feynman, R.P., Leighton, R.B., and Sands, M. (2006). The Feynman Lectures on Physics - volume III. Massachusetts, USA: Addison-Wesley, p. 21-19.

Goldenfeld, N., and Woese, C. (2011) Life is Physics: Evolution as a Collective Phenomenon Far from Equilibrium. Annu. Rev. Condens. Matter Phys. **2:** 375-399.

Karr, J.R., Sanghvi, J.C., Macklin, D.N., Gutschow, M.V., Jacobs, J.M., Bolival, B.*, et al.* (2012) A Whole-Cell Computational Model Predicts Phenotype from Genotype*. Cell* **150:** 389-401.

Kauffman, S.A. (2014) Prolegomenon to patterns in evolution. *BioSystems* **123:** 3-8.

Koonin, E. (2011). The Logic of Chance: The Nature and Origin of Biological Evolution. New Jersey, USA: FT Press, pp. 293-328.

Lynch, M., Koskella, B., and Schaack, S. (2006) Mutation pressure and the evolution of organelle genomic architecture*. Science* **311:** 1727-1730.

Machery, E. (2012) Why I stopped worrying about the definition of life... and why you should as well. *Synthese* **185:** 145-164.

McCutcheon, J.P., and von Dohlen, C.D. (2011) An Interdependent Metabolic Patchwork in the Nested Symbiosis of Mealybugs. *Current Biology* **21:** 1366-1372.

Nakabachi, A., Yamashita, A., Toh, H., Ishikawa, H., Dunbar, H.E., Moran, N.A., and Hattori, M. (2006) The 160-kilobase genome of the bacterial endosymbiont Carsonella*. Science* **314:** 267-267.

Perez-Brocal, V., Gil, R., Ramos, S., Lamelas, A., Postigo, M., Manuel Michelena, J.*, et al.* (2006) A small microbial genome: The end of a long symbiotic relationship? *Science* **314:** 312-313.

Philippe, N., Legendre, M., Doutre, G., Coute, Y., Poirot, O., Lescot, M.*, et al.* (2013) Pandoraviruses: Amoeba Viruses with Genomes Up to 2.5 Mb Reaching That of Parasitic Eukaryotes. *Science* **341:** 281-286.

Ridley, M. (2008) Francis Crick: Discoverer of the Genetic Code. London, UK: Harper Perennial, p. 208.

Smith, J.M. (1993a) The Theory of Evolution. Cambridge, UK: Cambridge University Press, p. 255.

Smith, J.M. (1993b) The Theory of Evolution. Cambridge, UK: Cambridge University Press, p. 264.

Van Holde, K.E., Curtis Johnson, W., Shing Ho, P. (2006) Principles of Physical Biochemistry, New Jersey, USA: Pearson Education, p. 281.

Van Leuven, J.T., Meister, R.C., Simon, C., and McCutcheon, J.P. (2014) Sympatric Speciation in a Bacterial Endosymbiont Results in Two Genomes with the Functionality of One. *Cell* **158:** 1270-1280.

Wilson, E.O. (1998) Consilience: The Unity of Knowledge. New York, USA: Random House.

Woese, C. (2005) Q & A. *Current Biology* **15:** R111-R112.

Woese, C. (2002) On the evolution of cells. *Proc Natl Acad Sci U S A* **99:** 8742-8747.

Wulf, A. (2015). The Invention of Nature. USA: Alfred A. Knopf, pp. 88.

Samenvatting

Dit proefschrift bestaat uit drie delen. Het eerste deel (H. 1-3) gaat over de manier waarop eiwitten bewegen in prokaryote cellen (bacteriën en archaea), het tweede deel (H. 4, 5) gaat over de manier waarop membraan eiwitten geproduceerd worden in bacteriën, en in het laastste deel (H. 6) geef ik mijn opinie over wetenschap en biologie in bredere zin.

