Microfluidics

DOI: 10.1002/anie.200901550

Microfluidic Confinement of Single Cells of Bacteria in Small Volumes Initiates High-Density Behavior of Quorum Sensing and Growth and Reveals Its Variability**

*James Q. Boedicker, Meghan E. Vincent, and Rustem F. Ismagilov**

Herein we show that upon confinement in small volumes, groups of *Pseudomonas aeruginosa* bacteria, containing as few as one to three cells, are able to initiate quorum sensing (QS) and achieve QS-dependent growth. In addition, we show that, at low numbers of cells, initiation of QS is highly variable within a clonal population. QS pathways are involved in critical functions of microorganisms, such as pathogenesis, development of biofilms, sporulation, acquisition of nutrients, conjugation, motility, and production of secondary metabolites such as antibiotics.[1] Experimental control of QS pathways through confinement may enable fundamental research into the role of QS in small groups of cells and could provide a tool for growing unculturable bacteria or inducing antibiotic production.

QS is by definition a high-density behavior, regulated by the number of cells per unit volume. QS is initiated by the accumulation of released signaling molecules, such as autoinducers (AIs), and is known to be initiated when the density of cells rises above a threshold level. There are two approaches to activating high-density behavior in cultures: 1) seed a macroscopic volume with bacteria and let them divide until they reach high density, or 2) take a few cells and confine them in a very small volume to enable accumulation of AIs (Figure 1 A). Activation of QS through strategy (1) has dominated and has led to the general view that QS is a process to coordinate the collective behavior of large groups of cells,[2,3] and the possibility that small groups of bacteria could initiate QS upon confinement is often overlooked. Nevertheless, strategy (2) is important to consider, because QS pathways are relevant to function, survival, and growth of small numbers of cells—for example, early in biofilm formation, in early stages of infection, or in soils.

Because confinement influences the diffusion of released signaling molecules, QS has been redefined in terms of diffusion[4] or efficiency[5] sensing. Spatial constraints and transport parameters, such as the flow rate through open systems, have been shown to play a role in the regulation of QS.[6–9] In addition, it has been shown that even small groups of *Staphylococcus aureus* cells inside a vesicle of a host cell are able to initiate QS,[10,11] and that initiation is required for escape from the vesicle.[11,12] In another example, the restriction of diffusion on dry leaf surfaces reduced the size of a quorum down to groups of dozens of cells.[13] In these experiments the cells were not only confined but they were also exposed to the host environment. It is not known in these systems to what extent host factors or confinement are playing a role in the initiation of QS.

To study high-density behavior of a few cells experimentally without influence from host factors requires the creation

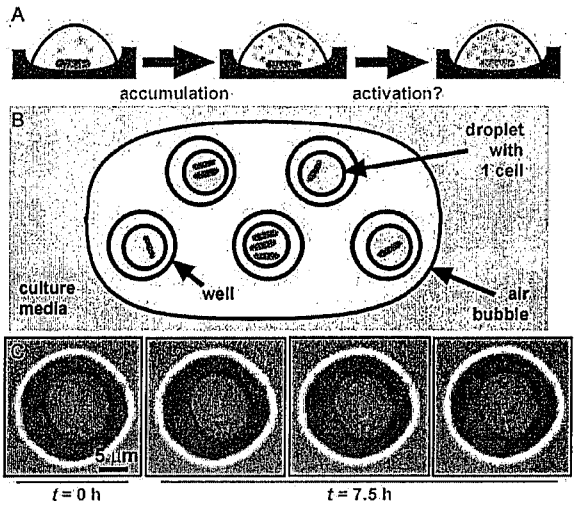

Figure 1. A hypothesis for activation of QS by confinement of single cells and a device for testing this hypothesis. A) *Pseudomonas aeruginosa* cells (gray rods) excrete auto inducers (AIs, orange circles) that can accumulate in the media. In a confined volume, the AIs do not diffuse away, and thus they reach high concentrations around the cell. We hypothesized that in a small volume, a single cell would be able to accumulate AIs above the critical concentration needed to initiate QS. In these experiments, QS was visualized by using a green-fluorescent protein (GFP) reporter gene for lasB (green rod). B) Schematic drawing of arrays of sub-picoliter droplets that contain confined bacteria covered by air. The entire device is sealed inside a petri dish (not shown). C) Small groups of bacteria can grow and divide when confined in droplets with volumes less than 1 pL. After 7.5 h, the original cell has divided, as shown by the increase in the number of cells in the droplet, and the bacteria are also motile, as indicated by the series of pictures, taken several seconds apart, showing movement of some of the cells in the droplet.

[*] J. Q. Boedicker, M. E. Vincent, Prof. Dr. R. F. Ismagilov
Department of Chemistry, University of Chicago
929 East 57th Street, Chicago, IL 60637 (USA)
Fax: (+1) 773-834-3544
E-mail: r-ismagilov@uchicago.edu
Homepage: http://ismagilovlab.uchicago.edu/index.html

[**] This research was supported by the NIH Director's Pioneer Award DP1 OD003584 to R.F.I. We thank S. Molin for generously providing the reporter strain, J. Shapiro and O. Zaborina for helpful discussion, and E. B. Haney for contributions to writing this manuscript.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.200901550.

Biophysical Journal Volume 97 October 2009 2368–2376

Quantitative Comparison of Different Fluorescent Protein Couples for Fast FRET-FLIM Acquisition

Sergi Padilla-Parra,* Nicolas Audugé, Hervé Lalucque, Jean-Claude Mevel, Maïté Coppey-Moisan,* and Marc Tramier

Institut Jacques Monod, Unite Mixte de Recherche 7592, and Centre National de la Recherche Scientifique, Université Paris-Diderot, Paris, France

ABSTRACT The fluorescent-protein based fluorescence resonance energy transfer (FRET) approach is a powerful method for quantifying protein-protein interactions in living cells, especially when combined with fluorescence lifetime imaging microscopy (FLIM). To compare the performance of different FRET couples for FRET-FLIM experiments, we first tested enhanced green fluorescent protein (EGFP) linked to different red acceptors (mRFP1-EGFP, mStrawberry-EGFP, HaloTag (TMR)-EGFP, and mCherry-EGFP). We obtained a fraction of donor engaged in FRET (f_D) that was far from the ideal case of one, using different mathematical models assuming a double species model (i.e., discrete double exponential fixing the donor lifetime and double exponential stretched for the FRET lifetime). We show that the relatively low f_D percentages obtained with these models may be due to spectroscopic heterogeneity of the acceptor population, which is partially caused by different maturation rates for the donor and the acceptor. In an attempt to improve the amount of donor protein engaged in FRET, we tested mTFP1 as a donor coupled to mOrange and EYFP, respectively. mTFP1 turned out to be at least as good as EGFP for donor FRET-FLIM experiments because 1), its lifetime remained constant during light-induced fluorescent changes; 2), its fluorescence decay profile was best fitted with a single exponential model; and 3), no photoconversion was detected. The f_D value when combined with EYFP as an acceptor was the highest of all tandems tested (0.7). Moreover, in the context of fast acquisitions, we obtained a minimal f_D (mf_D) for mTFP1-EYFP that was almost two times greater than that for mCherry-EGFP (0.65 vs. 0.35). Finally, we compared EGFP and mTFP1 in a biological situation in which the fusion proteins were highly immobile, and EGFP and mTFP1 were linked to the histone H4 (EGFP-H4 and mTFP1-H4) in fast FLIM acquisitions. In this particular case, the fluorescence intensity was more stable for EGFP-H4 than for mTFP1-H4. Nevertheless, we show that mTFP1/EYFP stands alone as the best FRET-FLIM couple in terms of f_D analysis.

INTRODUCTION

Quantitative microscopy techniques are performed in live cells to elucidate the spatial and temporal dynamics of protein interactions. Imaging of fluorescence resonance energy transfer (FRET) (1) using fluorescent proteins (FPs) (2) in living cells helps to detect and quantify protein-protein interactions to reveal the spatiotemporal dynamics of different biological systems (3). A suitable technique for detecting FRET is fluorescence lifetime imaging microscopy (FLIM). By using FLIM to quantify FRET, one can determine the fraction of donor in the interaction (f_D) (4–7). This parameter is related to the relative concentration of interacting protein and is particularly interesting in relation to biology. In the last few years, FLIM has emerged as a powerful tool to detect and quantify the dynamics of protein interactions. Different approaches have been developed that reduce acquisition times (8,9), provide a map of the molecular environment of a fluorophore in a rapid manner, and allow the related interaction to be followed as a function of time.

The sensitivity of a FRET experiment depends on the FP couple used. Frequently, comparisons between different FRET couples are performed using tandems with a mean determination of the FRET efficiency, E (10,11). However, when considering intermolecular interactions, the main parameter to compare for every FRET pair is the f_D value, which is only accessible with the use of FLIM. Additionally, in FRET-FLIM experiments, it is important to use a donor with the right photophysical properties, i.e., high photostability, absence of photoconversion, and single exponential behavior of its fluorescence decay profile. Enhanced green fluorescent protein (EGFP) is known to fulfill these requirements (12) and is frequently combined with red acceptors for FLIM experiments (9,13–17). However, red acceptors (e.g., mRFP1 (18) and mCherry (19)) exhibit relatively low f_D values that can lead to misinterpretation of quantitative data (14,20). This implies that when intermolecular interactions in a particular biological system are being considered, a mathematical fD correction should be done (9,14).

In an attempt to determine the best FRET couple for quantitative FRET-FLIM, we analyzed a set of different FRET standards formed by two FPs (donor and acceptor) linked by a polypeptide chain. We used FLIM to find the best couple of fluorophores for quantitative experiments based on f_D calculation. mTFP1-EYFP and mTFP1-mOrange were previously proposed as attractive FRET couples by Ai and co-workers (21). They also reported that the

Submitted May 13, 2009, and accepted for publication July 21, 2009.

*Correspondence: sergi_padilla@hotmail.com or coppey.maite@ijm.univ-paris-diderot.fr

Editor: Enrico Gratton.