Cellen bestaan uit moleculen en die moleculen zijn constant in beweging. Deze beweging wordt in stand gehouden door het behoud van energie. Omdat al deze moleculen constant op elkaar botsen heeft de beweging van deze moleculen een grote mate van willekeur, zowel in richting als in afgelegde afstand. Deze soort van beweging wordt diffusie genoemd. Het bewegen van moleculen, wat ook plaatsvindt in alle andere voor ons bekende objecten, wordt in het dagelijks leven ervaren als temperatuur. De beweging der moleculen is niet geheel willekeurig. De *gemiddelde* afstand die een molecuul aflegt binnen een bepaalde tijd ligt vast en is afhankelijk van de eigenschappen van het molecuul zelf, bijvoorbeeld de omvang; en van de omgeving, bijvoorbeeld de stroperigheid van de vloeistof. Het wonderlijke is dat de gemiddelde afgelegde afstand en allerlei andere eigenschappen van de beweging van moleculen zich laten onderscheiden door een enkele waarde: de diffusiecoëfficiënt. Hoe hoger de diffusiecoëfficiënt hoe sneller de beweging. Wij hebben ons bezig gehouden met het bepalen van diffusiecoëfficiënten van eiwitten in prokaryote cellen.

Het doen en laten van een cel is grotendeels afhankelijk van de eiwitten die zich in die cel bevinden. Bijvoorbeeld het kopiëren van het DNA, het delen van een cel, het transporteren van voedingstoffen naar binnen en afvalstoffen naar buiten, en het bepalen van de vorm van de cel. Voor veel taken die de eiwitten uitvoeren is het van belang dat de verschillende eiwitten elkaar kunnen vinden. Zij vinden elkaar, bij toeval, door hun diffusie. En de snelheid waarmee zij elkaar vinden is direct gerelateerd aan de diffusiecoëfficiënt. Dus door het meten van diffusiecoëfficiënten van eiwitten kunnen wij iets zeggen over hoe snel die eiwitten hun taken kunnen uitvoeren. Daarnaast kunnen wij, wat in mijn ogen veel interessanter is, iets zeggen over welke taken wel of niet mogelijk zijn voor cellen.

Toen ik begon met het bestuderen van diffusie waren van vele eiwitten in de bacterie *Escherichia coli* al diffusiecoëfficiënten bepaald. Van de vele andere soorten bacteriën (en archaea) was vrijwel niks bekend op het gebied van diffusie. Wij kozen een in ons lab veel gebruikte bacterie, *Lactococcus lactis*, om te zien of diffusiecoëfficiënten variëren over verschillende soorten. *Lactococcus lactis* is een relatief eenvoudig eencellig organisme met een vorm van een ietwat uitgerekte bol met een diameter van ongeveer een micrometer (een miljoenste meter). Wij hebben de *diffusiecoëfficiënt* bepaald van vier verschillende eiwitten: GFP, β-Galactosidase-GFP, LacSΔIIA-GFP, en BcaP-GFP. GFP is een klein eiwit dat zich in het binnencompartiment van de cel (cytoplasma) bevindt. β-Galactosidase-GFP is een groot eiwit dat zich eveneens in het cytoplasma bevindt. LacSΔIIA-GFP en BcaP-GFP bevinden zich in de celmembraan.

Om de *diffusiecoëfficiënten* te kunnen meten moeten de specifieke eiwitten die wij willen bestuderen zichtbaar gemaakt worden onder de microscoop. Hiervoor kunnen wij handig gebruik maken van een eigenschap van het eiwit GFP. GFP, ook wel <u>G</u>reen <u>F</u>luorescent <u>P</u>rotein genoemd, fluoresceert; wat inhoudt dat het licht van een bepaalde golflengte kan absorberen en vervolgens licht van een andere golflengte weer kan uitsturen. Van het uitgestuurde licht wordt een foto gemaakt met de microscoop. Op die plaatjes kan de positie van de eiwitten binnen de cellen worden bepaald aan de hand van waar het licht zich bevindt. Nu kun je ook begrijpen waarom alle daarnet genoemde eiwitten GFP in de naam hebben. Deze eiwitten zijn namelijk allemaal zo gemodificeerd dat ze fluoresceren, zodat wij ze kunnen volgen met de microscoop. Het bepalen van de *diffusiecoëfficiënt* vanuit de microscoop plaatjes gaat als volgt. Voordat wij foto's maken schakelen wij de fluorescentie aan een zijde van de

cel uit met behulp van laserlicht. Vervolgens gaan de andere, nog fluorescerende, eiwitten zich herverdelen. Dit herverdelen is een gevolg van de willekeurige beweging van de individuele eiwitten en gaat met een snelheid die afhangt van de *diffusiecoëfficiënt*. Wij kunnen de foto's die we gemaakt hebben van het herverdelen nu vergelijken met een wiskundige formule die hetzelfde herverdelen beschrijft. Uit de vergelijking van fluorescentie foto's en wiskundige formule halen wij de *diffusiecoëfficiënt*.