0006-3495/09/10/2368/9 $2.00

doi: 10.1016/j.bpj.2009.07.044

Optical Saturation as a Versatile Tool to Enhance Resolution in Confocal Microscopy

Jana Humpolíčková,[†]* Aleš Benda,[†] and Jörg Enderlein[‡]
[†]J. Heyrovský Institute of Physical Chemistry, Academy of Sciences of the Czech Republic, Prague, Czech Republic; and [‡]Third Institute of Physics, Georg-August-Universität Göttingen, Göttingen, Germany

ABSTRACT One of the most actively developing areas in fluorescence microscopy is the achievement of spatial resolution below Abbe's diffraction limit, which restricts the resolution to several hundreds of nanometers. Most of the approaches in use at this time require a complex optical setup, a difficult mathematical treatment, or usage of dyes with special photophysical properties. In this work, we present a new, to our knowledge, approach in confocal microscopy that enhances the resolution moderately but is both technically and computationally simple. As it is based on the saturation of the transition from the ground state to the first excited state, it is universally applicable with respect to the dye used. The idea of the method presented is based on a principle similar to that underlying saturation excitation microscopy, but instead of applying harmonically modulated excitation light, the fluorophores are excited by picosecond laser pulses at different intensities, resulting in different levels of saturation. We show that the method can be easily combined with the concept of triplet relaxation, which by tuning the dark periods between pulses helps to suppress the formation of a photolabile triplet state and effectively reduces photobleaching. We demonstrate our approach imaging GFP-labeled protein patches within the plasma membrane of yeast cells.

INTRODUCTION

Fluorescence microscopy represents one of the most powerful imaging techniques in molecular biology, since 1), it allows for specific labeling, and 2), compared to standard light microscopy, it is enormously sensitive, yielding high image contrast. Moreover, since fluorescence excitation is rather noninvasive, fluorescence microscopy is a perfect tool for imaging in living systems. However, as with any optical imaging technique, fluorescence microscopy is diffraction-limited in resolution, and depending on the excitation and emission wavelength used, as well as the numerical aperture of the objective, its resolution is usually on the order of several hundreds of nanometers. To break that diffraction limit, special methods have been developed: stimulated emission depletion (1) reduces the volume from which fluorescence is generated by depleting excited molecules within a ringlike region at the edges of the diffraction-limited excitation spot. The final resolution is given by the saturation level of the depleting process, i.e., by the depletion intensity. This approach has been shown to achieve up to 12 times better (2) spatial resolution in a lateral dimension. By combining it with 4Pi-microscopy by using two objectives (3), it can also be used to significantly improve the axial resolution. The price of this superior performance is an enormous technical complexity, which has until now, prevented a wider distribution of stimulated emission depletion microscopy, although recent progress in the technology may perhaps change that (4). Another set of methods is based on the precise localization of stochastically activated single fluorophores and can achieve lateral resolution of ~20 nm (5–7). Nowadays, it has been shown, even common fluorescence dyes can be utilized for this kind of photoswitching microscopy (8,9), which makes these approaches even more attractive.

A third set of techniques makes use of saturation of the transition from the ground to the first excited state and the resulting nonlinearity between excitation and fluorescence intensity. For wide-field imaging, structured (10) or patterned (11) illumination is introduced, and the information on higher resolution is deduced from higher-order harmonics that appear in the Fourier transform of the image. A similar idea was also implemented for confocal microscopes, where it is known as saturation excitation (SAX) microscopy. Instead of the spatial modulation of light, a temporally modulated excitation is applied (12,13). In theory, the resolution enhancement of both these approaches can go to infinity, although the real performance is substantially dependent on the brightness and photostability of the fluorophores used. The impact of photostability is even more crucial, since the excitation intensities have to be rather large to reach sufficient saturation.

In this article, we present an alternative to SAX microscopy that, instead of using temporal harmonic light modulation, acquires the entire saturation curve at every scanning step. The amount and order of nonlinearity is obtained from fitting of the saturation curves to a theoretical model that can be represented by an infinite Taylor series. The method is both experimentally and mathematically simple, improves both lateral and axial resolution in a similar manner, and is dye-independent. The experimentally improved resolution is moderate, but comparable with resolution using SAX microscopy. The most important benefit of the presented

Submitted June 2, 2009, and accepted for publication August 3, 2009.

Jana Humpolíčková and Aleš Benda contributed equally to this work.

*Correspondence: jana.humpolickova@jh-inst.cas.cz

Editor: Levi A. Gheber.

0006-3495/09/11/2623/7 $2.00

doi: 10.1016/j.bpj.2009.08.002

Biophysical Journal Volume 97 November 2009 L22–L24

Long Fluorescence Lifetime Molecular Probes Based on Near Infrared Pyrrolopyrrole Cyanine Fluorophores for In Vivo Imaging

Mikhail Y. Berezin,† Walter J. Akers,† Kevin Guo,† Georg M. Fischer,§ Ewald Daltrozzo,§ Andreas Zumbusch,§ and Samuel Achilefu†‡*

†Department of Radiology and ‡Department of Biochemistry & Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri, and §Department of Chemistry, University of Konstanz, Konstanz, Germany

ABSTRACT Fluorescence lifetime (FLT) properties of organic molecules provide a new reporting strategy for molecular imaging in the near infrared (NIR) spectral region. Unfortunately, most of the NIR fluorescent dyes have short FLT typically clustered below 1.5 ns. In this study, we demonstrate that a new class of NIR fluorescent dyes, pyrrolopyrrole cyanine dyes, have exceptionally long FLTs ranging from 3 to 4 ns, both in vitro (dimethyl sulfoxide and albumin/water solutions) and in vivo (mice). These results provide a new window for imaging molecular processes, rejecting backscattered light and autofluorescence, and multiplexing imaging information with conventional NIR fluorescent dyes that absorb and emit light at similar wavelengths.

Received for publication 10 July 2009 and in final form 17 August 2009.

*Correspondence: achilefus@mir.wustl.edu

A variety of near-infrared (NIR) fluorescent molecular probes have been developed for in vivo imaging applications, mostly driven by the deep penetration of NIR light in tissue (1,2). Since optical imaging is often based on fluorescence intensity, requirements for ideal NIR fluorescent dyes include high molar absorptivity, photostability, and fluorescence quantum yield to improve image quality. However, the dependence of fluorescence intensity on many intrinsic and extrinsic variables presents a serious challenge in data analysis from in vivo studies. To overcome these problems, advanced optical imaging systems and image reconstruction algorithms have been developed to provide quantitative fluorescence intensity maps of dye distribution in small animals (3,4). In contrast, the fluorescence lifetimes (FLT) of organic dyes are less dependent on dye concentration and can be sensitive to the dye's microenvironment. In addition, FLT is much less affected by concentration artifacts and photobleaching (5), light scattering (6,7), excitation intensity, or sample turbidity (8,9). These features have motivated the use of FLT as a complementary in vivo optical imaging method (10–13). In combination with intensity maps, changes in FLT can reliably report molecular events in a specific region of interest.

Beyond improving the tissue penetration depth of light, imaging in the NIR region is particularly favorable for FLT measurements. In the visible wavelengths between 500 and 650 nm, autofluorescence FLTs ranges from 1–10 ns (14,15). Consequently, in vivo molecular imaging would require that the lifetime of exogenous probes must be significantly higher than that of autofluorescence (16,17). Since autofluorescence is negligible between 750 and 850 nm, molecular probes with diverse FLTs are suitable for NIR FLT imaging, depending on instrument capability and FLT resolution. While the NIR window facilitates the use of NIR fluorescent dyes with a wide FLT range, virtually all the NIR fluorescent organic dyes currently used for molecular imaging exhibit FLTs of <1.5 ns (18), with very few exceptions such as phthalocyanines (19), thereby confining the range of imaging applications with these probes.

A major obstacle to developing NIR dyes with long FLTs has been attributed to the energy gap law (21,22), which states that radiationless transitions at longer wavelengths increase due to vibrational overlaps between the ground and excited states, thus causing a decrease in quantum yield and fluorescence lifetime (23). Recently, the energy gap law was challenged by new generations of fluorescent dyes based on the pyrrolopyrrole cyanine (PPCy) skeleton (24,25). The new dyes exhibit high quantum yields (>0.50) in the NIR range, rendering them suitable for in vivo imaging. Herein, we report that these dyes have long FLTs, which expand the FLT window for imaging molecular processes. We also demonstrate what to our knowledge are the first applications of these dyes for in vivo imaging.

PPCy compounds are interesting for in vivo use because similar diketopyrrolopyrroles are generally nontoxic by either the oral route (LD_{50} > 5000 mg/kg) or by skin contact (LD_{50} > 2000 mg/kg), are not irritating to either the skin or eyes, are not mutagenic (26), and have been approved by the FDA for food contact. Particularly, the NIR PPCy dyes 4a and 4a′ (Fig. 1) were selected for this study because we anticipated that they would have similar toxicity profiles as basic PPCy dyes and their compact structure produced promising steady-state optical properties such as high molar absorptivities, quantum yields and photostability (24,25). For comparative purpose, we used indocyanine green (ICG) dye as a reference dye for this study since it is widely used in molecular imaging and

Editor: Alberto Diaspro.

doi: 10.1016/j.bpj.2009.08.022

Cell

Detecting Folding Intermediates of a Protein as It Passes through the Bacterial Translocation Channel

Hiroshi Kadokura[1,2,*] and Jon Beckwith[1,*]
[1]Department of Microbiology and Molecular Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, USA
[2]Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan
*Correspondence: hkadokura@bs.naist.jp (H.K.), jbeckwith@hms.harvard.edu (J.B.)
DOI 10.1016/j.cell.2009.07.030

SUMMARY

Most bacterial exported proteins cross the cytoplasmic membrane as unfolded polypeptides. However, little is known about how they fold during or after this process due to the difficulty in detecting folding intermediates. Here we identify cotranslational and posttranslational folding intermediates of a periplasmic protein in which the protein and DsbA, a periplasmic disulfide bond-forming enzyme, are covalently linked by a disulfide bond. The cotranslational mixed-disulfide intermediate is, upon further chain elongation, resolved, releasing the oxidized polypeptide, thus allowing us to follow the folding process. This analysis reveals that two cysteines that are joined to form a structural disulfide can play different roles during the folding reaction and that the mode of translocation (cotranslational verse posttranslational) can affect the folding process of a protein in the periplasm. The latter finding leads us to propose that the activity of the ribosome (translation) can modulate protein folding even in an extracytosolic compartment.

INTRODUCTION

In living organisms, most exported proteins cross the cytoplasmic membrane through the Sec channel (Wickner and Schekman, 2005). In order for it to go through the channel, a protein must be maintained in an unfolded state (Liu et al., 1988; Osborne and Rapoport, 2007). However, poorly understood is the nature of the folding process of a protein during and after its passage through the channel (Akiyama and Ito, 1993; Ureta et al., 2007).

A critical step in the folding of many exported proteins is the formation of disulfide bonds (Sevier and Kaiser, 2006). Thus, one convenient way to study aspects of folding of an exported protein in vivo is by following the appearance of disulfide bonds in that protein (Jansens et al., 2002).

Disulfide bonds are generally introduced into exported proteins by members of the thioredoxin superfamily in both prokaryotes and eukaryotes (Kadokura et al., 2003; Sevier and Kaiser, 2006). In the periplasm of *Escherichia coli*, the formation of disulfide bonds is catalyzed by the protein DsbA, a thioredoxin superfamily member, which oxidizes substrates by donating its disulfide bond to a pair of cysteines on a substrate (Bardwell et al., 1991; Kamitani et al., 1992). The process likely begins with the attack by a deprotonated cysteine of the substrate protein on the disulfide bond of DsbA, leading to a short-lived complex of DsbA and substrate linked together by an intermolecular disulfide bond (Figure 1A) (Kadokura et al., 2004). Next, a deprotonated second cysteine of the substrate attacks the mixed-disulfide to resolve the complex, releasing oxidized substrate and reduced DsbA. Results obtained in vitro are consistent with this model (Darby and Creighton, 1995; Frech et al., 1996). However, due to the difficulty of detecting the disulfide-linked enzyme-substrate complex, the process leading to formation of a disulfide bond in a folding protein via this complex has not been followed in vivo. Additionally, many questions remain about the mechanism of disulfide bond formation in vivo. We do not know how formation of disulfide bonds is coordinated with protein translocation and other folding processes. We also do not know which cysteines in substrates are used to form the intermediate mixed-disulfide complex with DsbA.

We previously reported a mutation that generated a Pro151 to Thr change in DsbA, which slows down the resolution of covalent mixed-disulfide complexes between DsbA and its substrates and, thus, allows the detection of these intermediates (Kadokura et al., 2004). Here we use this mutant and wild-type cells to follow the oxidative folding of a periplasmic protein in vivo. This approach enables us to (1) delineate the two steps of electron transfer that lead to the formation of a disulfide bond in a folding protein by a thioredoxin superfamily member (Figure 1A) and (2) specify the substrate cysteine that is used to form a disulfide-linked enzyme-substrate intermediate. Our results show that the two cysteines that form a structural disulfide bond can play different roles in the oxidative protein-folding reaction catalyzed by a thioredoxin superfamily member in vivo. Moreover, we provide evidence that the mode of translocation can affect the protein-folding process in this compartment. Our detection and characterization of the intermediates reveal novel features of protein folding in an extracytosolic compartment.

Molecular Cell
Forum

How To Choose a Good Scientific Problem

Uri Alon[1,*]
[1]Department Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel
*Correspondence: urialon@weizmann.ac.il
DOI 10.1016/j.molcel.2009.09.013

Choosing good problems is essential for being a good scientist. But what is a good problem, and how do you choose one? The subject is not usually discussed explicitly within our profession. Scientists are expected to be smart enough to figure it out on their own and through the observation of their teachers. This lack of explicit discussion leaves a vacuum that can lead to approaches such as choosing problems that can give results that merit publication in valued journals, resulting in a job and tenure.

The premise of this essay is that a fuller discussion of our topic, including its subjective and emotional aspects, can enrich our science, and our well-being. A good choice means that you can competently discover new knowledge that you find fascinating and that allows self-expression.

We will discuss simple principles of choosing scientific problems that have helped me, my students, and many fellow scientists. These principles might form a basis for teaching this subject generally to scientists.

Starting Point: Choosing a Problem Is an Act of Nurturing

What is the goal of starting a lab? It is sometimes easy to pick up a default value, common in current culture, such as "The goal of my lab is to publish the maximum number of papers of the highest quality."

However, in this essay, we will frame the goal differently: "A lab is a nurturing environment that aims to maximize the potential of students as scientists and as human beings."

Choices such as these are crucial. From values—even if they are not consciously stated—flow all of the decisions made in the lab, big and small: how the lab looks, when students can take a vacation, and (as we will now discuss) what problems to choose. Within the nurturing lab, we aim to choose a problem for our students (and for ourselves) in order to foster growth and self-motivated research.

The Two Dimensions of Problem Choice

To choose a scientific problem, let us begin with a simple graph, as a starting point for discussion (Figure 1). We will compare problems by imagining two axes. The first is *feasibility*—that is, whether a problem is hard or easy, in units such as the expected time to complete the project. This axis is a function of the skills of the researchers and of the technology in the lab. It is important to remember that problems that are easy on paper are often hard in reality, and that problems that are hard on paper are nearly impossible in reality.

The second axis is *interest*: the increase in knowledge expected from the project. We generally value science that ventures deep into unknown waters. Problems can be ranked in terms of the distance from the known shores, by the amount in which they increase verifiable knowledge. We will call this the interest of the problem.

In a forthcoming section, we will discuss the subjective nature of the interest axis. But first, let us first consider aspects of problem choice using our diagram.

Looking at the range of problems in this two-dimensional space, one sees that many projects in current research are of the easy-but-not-too-interesting variety, also known as "low-hanging fruit." Many other projects in science today are unfortunately both difficult and have low interest, partially stemming from a view that hard equals good. A few problems are grand challenges: tough problems with the potential to considerably advance understanding. But most often we would like problems in the top-right quadrant, both feasible and with high interest, likely to extend our knowledge significantly.

The diagram suggests a way to choose between problems, using the Pareto front principle of optimization theory. If problem A is better on both axes than problem B, one can erase B from the diagram. Applying this criterion to all problems, one is left only with problems for which there are no problems clearly better in both feasibility and interest. These remaining problems are on the Pareto front.

To decide which problem to select along the front depends on how we weigh the two axes. For example, a beginning graduate student needs a problem that is easy; positive feedback can thus be rapidly provided, bolstering confidence. These problems are on the bottom right of the Pareto front. The second problem in graduate school can move up the interest axis. Postdocs need projects in the top-right quadrant, since time is limited. Beginning PIs, who need to select a field on which to spend many years and with which to train students, may seek a grand challenge that can be divided into many good, smaller projects. Thus, the optimal problems move along the Pareto front as a function of the life stages of the scientist.

Take Your Time

A common mistake made in choosing problems is taking the first problem that comes to mind. Since a typical project takes years even it if seems doable in months, rapid choice leads to much frustration and bitterness in our profession. It takes time to find a good problem, and every week spent in choosing one can save months or years later on.

In my lab, we have a rule for new students and postdocs: *Do not commit to a problem before 3 months have elapsed*. In these 3 months the new

Molecular Cell
Forum

How To Choose a Good Scientific Problem

Uri Alon[1,*]
[1]Department Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel
*Correspondence: urialon@weizmann.ac.il
DOI 10.1016/j.molcel.2009.09.013

Choosing good problems is essential for being a good scientist. But what is a good problem, and how do you choose one? The subject is not usually discussed explicitly within our profession. Scientists are expected to be smart enough to figure it out on their own and through the observation of their teachers. This lack of explicit discussion leaves a vacuum that can lead to approaches such as choosing problems that can give results that merit publication in valued journals, resulting in a job and tenure.

The premise of this essay is that a fuller discussion of our topic, including its subjective and emotional aspects, can enrich our science, and our well-being. A good choice means that you can competently discover new knowledge that you find fascinating and that allows self-expression.

We will discuss simple principles of choosing scientific problems that have helped me, my students, and many fellow scientists. These principles might form a basis for teaching this subject generally to scientists.

Starting Point: Choosing a Problem Is an Act of Nurturing

What is the goal of starting a lab? It is sometimes easy to pick up a default value, common in current culture, such as "The goal of my lab is to publish the maximum number of papers of the highest quality."

However, in this essay, we will frame the goal differently: "A lab is a nurturing environment that aims to maximize the potential of students as scientists and as human beings."

Choices such as these are crucial. From values—even if they are not consciously stated—flow all of the decisions made in the lab, big and small: how the lab looks, when students can take a vacation, and (as we will now discuss) what problems to choose. Within the nurturing lab, we aim to choose a problem for our students (and for ourselves) in order to foster growth and self-motivated research.

The Two Dimensions of Problem Choice

To choose a scientific problem, let us begin with a simple graph, as a starting point for discussion (Figure 1). We will compare problems by imagining two axes. The first is *feasibility*—that is, whether a problem is hard or easy, in units such as the expected time to complete the project. This axis is a function of the skills of the researchers and of the technology in the lab. It is important to remember that problems that are easy on paper are often hard in reality, and that problems that are hard on paper are nearly impossible in reality.

The second axis is *interest*: the increase in knowledge expected from the project. We generally value science that ventures deep into unknown waters. Problems can be ranked in terms of the distance from the known shores, by the amount in which they increase verifiable knowledge. We will call this the interest of the problem.

In a forthcoming section, we will discuss the subjective nature of the interest axis. But first, let us first consider aspects of problem choice using our diagram.