De *diffusiecoëfficiënten* van de vier bestudeerde eiwitten in *Lactococcus lactis*, GFP, β-Galactosidase-GFP, LacSΔIIA-GFP, en BcaP-GFP, zijn vrijwel identiek aan wat gevonden is voor vergelijkbare eiwitten in *Escherichia coli*. Deze metingen zijn gedaan op cellen die zich in veel gebruikte groeimedia bevonden. Wij hebben wel een verschil tussen *Lactococcus lactis* en *Escherichia coli* gevonden als we de *diffusiecoëfficiënten* bepalen in andere omstandigheden. Als je *Lactococcus lactis* of *Escherichia coli* cellen blootstelt aan hoge zoutconcentraties wordt er water ontrokken aan het cytoplasma. Hierdoor wordt het cytoplasmatisch volume kleiner en komen de eiwitten dichter op elkaar te zitten. Wij hebben laten zien dat onder deze omstandigheden de *diffusiecoëfficiënt* van GFP in het cytoplasma van *Lactococcus lactis* lager is. Eerder is dit al voor *Escherichia coli* laten zien. Het verschil tussen deze twee organismen duikt op als je de *diffusiecoëfficiënt* relateerd aan het celvolume. Het blijkt dat de *diffusiecoëfficiënt* van GFP in *Lactococcus lactis* veel sneller afneemt met een kleiner wordend volume dan in *Escherichia coli*. De reden voor dit verschil is duister maar heeft mogelijk iets te doen met hoeveel eiwitten er in *Lactococcus lactis* zitten of hoe deze eiwitten zijn georganiseerd.

Om dit verschil te kunnen begrijpen, maar ook om algemeen inzicht te verkrijgen over diffusie in cellen, hebben we ook de *diffusiecoëfficiënt* bepaald van eiwitten met verschillende soorten oppervlakken. Of een eiwit zich aan een ander eiwit kan binden hangt af van de chemische eigenschappen van de oppervlakken van deze eiwitten. Als een klein eiwit aan een groot eiwit bindt wordt de diffusiesnelheid van het kleine eiwit lager. Wij hebben een set van GFP's verkregen die veranderen in netto lading van -30 tot en met +25. Van deze eiwitten hebben wij vervolgens de *diffusiecoëfficiënten* bepaald in de bacteriën *Escherichia coli* en *Lactococcus lactis*, en de archaeon *Haloferax volcanii*. Uit deze metingen blijkt dat negatief geladen eiwitten snel diffunderen, en dat hoe positiever het eiwit hoe langzamer de diffusie. Dit effect is het sterkst in *Escherichia coli*, waar de +25 GFP wel 100x langzamer diffundeert dan de -7 GFP, en het zwakst in *Haloferax volcanii*. De afname van de diffusiesnelheid van +25 GFP in de volgorde *Haloferax volcanii*, *Lactococcus lactis*, *Escherichia coli*, wordt veroorzaakt door de verschillen in het aantal kleine geladen moleculen/atomen, dat in diezelfde volgorde afneemt in de drie organismen. Kleine geladen moleculen en atomen kunnen de geladen oppervlakken van de eiwitten afdekken en daarmee de interacties die de diffusie vertragen teniet doen.

Een census van de componenten die zich in de cel bevinden laat zien dat alleen de membraan eiwitten, DNA, mRNA, en ribosomen langzaam genoeg diffunderen om de diffusiesnelheid in de mate te laten afnemen die wij hebben gemeten. In een serie van vervolg experimenten in *Escherichia coli* hebben wij laten zien dat het voornamelijk de ribosomen zijn die de positieve GFP's afremmen. Omdat ribosomen zeer belangrijk zijn voor het functioneren van een cel is het waarschijnlijk dat het aantal positieve eiwitten zo klein mogelijk wordt gehouden. Want als er allerlei eiwitten op de ribosomen blijven plakken kunnen de ribosomen hun taak niet meer uitvoeren. Het langzaam diffunderen van positieve eiwitten is waarschijnlijk iets dat voorkomt in alle cellulaire organismen aangezien al deze organismen ribosomen hebben. Aan de hand van de lettervolgorde in het DNA kunnen voorspellingen gedaan worden over de lading van eiwit oppervlakken. Dit laat zien dat in *Escherichia coli*, *Lactococcus lactis*, en *Haloferax volcanii* de meeste eiwitten negatief geladen zijn. Er zijn echter ook organismen,

bijvoorbeeld *Buchnera aphidicola*, waarvan we voorspellen dat de meeste eiwitten positief geladen zijn. Hoe deze organismen kunnen bestaan met zo'n positieve set aan eiwitten is een raadsel.