Looking at the range of problems in this two-dimensional space, one sees that many projects in current research are of the easy-but-not-too-interesting variety, also known as "low-hanging fruit." Many other projects in science today are unfortunately both difficult and have low interest, partially stemming from a view that hard equals good. A few problems are grand challenges: tough problems with the potential to considerably advance understanding. But most often we would like problems in the top-right quadrant, both feasible and with high interest, likely to extend our knowledge significantly.

The diagram suggests a way to choose between problems, using the Pareto front principle of optimization theory. If problem A is better on both axes than problem B, one can erase B from the diagram. Applying this criterion to all problems, one is left only with problems for which there are no problems clearly better in both feasibility and interest. These remaining problems are on the Pareto front.

To decide which problem to select along the front depends on how we weigh the two axes. For example, a beginning graduate student needs a problem that is easy; positive feedback can thus be rapidly provided, bolstering confidence. These problems are on the bottom right of the Pareto front. The second problem in graduate school can move up the interest axis. Postdocs need projects in the top-right quadrant, since time is limited. Beginning PIs, who need to select a field on which to spend many years and with which to train students, may seek a grand challenge that can be divided into many good, smaller projects. Thus, the optimal problems move along the Pareto front as a function of the life stages of the scientist.

Take Your Time

A common mistake made in choosing problems is taking the first problem that comes to mind. Since a typical project takes years even if it seems doable in months, rapid choice leads to much frustration and bitterness in our profession. It takes time to find a good problem, and every week spent in choosing one can save months or years later on.

In my lab, we have a rule for new students and postdocs: *Do not commit to a problem before 3 months have elapsed.* In these 3 months the new

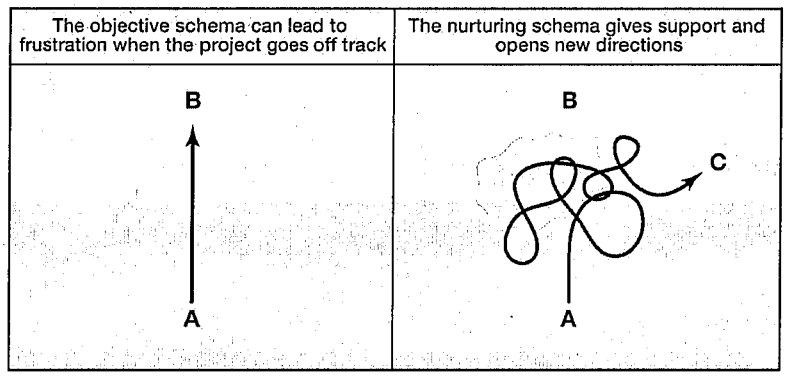

Figure 2. The Objective and Nurturing Schemas of Research
The nurturing schema includes "the cloud"—a period of time in which basic assumptions break down.

at point A, which is the question, and proceeds by the shortest path to point B, the answer. There is a danger, if one accepts this schema, to regard students as a means to an end (an arrow to B). Furthermore, for those that hold this schema, any deviation from the path (experiments that don't work, students that become depressed, etc.) is intolerable. Deviation causes stress because of the cognitive dissonance between reality and the mental schema.

However, one can adopt a second schema, one that resembles more the course of most projects. As before, one starts at point A and moves toward the goal at point B. Soon enough, things move off course, and the path meanders and loops back. Experiments stop working, all assumptions seem wrong, and nothing makes sense. The researcher has entered a phase linked with negative emotions that may be called "the cloud."

Then, in the midst of confusion, one senses a new problem in the materials at hand. Let's call this new problem C. If C is more interesting and feasible than B, one can choose to go toward it. After a few more detours, C is reached. The researchers can pause to celebrate before taking time to think about the next problem.

In this second schema, the meandering of research is seen as an integral part of our craft, rather than a nuisance. The mentors' task is to support students through the cloud that seems to guard the entry into the unknown. And, with this schema, we have more space to see that problem C exists and may be more worthwhile than continuing to plod toward B.

In the nurturing schema, we celebrate the courage and openness of scientists. Sailing into the unknown again and again takes courage; seeing there something different from expectations, and usually more rich and strange, requires uncommon openness.

In summary, take your time (recall the 3 Month Rule) to find among the problems available the one that is most feasible and most interesting to you rather than to others. A good project draws upon your skills to achieve self-expression.

ACKNOWLEDGMENTS

The ideas in this essay were presented to me as gifts in conversations and books, or are the fruit of learning from my mistakes, and are collected here and again offered as a gift. Especially memorable are discussions with Ron Milo; Galit Lahav; Becky Ward; Yuvalal Liron; Michael Elowitz; Angela DePace; Evelyn Fox Keller and her writings, especially *Reflections on Gender and Science*; and with members of my lab and colleagues who told me stories of mentoring and problem choice. I would also like to thank my parents; Galia Moran and our daughter Gefen; and mentors I. Balberg, Dov Shvarts, David Mukamel, and Stan Leibler; Harvard's Positive Psychology taught by Tal Ben-Shahar, 2008; Dan MacAdams for his books *The Person: A New Introduction to Personality Psychology* and *The Narrative Study of Lives*; Amir Orian and The Open Circle approach to theatre and creative arts, classes of 2005/2006; Jonathan Fox for Playback theatre and his book *Acts of Service*; Jerome Bruner for his book *Acts of Meaning*, Erik Erikson for *Childhood and Society*; The Weizmann Institute for providing freedom to play; Mark Kirschner and the Harvard Medical School Department of Systems Biology for hospitality and a place to discuss these ideas with a well-prepared audience; and critical remarks by audience members in Janelia Farms who helped sharpen the message.

Molecular Cell
Review

Transcription Dynamics

Gordon L. Hager,[1,2,*] James G. McNally,[1,2,*] and Tom Misteli[1,2,*]
[1]National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA
[2]The authors contributed equally to this work
*Correspondence: hagerg@mail.nih.gov (G.L.H.), mcnallyj@mail.nih.gov (J.G.M.), mistelit@mail.nih.gov (T.M.)
DOI 10.1016/j.molcel.2009.09.005

All aspects of transcription and its regulation involve dynamic events. The basal transcription machinery and regulatory components are dynamically recruited to their target genes, and dynamic interactions of transcription factors with chromatin—and with each other—play a key role in RNA polymerase assembly, initiation, and elongation. These short-term binding dynamics of transcription factors are superimposed by long-term cyclical behavior of chromatin opening and transcription factor-binding events. Its dynamic nature is not only a fundamental property of the transcription machinery, but it is emerging as an important modulator of physiological processes, particularly in differentiation and development.

Introduction

Transcription is the key step in the regulation of gene expression. The transcription process involves several distinct steps. The core promoter structure contains sequences that serve to anchor a set of protein complexes generally referred to as the general transcription factors (Smale and Kadonaga, 2003; Kornberg, 2005). These factors lead to the formation of a preinitiation complex, which can be quite stable when assembled in vitro. RNA polymerases recognize in turn these multimeric complexes and initiate the synthesis of RNA. The rate of initiation is generally considered the primary regulatory step of transcription, but alternative processes are now known to modulate the accumulation of transcripts. Enhancers are positioned at distant sites in chromatin and modify the rate of initiation complex formation by mechanisms that are still poorly understood (Roeder, 2005). The elongating polymerase complex can be transiently retarded, leading to a "paused" complex that may be subject to reactivation (Price, 2008). The elongation complex brings with it several activities that modulate its postinitiation activity (Shilatifard et al., 2003; Price, 2008). Furthermore, during RNA chain extension, the elongating complex must pass through nucleosome structures, and activities are specifically recruited that allow for their transient disassembly and reformation (Carrozza et al., 2005). Complex histone modifications also take place during elongation (Guenther et al., 2007), and appear to interact with the extending polymerase, affecting its progress. Lastly, the termination and polyadenylation of transcripts represents a final step in the generation of the primary transcript (Rosonina et al., 2006).

The process of transcription is intrinsically dynamic. Yet, most of what we know about how the transcription machinery is assembled, how it initiates, and how it elongates along a gene comes from mostly static biochemical investigations. While these methods have been invaluable in defining the key factors involved in the transcription process and their interactions, they are not ideally suited to gain insight into the real-time kinetics of transcription. The reliance on purification approaches and in vitro reconstitution also raises the question of how accurately findings using these methods reflect the complex environment in which transcription takes place in an intact living cell. The application of recently developed cell biological methods, mostly based on in vivo imaging, to the study of transcription in its natural context and in real-time has overcome some of these limitations (Misteli, 2001; Darzacq et al., 2009). These new methods are now providing first insights into how transcription occurs in a live cell nucleus.

How Transcription Factors Find Their Targets: 3D Genome Scanning

The basis of all transcriptional activity and regulation is the recruitment of transcription complexes to target genes. The basal transcription machinery associates with well-defined binding sites in promoter regions, and regulatory factors bind to specific sites in control elements in the vicinity and, at times, at long distances away, from target genes. Specific binding sites for both the basal machinery, as well as gene-specific regulators, are exceedingly sparse in the genome compared to the number of nonspecific binding sites with which a given transcription factor (TF) may interact. Conservatively, assuming an average mammalian core promoter size of ~150 nt, promoter regions make up less than 0.1% of the human genome, and many TFs have only a few specific binding sites in the genome. How then do TFs find, often rapidly and in response to tightly controlled physiological signaling cascades, their few specific binding sites in the vast sea of nontarget sites in the genome? The key to efficient recruitment of the transcription machinery to its target site are two fundamental dynamic properties of TFs: their ability to rapidly diffuse through the nucleus and their propensity to very transiently bind to chromatin.