Het tweede deel van het proefschrift gaat over de productie van membraaneiwitten in *Escherichia coli* en *Lactococcus lactis*. Net als alle andere eiwitten zijn membraaneiwitten gecodeerd in het DNA. De informatie in het deel van het DNA dat voor het membraa eiwit codeert wordt gekopieerd door een RNA polymerase. Zo'n kopie heet een mRNA. Dit mRNA wordt vervolgens gebruikt door ribosomen om één of meerdere eiwitten te produceren. In het geval van membraaneiwitten wordt de productie van het eiwit vergezeld met het in de membraan zetten van datzelfde eiwit door het Sec translocon.

Wij hebben een techniek bedacht om de productie van membraaneiwitten van eenzelfde mRNA te kunnen volgen op het niveau van enkele moleculen en in een groeiende en delende cel. Het volgen van eiwitten in cellen is mogelijk dankzij de al eerder besproken fluorescerende eiwitten. Het probleem is echter dat de productie van een membraan eiwit tientallen seconden duurt terwijl een fluorescent eiwit pas fluorescent wordt na een aantal minuten. Met als gevolg dat als wij een fusie maken van een membraan eiwit en een fluorescerend eiwit wij niks zien van de timing van de productie maar alleen de toename van het aantal membraan eiwitten. Dit probleem kan omzeilt worden door de fluorescerende eiwitten voortijdig te produceren in het cytoplasma en die vervolgens te laten binden aan de net geproduceerde membraaneiwitten met behulp van een interactie eiwit. Door verschillen in de diffusiesnelheid van cytoplasmatische eiwitten en membraan eiwitten, en door de sluitertijd van de microscoop camera goed te kiezen, kan het binden van het fluorescente eiwit aan het membraan eiwit zichtbaar worden gemaakt als het onstaan van een fluorescente stip. De tijd tussen het ontstaan van de stippen kan gerelateerd worden aan de tijd tussen de productie van verschillende eiwitten vanaf één mRNA. Dit was het idee. Wij hebben op veel manieren geprobeerd om het idee in de praktijk tot stand te brengen. Wij hebben verschillende fluorescente eiwitten, membraan eiwitten, interactie eiwitten, en organismen geprobeerd. Wij hebben ook geprobeerd om vanaf de buitenkant de membraaneiwitten te binden en hebben geprobeerd meerdere fluorescente eiwitten aan een membraan eiwit te laten binden. Geen van deze methodes was succesvol. Dit komt voornamelijk door de aanwezigheid van fluctuaties in het fluorescentiesignaal, wat stippen laat verschijnen en verdwijnen onafhankelijk van wat er met de membraan eiwitten gebeurd.

In het laatste onderdeel van dit proefschrift heb ik mij uitgelaten over de wetenschap en biologie in bredere zin. De enige manier om iets te weten te komen over de wereld waarin wij leven is door gebruik te maken van logica en observatie. Deze houding, en de daaruit voortkomende vakgebieden zoals de natuurwetenschappen, zijn een cruciaal onderdeel van onze cultuur en zingeving. De biologie heeft hier, in haar bijna oneindige diversiteit, veel aan bij te dragen. Ik zie het als de taak van de bioloog, niet alleen om die diversiteit te kunnen observeren voor zover die hier op aarde voorkomt of voorkwam, maar ook om alle mogelijke diversiteit te kunnen navigeren. Dit betekent ook dat het voor zover als mogelijk toegankelijk gemaakt moet worden voor een individu. Ik zie een nadruk op diversiteit ook als een sleutel voor het ontrafelen van wat leven is.

Summary

This thesis consists of three parts. Part one (Ch. 1-3) is about the movement of proteins in prokaryotic cells (bacteria and archaea), part two (Ch. 4, 5) is about the production of membrane proteins in bacteria, and in the last part (Ch. 6) I express, in a broad sense, my view of science and biology.