Diffusion is the prime means by which TFs move through the nucleus (Misteli, 2001; Gorski et al., 2006). FRAP (fluorescence recovery after photobleaching) experiments in which the motion of a fluorescently tagged TF is traced in living cells have revealed that most TFs move rapidly within the nucleus (Phair et al., 2004; Sprague et al., 2004; Hoogstraten et al., 2002; Stenoien et al., 2001). TF motion is not directional and does not require energy. Measured diffusion coefficients for TFs range from ~0.5 to 5 $\mu m^2 s^{-1}$ depending on a molecule's shape, size, and its interactions with chromatin (Gorski et al., 2006). To put this into perspective, this diffusion behavior allows a molecule to traverse

J|A|C|S
ARTICLES

Published on Web 09/29/2009

Single Molecule Nanocontainers Made Porous Using a Bacterial Toxin

Burak Okumus,[†,§] Sinan Arslan,[†] Stephanus M. Fengler,[⊥] Sua Myong,[†] and Taekjip Ha*[,†]

Center for Biophysics and Computational Biology, University of Illinois, Urbana, Illinois 61801, Department of Physics and Center for the Physics of Living Cells, Urbana, Illinois 61801, Department of Theoretical and Computational Biophysics, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11 D-3707, Göttingen, Germany, Institute for Genomic Biology, University of Illinois, Champaign, Illinois 61801, and Howard Hughes Medical Institute, Urbana, Illinois 61801

Received May 25, 2009; E-mail: tjha@illinois.edu

Abstract: Encapsulation of a biological molecule or a molecular complex in a vesicle provides a means of biofriendly immobilization for single molecule studies and further enables new types of analysis if the vesicles are permeable. We previously reported on using DMPC (dimyristoylphosphatidylcholine) vesicles for realizing porous bioreactors. Here, we describe a different strategy for making porous vesicles using a bacterial pore-forming toxin, α-hemolysin. Using RNA folding as a test case, we demonstrate that protein-based pores can allow exchange of magnesium ions through the vesicle wall while keeping the RNA molecule inside. Flow measurements indicate that the encapsulated RNA molecules rapidly respond to the change in the outside buffer condition. The approach was further tested by coencapsulating a helicase protein and its single-stranded DNA track. The DNA translocation activity of *E. coli* Rep helicase inside vesicles was fueled by ATP provided outside the vesicle, and a dramatically higher number of translocation cycles could be observed due to the minuscule vesicle volume that facilitates rapid rebinding after dissociation. These pores are known to be stable over a wide range of experimental conditions, especially at various temperatures, which is not possible with the previous method using DMPC vesicles. Moreover, engineered mutants of the utilized toxin can potentially be exploited in the future applications.

Introduction

Single molecule fluorescence techniques are revolutionizing biological inquiries both in vitro and in vivo.[1−4] In order to observe single molecule reactions for an extended period, molecular movements via diffusion have to be slower than the time scale of the method used to track their movements. Although it is possible to track particles in one dimension, for example cytoskeleton motor proteins moving on actin filaments[5] or microtubules,[6] or in two dimensions, as in the case of membrane protein diffusion,[7] there are many other examples where three-dimensional diffusion would preclude an extended observation time window. Therefore, several methodologies have been developed to keep the molecules of interest under observation for prolonged periods. Immobilization via biotin−streptavidin anchors on bare[8] or polymer-grafted surfaces[9] and confinement within agarose or polyacrylamide gels[10] are very useful strategies, yet care should be taken to ensure minimal perturbation to the activity of biological molecules. A recent report[11] described an elegant technology for the confinement of individual molecules in free solution by means of anti-Brownian electrokinetic trapping (ABEL). ABEL, however, does not allow changing the solution condition while keeping the molecule trapped. An alternative approach entraps a molecule within a small unilamellar vesicle (SUV) which is then tethered on surfaces.[12−14] Vesicle encapsulation has been used to study

[†] University of Illinois.
[§] Present address: Department of Systems Biology, Harvard Medical School, Boston, MA 02115.
[⊥] Max Planck Institute for Biophysical Chemistry.

(1) Joo, C.; McKinney, S. A.; Nakamura, M.; Rasnik, I.; Myong, S.; Ha, T. *Cell* **2006**, *126*, 515–27.
(2) Joo, C.; Balci, H.; Ishitsuka, Y.; Buranachai, C.; Ha, T. *Annu. Rev. Biochem.* **2008**, *77*, 51–76.
(3) Moerner, W. E. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 12596–602.
(4) Xie, X. S.; Choi, P. J.; Li, G. W.; Lee, N. K.; Lia, G. *Annu. Rev. Biophys.* **2008**, *37*, 417–44.
(5) Yildiz, A.; Forkey, J. N.; McKinney, S. A.; Ha, T.; Goldman, Y. E.; Selvin, P. R. *Science* **2003**, *300*, 2061–5.
(6) Yildiz, A.; Tomishige, M.; Vale, R. D.; Selvin, P. R. *Science* **2004**, *303*, 676–8.
(7) Lommerse, P. H.; Snaar-Jagalska, B. E.; Spaink, H. P.; Schmidt, T. *J. Cell Sci.* **2005**, *118*, 1799–809.
(8) Lee, J. Y.; Okumus, B.; Kim, D. S.; Ha, T. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 18938–43.
(9) Amirgoulova, E. V.; Groll, J.; Heyes, C. D.; Ameringer, T.; Rocker, C.; Moller, M.; Nienhaus, G. U. *ChemPhysChem* **2004**, *5*, 552–555.
(10) Dickson, R. M.; Norris, D. J.; Tzeng, Y. L.; Moerner, W. E. *Science* **1996**, *274*, 966–9.
(11) Cohen, A. E.; Moerner, W. E. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 4362–5.
(12) Okumus, B.; Wilson, T. J.; Lilley, D. M.; Ha, T. *Biophys. J.* **2004**, *87*, 2798–806.
(13) Boukobza, E.; Sonnenfeld, A.; Haran, G. *J. Phys. Chem. B* **2001**, *105*, 12165–12170.
(14) Benitez, J. J.; Keller, A. M.; Ochieng, P.; Yatsunyk, L. A.; Huffman, D. L.; Rosenzweig, A. C.; Chen, P. *J. Am. Chem. Soc.* **2008**, *130*, 2446–2447.

10.1021/ja9042356 CCC: $40.75

JB Accepts, published online ahead of print on 6 November 2009
J. Bacteriol. doi:10.1128/JB.01245-09

1 **10-28-09**

2 **P1 plasmid segregation: accurate re-distribution by dynamic plasmid**

3 **pairing and separation**

4

5 Running title

6 P1 plasmid segregation.

7

8 **Manjistha Sengupta, Henrik Jorck Nielsen, Brenda Youngren and**

9 **Stuart Austin***

10

11 *Gene Regulation and Chromosome Biology Laboratory, National Cancer Institute, CCR,*

12 *NCI-Frederick, Frederick, Maryland 21702-1201, USA.*

13

14

15 **Corresponding author.* Mailing address: Gene Regulation and Chromosome Biology

16 Laboratory, National Cancer Institute, CCR, NCI-Frederick, Frederick, MD 21702-1201.

17 Phone: (301) 846 1266. Fax: (301) 846 6988. E-mail: austinst@mail.nih.gov

18

19

20

21

JOURNAL OF BACTERIOLOGY, Nov. 2009, p. 6779–6781 Vol. 191, No. 22
0021-9193/09/$12.00 doi:10.1128/JB.01150-09

COMMENTARIES

A New Twist on a Classic Paradigm: Illumination of a Genetic Switch in *Vibrio cholerae* Phage CTXΦ[▽]

Bryce E. Nickels*

Department of Genetics and Waksman Institute, Rutgers University, Piscataway, New Jersey 08854

Regulatory circuits that control the life cycles of bacteriophage have served as important models for understanding complex regulatory networks in all organisms. Many lysogenic bacteriophage can exist in a dormant state in which the phage DNA is integrated into the bacterial chromosome. Maintenance of this "prophage state" is achieved through finely tuned regulatory circuits built in a manner that enables prophage induction, i.e., the production of phage particles, to occur in a rapid and efficient manner upon exposure to acute cellular stress. The paradigm for such regulatory circuits is the classic "genetic switch" of bacteriophage λ (8), in which prophage induction is governed by a phage-encoded DNA binding protein, λCI, which represses expression of the genes that control phage production and cell lysis. Proper functioning of the λ switch requires that the intracellular levels of λCI be precisely controlled, which is accomplished through the ability of λCI to regulate expression of its own gene through an auto-feedback loop. In this issue of the Journal of Bacteriology, Kimsey and Waldor illuminate features of a regulatory circuit that controls production of the filamentous bacteriophage CTXΦ in *Vibrio cholerae* (5), which bears similarities to the genetic switch of λ. The CTXΦ switch is governed by two transcription factors, the host-encoded SOS response regulator LexA and the phage-encoded repressor RstR. Prior work had suggested that the intracellular levels of RstR must be precisely controlled to ensure the proper functioning of the CTXΦ switch. Kimsey and Waldor's study now provides evidence that control of the intracellular levels of RstR is mediated, in part, through a unique LexA-dependent auto-feedback loop. Kimsey and Waldor propose that these novel features of the CTXΦ switch allow the circuit to exhibit transient, reversible behavior upon induction.

***Vibrio cholerae* bacteriophage CTXΦ.** The lysogenic filamentous phage CTXΦ, which carries genes encoding cholera toxin, infects the gram-negative bacterium *Vibrio cholerae*, the causative agent of epidemic cholera (12). The activity of cholera toxin is primarily responsible for the profuse secretory diarrhea that is the hallmark of cholera and which enables the dissemination of *Vibrio cholerae*. Thus, the horizontal transfer of CTXΦ is an important contributor to the emergence of new pathogenic strains of *Vibrio cholerae*.

CTXΦ is an unusual filamentous bacteriophage because its DNA becomes incorporated into the *Vibrio cholerae* genome, leading to the establishment of a lysogenic program of gene expression (7). In contrast to other lysogenic phage such as bacteriophage λ, CTXΦ does not enter into a lytic cycle and kill its host. Rather, CTXΦ and *Vibrio cholerae* seem to have coevolved to the mutual benefit of both organisms (7). In particular, the replication and dissemination of the host *Vibrio cholerae* are facilitated by the production of cholera toxin from the CTXΦ chromosome. Furthermore, *Vibrio cholerae*-encoded factors are required throughout the CTXΦ life cycle, enabling the stable integration of CTXΦ DNA into the bacterial chromosome and the production and secretion of phage particles.

The CTXΦ switch: the roles of RstR and LexA. In the context of the CTXΦ lysogen, the genes required for CTXΦ replication and morphogenesis, under the control of phage promoter P_A, are repressed. During prophage induction, triggered by agents that cause DNA damage, P_A transcription is derepressed, resulting ultimately in the production of phage particles that are secreted from the cell (2, 7, 11). Because prophage induction does not result in cell lysis, CTXΦ presumably is able to reestablish the lysogenic program and once again enter into a quiescent state. Thus, unlike the genetic switch of λ, which once flipped, irreversibly commits the phage to the lytic program, the CTXΦ switch must have distinct features that enable it to display reversible behavior upon induction.

Prior work has established that control of gene expression from P_A is governed by two DNA-binding proteins, the phage-encoded RstR and the host-encoded SOS response regulator LexA (Fig. 1A). In the context of the lysogen, RstR and LexA function in an independent manner to repress P_A transcription. RstR inhibits transcription from P_A by binding to the following three sites in the P_A promoter region (4): a high-affinity operator, *O1*, and two weaker operators, *O2* and *O3*. LexA inhibits transcription from P_A by binding a single site (SOS box) that overlaps with the RstR *O2* operator (9), and as expected based on this overlap, LexA binding to the SOS box and RstR binding to the *O2* operator are mutually exclusive (9).

Prophage induction in response to DNA damage occurs via the RecA-stimulated autocleavage of LexA, which leads to partial derepression of P_A (9). This finding suggested that under normal growth conditions, the intracellular levels of LexA and RstR are such that LexA occupies the SOS box DNA. Furthermore, because the accumulation of high levels of RstR would cause RstR to occupy the *O2* site and prevent LexA from binding to the SOS box, these findings raised the

* Mailing address: Department of Genetics and Waksman Institute, Rutgers University, Piscataway, NJ 08854. Phone: (732) 445-6852. Fax: (732) 445-5735. E-mail: bnickels@waksman.rutgers.edu.

▽ Published ahead of print on 11 September 2009.

JOURNAL OF BACTERIOLOGY, Nov. 2009, p. 6665–6674 Vol. 191, No. 21
0021-9193/09/$12.00 doi:10.1128/JB.00344-09

Transcriptional Regulation of the *Escherichia coli* Gene *rraB*, Encoding a Protein Inhibitor of RNase E▽‡

Li Zhou,[1]† Meng Zhao,[2]†§ Rachel Z. Wolf,[2] David E. Graham,[2] and George Georgiou[1,2,3,4]*

Section of Microbiology and Molecular Genetics,[1] Institute for Cell and Molecular Biology,[2] Department of Chemical Engineering,[3] and Department of Biomedical Engineering,[4] University of Texas at Austin, Austin, Texas 78712

Received 11 March 2009/Accepted 11 August 2009

The *Escherichia coli* RNA degradosome is a protein complex that plays a critical role in the turnover of numerous RNAs. The key component of the degradosome complex is the endoribonuclease RNase E, a multidomain protein composed of an N-terminal catalytic region and a C-terminal region that organizes the other protein components of the degradosome. Previously, the RNase E inhibitors RraA and RraB were identified genetically and shown to bind to the C-terminal region of RNase E, thus affecting both the protein composition of the degradosome and the endonucleolytic activity of RNase E. In the present work, we investigated the transcriptional regulation of *rraB*. *rraB* was shown to be transcribed constitutively from its own promoter, P*rraB*. Transposon mutagenesis and screening for increased β-galactosidase activity from a chromosomal P*rraB-lacZ* transcriptional fusion resulted in the isolation of a transposon insertion in *glmS*, encoding the essential enzyme glucosamine-6-phosphate synthase that catalyzes the first committed step of the uridine 5′-diphospho-*N*-acetyl-glucosamine (UDP-GlcNAc) pathway, which provides intermediates for peptidoglycan biogenesis. The *glmS852*::Tn*5* allele resulted in an approximately 50% lower intracellular concentration of UDP-GlcNAc and conferred a fivefold increase in the level of *rraB* mRNA. This allele also mediated a twofold increase in β-galactosidase activity from a chromosomal fusion of the 5′ untranslated region of the *rne* gene to *lacZ*, suggesting that a reduction in cellular concentration of UDP-GlcNAc and the resulting increased expression of RraB might modulate the action of RNase E.

The endoribonuclease RNase E plays a central role in RNA metabolism, including the processing of rRNAs and tRNAs (30, 42, 43); the degradation of small regulatory RNAs; and, most importantly, the turnover of numerous cellular mRNAs in *Escherichia coli* (5, 13, 23, 33). Homologous RNase E has been identified in more than 50 bacteria, archaea, and plants (28). The 1,061-amino-acid *E. coli* RNase E protein can be divided into functional portions, from the N terminus to the C terminus, as follows. The N-terminal half (amino acid residues 1 to 529) contains the endonuclease active site (amino acid residues 1 to 395) and a zinc finger region (amino acid residues 400 to 415) (8, 9). The central region includes the membrane anchoring segment A (amino acid residues 565 to 582) and its flanking portions (25) as well as an arginine-rich RNA binding site (amino acid residues 604 to 688) (34). The C-terminal half (amino acid residues 734 to 1061) is an unstructured scaffold domain that contains binding sites for the other core degradosome components, namely, polynucleotide phosphorylase, the RhlB helicase, and the glycolytic enzyme enolase (11, 35, 45, 46, 53). Previous studies indicate that the assembled degradosome complex is necessary for normal mRNA degradation and the degradosome components functionally interact during decay of at least some RNAs in *E. coli* (4, 24).

The cellular level and activity of RNase E in *E. coli* are under complex regulation. First, the efficiency of RNase E cleavage depends on structural features of RNA substrates and the factors that affect the accessibility of putative cleavage sites. A 5′ monophosphate in substrate RNAs serves as an allosteric activator of the endonuclease activity (10, 21). Interactions of RNA substrates with Hfq and small RNAs exert an important role on the cleavage of certain mRNAs (1, 7, 39). Second, RNase E autoregulates its synthesis by modulating the decay of its own mRNA. Previous studies showed that the 5′ untranslated region (UTR) of *rne* mRNA is subject to RNase E degradation; i.e., a higher cellular level of RNase E results in faster degradation of its transcript, which in turn leads to a reduction in the RNase E level (21, 40). Third, the membrane localization of RNase E and its association with the bacterial cytoskeleton may affect its function through various mechanisms (25, 31, 51). Finally, our lab has shown that the activity of RNase E is affected globally by the endoribonuclease binding proteins RraA and RraB, which control the decay and abundance of large numbers of bacterial mRNAs in *trans* (16, 29, 54, 56).

RraB, previously annotated as YjgD in the NCBI database, is a 15.6-kDa protein that interacts with RNase E and inhibits RNase E endonucleolytic cleavages in *E. coli*. In contrast to RraA homologues, which exist in numerous bacterial genomes, RraB is found only in gammaproteobacteria, suggesting that the latter protein may have a more specialized role in modulating RNA degradation. Similar to RraA, RraB does not alter RNase E cleavage site specificity or interact detectably with the substrate RNAs. RraB also possesses the same affinity for

* Corresponding author. Mailing address: University of Texas at Austin, Austin, TX 78712. Phone: (512) 471-6975. Fax: (512) 471-7963. E-mail: gg@che.utexas.edu.

† These authors contributed equally to this work.

§ Present address: Internal Medicine Residency Program, University of Texas Medical Branch Austin Programs, University Medical Center at Brackenridge, Austin, TX 78701.

‡ Supplemental material for this article may be found at http://jb.asm.org/.

▽ Published ahead of print on 28 August 2009.

Intraflagellar transport particle size scales inversely with flagellar length: revisiting the balance-point length control model

Benjamin D. Engel, William B. Ludington, and Wallace F. Marshall

Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA 94158

The assembly and maintenance of eukaryotic flagella are regulated by intraflagellar transport (IFT), the bidirectional traffic of IFT particles (recently renamed IFT trains) within the flagellum. We previously proposed the balance-point length control model, which predicted that the frequency of train transport should decrease as a function of flagellar length, thus modulating the length-dependent flagellar assembly rate. However, this model was challenged by the differential interference contrast microscopy observation that IFT frequency is length independent. Using total internal reflection fluorescence microscopy to quantify protein traffic during the regeneration of *Chlamydomonas reinhardtii* flagella, we determined that anterograde IFT trains in short flagella are composed of more kinesin-associated protein and IFT27 proteins than trains in long flagella. This length-dependent remodeling of train size is consistent with the kinetics of flagellar regeneration and supports a revised balance-point model of flagellar length control in which the size of anterograde IFT trains tunes the rate of flagellar assembly.

Introduction

How cells regulate organelle size is a fundamental question in cell biology. The eukaryotic flagellum (a term used interchangeably with cilium), with its easily measured linear geometry, provides an ideal model for studying organelle size control. Furthermore, ciliary length control has become increasingly medically relevant, with recent studies linking ciliary defects to a wide range of human disorders (Pazour and Rosenbaum, 2002; Bisgrove and Yost, 2006).