Cells consist of molecules and those molecules are constantly moving. This movement is perpetual because of the conservation of energy. These molecules bump into each other almost continually causing their motion to be ruled almost entirely by randomness, both in direction and distance travelled. This kind of movement is called diffusion. This molecular movement, which also occurs in all other known objects, is experienced by people as temperature. The movement of molecules is not entirely random. The *average* distance a molecule covers in a fixed time interval is non-random and depends on the properties of the diffusing molecule, for instance its size; and the environment, for instance the viscosity of the liquid. A peculiar property of diffusion is that the average distance travelled, along with all kinds of other properties of the motion of molecules, can be described by a single value: the diffusion coefficient. The higher the diffusion coefficient the faster the movement. We have measured diffusion coefficients of proteins in prokaryotic cells.

The behaviour of a cell is in large part dependent on the proteins present within that cell. For example the replication of DNA, the division of one cell into two, the transport of nutrients to the inside and waste to the outside of the cell, and determining the shape of the cell. For many of the tasks that proteins carry out it is important that different proteins can find each other. They do so, randomly, by their diffusion. And the rate at which they find each other is directly related to their diffusion coefficient. So by measuring diffusion coefficient we can make claims about the rate with which proteins can carry out their assigned tasks. And, more interestingly, we can say something about what tasks are possible or impossible for cells.

When I started to look into diffusion the bacterium *Escherichia coli* was already well studied, with diffusion coefficients known for many of its proteins. However, for many other species of bacteria (and archaea) little to nothing was known about the diffusion of their proteins. We chose the bacterium *Lactococcus lactis* to find out whether there are differences in protein diffusion rates between different species. *Lactococcus lactis* is a relatively simple single celled organism with the shape of a somewhat elongated sphere with a micro (millionth) meter diameter. We have determined diffusion coefficients of four different proteins: GFP, β-Galactosidase-GFP, LacSΔIIA-GFP, and BcaP-GFP. GFP is a small protein that is present in the cytoplasm. β-Galactosidase-GFP is a big protein that is also present in the cytoplasm. LacSΔIIA-GFP and BcaP-GFP are located in the membrane.

To measure the diffusion coefficients the specific proteins under study have to be selectively visualized under a microscope. This can be done by using a peculiar property of the protein GFP. GFP, also called Green Fluorescent Protein, fluoresces; meaning that it can absorb light of one wavelength and emit it in another wavelength. The emitted light is used in the microscope to generate an image. On these images the position of the proteins within the cells can be determined from where the light is localized. Now you can also understand why all the proteins mentioned earlier have GFP in their name. They are all modified to become fluorescent so that we can follow their positions under the microscope. Determining diffusion coefficients from microscope images works as follows. Before we start to make pictures of the fluorescence in the cell we take out the fluorescence of the proteins on one side of the cell with the help of laser light. The still fluorescent proteins rearrange themselves over time. This rearranging is a consequence of the random motion of individual molecules and happens with a rate that depends on the diffusion coefficient. The images we made of the rearrangement of the proteins

can be compared to a mathematical equation that describes the same rearrangement. From this comparison we get the diffusion coefficient.

The diffusion coefficients of the four target proteins in *Lactococcus lactis*, GFP, β-Galactosidase-GFP, LacSΔIIA-GFP, and BcaP-GFP, are pretty much identical to what has been found for similar proteins in *Escherichia coli*. These measurements were done on cells present in growth media that are used regularly in the lab. When studying cells that were embedded in different media we did find differences in the diffusion rates of proteins between *Escherichia coli* and *Lactococcus lactis*. When *Escherichia coli* and *Lactococcus lactis* cells are deposited in a high salt solution the cytoplasm loses water. The volume of the cytoplasm shrinks and the proteins are packed together more tightly. We have determined that under these circumstances the diffusion coefficient of GFP in *Lactococcus lactis* has decreased. This has also been shown previously for *Escherichia coli*. The difference between these two organisms appears when the relation between diffusion coefficient and cell volume is compared. It turns out that the diffusion coefficient of GFP decreases with a decrease in volume much faster in *Lactococcus lactis* than in *Escherichia coli*. We do not know the source of this difference. However, it may be related to the amount, or organization, of proteins present in *Lactococcus lactis*.