The unicellular biflagellate green alga *Chlamydomonas reinhardtii* is a powerful tool for studying flagellar biology (Randall et al., 1969; Silflow and Lefebvre, 2001). After abscission, flagella regenerate to their original length with reproducible kinetics (Fig. 1 A). Short flagella undergo a period of rapid growth then transition to a slow elongation phase as they near their steady-state length (Fig. 1 B). Although this phenomenon was first observed 40 yrs ago (Rosenbaum et al., 1969), the mechanisms that control the kinetics of flagellar regeneration are not well understood. Studies in *C. reinhardtii* have revealed that the assembly and maintenance of flagella depend on intraflagellar transport (IFT), the molecular motor-driven process of bidirectional protein traffic within the flagellum (Kozminski et al., 1993). IFT describes the movement of IFT trains (previously referred to as "particles" but renamed "trains" based on electron tomography; Pigino et al., 2009), heterogeneous linear protein arrays composed of heterotrimeric kinesin-2, cytoplasmic dynein-1b, axonemal cargo, and two varieties of IFT complexes (A and B; Piperno and Mead, 1997; Cole et al., 1998; Cole, 2003; Ou et al., 2007). Kinesin powers anterograde transport to the distal tip of the flagellum, where the trains are remodeled and subsequently undergo dynein-driven retrograde return to the base. IFT has been shown to be directly responsible for the transport of axonemal precursors to the site of flagellar assembly at the tip (Qin et al., 2004; Hou et al., 2007). Defects in kinesin, dynein, and most of the IFT proteins all lead to short or absent flagella (Huang et al., 1977; Piperno et al., 1998; Pazour et al., 1999, 2000; Porter et al., 1999; Deane et al., 2001; Pedersen et al., 2005). Additionally, reducing the speed and frequency of IFT leads to flagellar shortening (Kozminski et al., 1995; Iomini et al., 2001). Thus, IFT appears to play a central role in mediating flagellar length.

B.D. Engel and W.B. Ludington contributed equally to this paper.

Correspondence to Wallace F. Marshall: wallace.marshall@biochem.ucsf.edu

Abbreviations used in this paper: CCD, charge-coupled device; DIC, differential interference contrast; IFT, intraflagellar transport; KAP, kinesin-associated protein; ROI, region of interest; TAP, Tris-acetate-phosphate; TIRF, total internal reflection fluorescence.

The Rockefeller University Press $30.00
J. Cell Biol. Vol. 187 No. 1 81–89
www.jcb.org/cgi/doi/10.1083/jcb.200812084

Supplemental Material can be found at:
http://jcb.rupress.org/cgi/content/full/jcb.200812084/DC1

Lab on a Chip

RSC Publishing

Journals Lab on a Chip Advance Articles DOI: 10.1039/b917503b

Lab Chip, 2009 | DOI: 10.1039/b917503b | Communication

Counting single molecules in sub-nanolitre droplets†

T. D. Rane‡ [a], C. M. Puleo‡ [a], K. J. Liu [a], Y. Zhang [a], A. P. Lee [b] and T. H. Wang *[ac]

[a] *Johns Hopkins University, Department of Biomedical Engineering, 3400 N. Charles St., Clark Hall, Baltimore, MD, USA. E-mail: tushar@jhu.edu, cpuleo@jhu.edu, Kelvin@jhu.edu, yzhang68@jhmi.edu; Fax: +1 410-516-4771; Tel: +1 410-516-7576*

[b] *University of California, Irvine, Departments of Biomedical Engineering and Mechanical/Aerospace Engineering, ET 716F, Irvine, CA, USA. E-mail: aplee@uci.edu; Fax: +1 949-824-1727; Tel: +1 949-824-9691*

[c] *Johns Hopkins University, Departments of Mechanical Engineering and Biomedical Engineering, 3400 N. Charles St., Latrobe Hall Rm. 108, Baltimore, MD, USA. E-mail: thwang@jhu.edu; Fax: +1 410-516-4316; Tel: +1 410-516-7086*

Received 26th August 2009 , Accepted 2nd October 2009

First published on the web 9th October 2009

We demonstrate single biomolecule detection and quantification within sub-nanolitre droplets through application of Cylindrical Illumination Confocal Spectroscopy (CICS) and droplet confinement within a retractable microfluidic constriction.

Droplet-based microfluidic platforms offer the distinct advantages of fast sample mixing, limited reagent dispersion, and lower sample loss or contamination when compared to traditional microfluidic devices.[1,2] In addition, simple control strategies for kHz frequency droplet generation, transportation, storage, and sorting give the platform a propensity for high throughput analysis. While initial applications have been shown in a range of research disciplines, including biochemical analysis,[3] chemical and material synthesis,[4] and chemical reactions,[5] more recent applications have emerged that seek to expand this high throughput capability to the analysis of individual biological entities, such as single cells[6–8] or biomolecules.[9,10]

In single cell experiments, *in vitro* compartmentalization in droplets enables rapid accumulation of secreted cellular factors,[11] and provides both chemical isolation and a unique means of cell selection and control.[12] Alternately, single molecule compartmentalization enables nucleic acid analysis through single-copy DNA PCR, extending the high throughput capacity of droplet-based microfluidics to digital PCR assays.[7,9,10] However, the current dependence on amplification techniques (*e.g.* PCR or fluorogenic substrate)[1] for detection of low concentration biomolecules presents limitations in droplet-based platforms; these include decreased throughput and

Molecular Cell
Article

An SOS Inhibitor that Binds to Free RecA Protein: The PsiB Protein

Vessela Petrova,[1] Sindhu Chitteni-Pattu,[2] Julia C. Drees,[2,3] Ross B. Inman,[2] and Michael M. Cox[1,2,*]
[1]Program in Cellular and Molecular Biology
[2]Department of Biochemistry
University of Wisconsin, Madison, Madison, WI 53706, USA
[3]Present address: Department of Laboratory Medicine, University of California, San Francisco, San Francisco, CA 94143, USA
*Correspondence: cox@biochem.wisc.edu
DOI 10.1016/j.molcel.2009.07.026

SUMMARY

The process of bacterial conjugation involves the transfer of a conjugative plasmid as a single strand. The potentially deleterious SOS response, which is normally triggered by the appearance of single-stranded DNA, is suppressed in the recipient cell by a conjugative plasmid system centered on the product of the *psiB* gene. The F plasmid PsiB protein inhibits all activities of the RecA protein, including DNA binding, DNA strand exchange, and LexA protein cleavage. The proteins known to negatively regulate recombinases, such as RecA or Rad51, generally work at the level of dismantling the nucleoprotein filament. However, PsiB binds to RecA protein that is free in solution. The RecA-PsiB complex impedes formation of RecA nucleoprotein filaments on DNA.

INTRODUCTION

Conjugation is a mechanism of horizontal gene transfer that allows asexually reproducing bacteria to diversify their genomes. Conjugative plasmids aid the spread of antibiotic resistance in bacterial strains (Aguero et al., 1984 and references therein), making the process of conjugation an important target for fighting disease (Lujan et al., 2007). In *Escherichia coli*, conjugation occurs between a donor cell harboring a conjugative plasmid and a recipient cell that lacks a similar plasmid (Frost et al., 1994). During conjugation, rolling circle replication ensues, and a single strand of the plasmid is spooled into the recipient (Wolkow et al., 1996 and references therein).

The single-stranded DNA (ssDNA) that enters the recipient cell can trigger the SOS response, involving induction of DNA repair genes that are under the control of the LexA transcription repressor protein (Roca and Cox, 1997; Walker et al., 2000). The LexA protein undergoes autocatalytic cleavage stimulated by RecA protein filaments bound to ssDNA and ATP (Little, 1991; Little et al., 1980; Roca and Cox, 1997; Slilaty and Little, 1987). During SOS, cell division halts (Walker et al., 2000), and the induction of the mutagenic DNA polymerase V is possible if the response is not abated (Sommer et al., 1993; Walker et al., 2000). Therefore, the induction of SOS during conjugation is potentially detrimental to the cell.

Some plasmids can suppress unwarranted SOS induction during conjugation. The Plasmid SOS Interference/Inhibition (psi) phenomenon was first observed when the conjugative plasmid R100.1 suppressed the temperature-sensitive constitutive SOS phenotype of *recA*441 (Bagdasarian et al., 1980). Further studies with the R6-5 plasmid showed that a region encoding the PsiB protein alone was sufficient to suppress the SOS response (Bailone et al., 1988).

PsiB expression varies as a function of the promoters present on a given conjugative plasmid and of the cellular circumstances. The native F plasmid *psiB* promoter appears to be activated only during conjugation and only in the recipient (Bagdasarian et al., 1992). Early transcription of PsiB is likely mediated by an element of secondary structure in the incoming single strand that mimics an RNA polymerase promoter recognized by the Frpo sigma factor (Masai and Arai, 1997).

The strength of PsiB inhibition of SOS depends on the RecA allele and on the concentration of PsiB in the host cell. Constitutive SOS at high temperatures is suppressed in *recA*441 (*tif-1*) cells harboring the R100.1 plasmid. However, the SOS response was still induced in response to DNA damage. The RecA441 mutant protein binds better than does wild-type RecA to ssDNA coated by single-stranded DNA binding protein (SSB) at high temperatures (Lavery and Kowalczykowski, 1990), and this characteristic may be significant in overcoming the inhibition by PsiB protein. In a separate experiment, plasmids derived from the R6-5 plasmid, containing a *psi* region, were able to suppress induction of wild-type *recA* gene after DNA damage (Bagdasarian et al., 1986). Thus, the PsiB protein is less effective in suppressing SOS induction by RecA441 than by wild-type RecA. The effect of PsiB is also dose dependent (Bailone et al., 1988).

Biochemical characterization of the *psiB* gene product has been very limited. PsiB does not inhibit SOS in LexA-deficient cells, ruling out the possibility that it is a transcription repressor (Bagdasarian et al., 1986). On the basis of lack of sequence homology, it has been argued that PsiB is unlikely to act as a LexA mimic, competing directly for the RecA nucleoprotein filament necessary for LexA cleavage (Bailone et al., 1988). Other proposals include a direct PsiB inhibition of RecA function by competing for DNA-binding sites, stabilizing a protein that binds to DNA and competes with RecA (e.g., SSB), or directly binding to RecA and preventing it from binding to DNA (Bagdasarian et al., 1986; Bailone et al., 1988). On the basis of all results, particularly on the unpublished effects of certain RecA mutant proteins, Devoret and colleagues favored an inhibition mediated

Molecular Microbiology (2009) 73(6), 1043–1057 ■

doi:10.1111/j.1365-2958.2009.06830.x
First published online 26 August 2009

The transcriptionally active regions in the genome of *Bacillus subtilis*

Simon Rasmussen, Henrik Bjørn Nielsen and Hanne Jarmer*
Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, 2800 Lyngby, Denmark.