To obtain insight into this difference, but also to study diffusion in cells more generally, we determined diffusion coefficients of proteins that have different surfaces. Whether a protein can bind to another protein depends on the chemical properties of the surfaces of these proteins. If a small protein binds a large protein its diffusion rate drops. We have obtained a set of GFP's that differ in net surface charge from -30 to +25. We determined the diffusion coefficients of these proteins in the bacteria *Escherichia coli* and *Lactococcus lactis*, and the archaeon *Haloferax volcanii*. These measurements show that negatively charged proteins diffuse rapidly, and that the more positive a protein is the slower it diffuses. This effect is most prominent in *Escherichia coli*, in which the +25 GFP diffuses a 100-fold slower than -7 GFP, and least prominent in *Haloferax volcanii*. The decrease in the diffusion coefficient of +25 GFP in the order *Haloferax volcanii*, *Lactococcus lactis*, *Escherichia coli*, is caused by differences in the amount of small charged molecules and atoms, which decreases in the same order in the three organisms. Small charged molecules and atoms can cover the charged surfaces of proteins making the diffusion decreasing interactions between proteins weaker.

A census of the components that make up the cell shows that the decrease in diffusion coefficient found for +25 GFP can only be caused by the membrane, DNA, mRNA, and/or ribosomes. In a series of follow up experiments in *Escherichia coli* we have shown that the decrease in +25 GFP diffusion coefficient is caused mainly by ribosomes. And because ribosomes are central to the functioning of the cell it is likely that in evolution the number of positive proteins is kept small. The reason being that it would be difficult for the ribosomes to carry out their tasks when they are swarmed by positive proteins. The slow diffusion of positive proteins probably occurs in all cellular organisms as they all contain ribosomes. The sequence of letters in the full complement of DNA of an organism can be used to make predictions about the charge present on the surfaces of all proteins in that organism. This shows that in *Escherichia coli*, *Lactococcus lactis*, and *Haloferax volcanii* most proteins are negatively charged. However, there are organisms, for example *Buchnera aphidicola*, that are predicted to have mostly positively charged proteins. How these organisms can function properly with such a positive set of proteins is a mystery.

The second part of the thesis is about the production of membrane proteins by *Escherichia coli* and *Lactococcus lactis*. Like all other proteins, membrane proteins are encoded on the DNA. The information in the part of the DNA that encodes the membrane protein is copied by an RNA polymerase producing an mRNA. This mRNA is used by ribosomes to produce one or more copies of

the membrane protein. While membrane proteins are being synthesized by the ribosomes they are also being inserted into the membrane by the Sec translocon.

We imagined a way to follow the production of individual membrane proteins from a single mRNA in cells that are still growing and dividing. It is possible to follow the production of proteins by using the already mentioned fluorescent proteins. The problem is that the production of a membrane protein takes tens of seconds whereas the fluorescent proteins only become fluorescent minutes after being produced. The upshot is that if we would make a fusion between the membrane protein and fluorescent protein we would not get any relevant data on the timing of production events but only on the increase in total number of these membrane proteins. This problem can be evaded by having the fluorescent proteins present in the cytoplasm in the fluorescent state before the membrane protein is produced. The fluorescent proteins then interact with the membrane proteins if both have been equipped with interaction domains. Because of differences in the diffusion rate of cytoplasmic and membrane proteins, and by choosing the exposure time in the microscope appropriately, attachment of the fluorescent protein to the membrane protein can be visualized as the appearance of a fluorescent spot. The time between the appearances of spots can be related to the time between the productions of different proteins from the same mRNA. This was the idea. We have made many attempts to turn this idea into reality. We have tried different variants of fluorescent proteins, various membrane proteins, different interaction proteins, and two different organisms. We have also tried to bind fluorescent molecules to the membrane proteins from the outside and to attach multiple fluorescent proteins to the same membrane protein. None of these methods was successful. The main cause of the failures are fluctuations in the fluorescence signal, which makes spots appear and disappear independent of what happens to the membrane proteins.

The last part of this thesis I have given some of my opinions in science and biology on general. The only way we can know something about the world we live in is by using logic and observation. This attitude, and the fields like the natural sciences that emerge from it, are a crucial element of our culture and sense of purpose. Biology, with its study of the staggering diversity of life, has much to contribute to this. In my eyes the purpose of the biologist is to not only be witness to what diversity is present in the world now or has existed in the past, but also to be able to navigate what could have been and what can be. This also means that the diversity of life should be made as far as possible accessible to an individual. I also see an emphasis on diversity as a key to understand what life is.