Summary

The majority of all genes have so far been identified and annotated systematically through *in silico* gene finding. Here we report the finding of 3662 strand-specific transcriptionally active regions (TARs) in the genome of *Bacillus subtilis* by the use of tiling arrays. We have measured the genome-wide expression during mid-exponential growth on rich (LB) and minimal (M9) medium. The identified TARs account for 77.3% of the genes as they are currently annotated and additionally we find 84 putative non-coding RNAs (ncRNAs) and 127 antisense transcripts. One ncRNA, *ncr22*, is predicted to act as a translational control on *cstA* and an antisense transcript was observed opposite the housekeeping sigma factor *sigA*. Through this work we have discovered a long conserved 3′ untranslated region (UTR) in a group of membrane-associated genes that is predicted to fold into a large and highly stable secondary structure. One of the genes having this tail is *efeN*, which encodes a target of the twin-arginine translocase (Tat) protein translocation system.

Introduction

The bacterial genome is a highly compact structure. Both strands are densely covered by genes, of which a large part is organized into the even more gene-dense arrangements of operons. Recent technological advances have allowed for an empirical assessment of the prevalence of transcriptionally active regions (TARs) across an entire genome – by the use of either high-throughput sequencing of RNA-derived cDNA (Nagalakshmi *et al.*, 2008) or high-density oligo-nucleotide tiling arrays (Tjaden *et al.*, 2002; Bertone *et al.*, 2004; David *et al.*, 2006; Li *et al.*, 2006; Reppas *et al.*, 2006). Where the studies of Tjaden and co-workers and Reppas and co-workers investigated the transcriptional landscape of *Escherichia coli*, we report here the first findings of a high-density tiling-array study performed on the Gram-positive *Bacillus subtilis*. *B. subtilis* was first described in 1835 by the German scientist Christian Gottfried Ehrenberg as the hay/grass-associated bacterium, *Vibrio subtilis* (Ehrenberg, 1835). In 1872 another German scientist, Ferdinand Julius Cohn, renamed it *Bacillus subtilis* (Cohn, 1872). In 1876 Cohn showed for the first time that *B. subtilis* is capable of changing into an endospore state, and hereby surviving environmental changes not suitable for vegetative growth (Cohn, 1876). In 1930 the American bacteriologist, Harold Joel Conn, published a description of the Marburg strain of *B. subtilis* (American Type Culture Collection No. 6051) (Conn, 1930; Teas, 1949) and in 1947 this particular strain was subjected to both X-rays and UV light by Burkholder and Giles (Burkholder and Giles, 1947; Teas, 1949). Charles Yanofsky provided a number of stable auxotrophs, which had been isolated from these experiments, to John Spizizen (Spizizen, 1984), which studied their ability to develop natural competence (Spizizen, 1958; Zeigler *et al.*, 2008). Further investigations resulted in the development of a highly efficient two-step protocol for transformation of the #168 strain (Anagnostopoulos and Spizizen, 1961), a success drawing the attention of the research community to such an extent that this strain was selected as the *B. subtilis* model strain. Today *B. subtilis* is widely used as an industrial production strain, and has even been shown to possess probiotic properties (Huang *et al.*, 2008). And now, more than 10 years after fully sequencing and annotating the genome the first time (Kunst *et al.*, 1997) and only shortly after the recent re-sequencing (Barbe *et al.*, 2009), we experimentally validate and extend these efforts.

Results and discussion

Identification of transcriptionally active regions (TARs)

Hybridization of labelled RNA to densely tiled microarrays allows for a high-resolution mapping of genome-wide expression on both strands, and we have found that

Accepted 23 July, 2009. *For correspondence. E-mail hanne@cbs.dtu.dk; Tel. (+45) 45 25 61 48; Fax (+45) 45 93 15 85.

Molecular Systems Biology 5; Article number 318; doi:10.1038/msb.2009.75
Citation: *Molecular Systems Biology* 5:318

www.molecularsystemsbiology.com

molecular systems biology

REPORT

Listening to the noise: random fluctuations reveal gene network parameters

Brian Munsky[1,*], Brooke Trinh[2] and Mustafa Khammash[3,*]

[1] Computer, Computational, and Statistical Sciences Division (CCS), and the Theoretical (T) Division at Los Alamos National Laboratory, Los Alamos, NM, USA, [2] Department of Molecular, Cellular and Developmental Biology, University of California, Santa Barbara, CA, USA and [3] Department of Mechanical Engineering and Center for Control, Dynamical Systems and Computations, University of California, Santa Barbara, CA, USA
* Corresponding authors. B Munsky, CCS-3 and CNLS, Los Alamos National Lab, Los Alamos, NM 87545, USA. Tel.: +1 505 665 6691; Fax: +1 505 665 7652; E-mail: brian.munsky@gmail.com or M Khammash, Department of Mechanical Engineering, University of California, Engineering II Building Room 2333, Santa Barbara, CA 93106, USA. Tel.: +1 805 893 4967; Fax: +1 805 893 8651; E-mail: khammash@engr.ucsb.edu

Received 20.2.09; accepted 8.9.09

The cellular environment is abuzz with noise originating from the inherent random motion of reacting molecules in the living cell. In this noisy environment, clonal cell populations show cell-to-cell variability that can manifest significant phenotypic differences. Noise-induced stochastic fluctuations in cellular constituents can be measured and their statistics quantified. We show that these random fluctuations carry within them valuable information about the underlying genetic network. Far from being a nuisance, the ever-present cellular noise acts as a rich source of excitation that, when processed through a gene network, carries its distinctive fingerprint that encodes a wealth of information about that network. We show that in some cases the analysis of these random fluctuations enables the full identification of network parameters, including those that may otherwise be difficult to measure. This establishes a potentially powerful approach for the identification of gene networks and offers a new window into the workings of these networks.
Molecular Systems Biology 5: 318; published online 13 October 2009; doi:10.1038/msb.2009.75
Subject Categories: computational methods; simulation and data analysis
Keywords: gene regulatory networks; stochastic biological processes; system identification

Introduction

Computational modeling in biology seeks to reduce complex systems to their essential components and functions, thereby arriving at a deeper understanding of biological phenomena. However, measuring or estimating key model parameters can be difficult when measurement noise corrupts experimental data. Thus, when cellular variability or 'noise' (Elowitz *et al*, 2002) leads to measurement fluctuations, it may appear deleterious. However, this is not the case. Just as white noise inputs help to identify dynamical system parameters (Ljung, 1999), so too can characterization of noise dynamics elucidate natural mechanisms. For example, steady state noise characteristics can distinguish between different logical structures, such as AND or OR gates (Warmflash and Dinner, 2008). At the same time, temporal measurements of transient dynamics can aid in the construction of reaction pathways (Arkin *et al*, 1997). In combination, noise and temporal analyses yield powerful tools for parameter identification. For example, the averages of correlations in cell expression at many time points reveal feed-forward loops in the galactose metabolism genes of *Escherichia coli* (Dunlop *et al*, 2008). Similarly, manipulating certain gene network transcription rates while observing the response of statistical cumulants can help to identify reaction rates for some gene regulatory networks (Raffard *et al*, 2008). In this study, we examine the possibility of identifying system parameters and mechanisms directly from single-cell distributions, such as those obtainable with flow cytometry, without time-varying control and at only a handful of different time points. We prove that the analysis of variability provides more information than the mean behavior alone. Furthermore, we illustrate potential of our approach using numerical and experimental analyses of common gene regulatory networks.

Results and discussion

Gene expression model

We adopt the gene expression model used in the study carried out by Thattai and van Oudenaarden (2001), which is characterized by random integer numbers of mRNA and protein molecules: R and P, respectively. Transcription, translation, and degradation events change the system state

Molecular Systems Biology 5; Article number 313; doi:10.1038/msb.2009.73
Citation: *Molecular Systems Biology* 5:313

www.molecularsystemsbiology.com

molecular systems biology

REPORT

Quantitative effect of scaffold abundance on signal propagation

Stephen A Chapman and Anand R Asthagiri*

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, USA
* Corresponding author. Division of Chemistry and Chemical Engineering, California Institute of Technology, 1200 E California Blvd, MC 210-41, Pasadena, CA 91125, USA. Tel.: +1 626 395 8130; Fax: +1 626 568 8743; E-mail: anand@cheme.caltech.edu

Received 19.11.08; accepted 9.9.09

Protein scaffolds bring together multiple components of a signalling pathway, thereby promoting signal propagation along a common physical 'backbone'. Scaffolds play a prominent role in natural signalling pathways and provide a promising platform for synthetic circuits. To better understand how scaffolding quantitatively affects signal transmission, we conducted an *in vivo* sensitivity analysis of the yeast mating pathway to a broad range of perturbations in the abundance of the scaffold Ste5. Our measurements show that signal throughput exhibits a biphasic dependence on scaffold concentration and that altering the amount of scaffold binding partners reshapes this biphasic dependence. Unexpectedly, the wild-type level of Ste5 is ~10-fold below the optimum needed to maximize signal throughput. This sub-optimal configuration may be a tradeoff as increasing Ste5 expression promotes baseline activation of the mating pathway. Furthermore, operating at a sub-optimal level of Ste5 may provide regulatory flexibility as tuning Ste5 expression up or down directly modulates the downstream phenotypic response. Our quantitative analysis reveals performance tradeoffs in scaffold-based modules and defines engineering challenges for implementing molecular scaffolds in synthetic pathways.
Molecular Systems Biology 5: 313; published online 13 October 2009; doi:10.1038/msb.2009.73
Subject Categories: metabolic and regulatory networks; signal transduction
Keywords: MAP kinase signalling; pheromone; *Saccharomyces cerevisiae*; signal throughput; Ste5 scaffold

Introduction

Protein scaffolds bind concomitantly to multiple components of a signalling pathway, thereby organizing signal transmission onto a common physical backbone. Scaffold-based modules are broadly used to propagate signals that regulate cell cycle, proliferation, differentiation and motility in species ranging from yeast to human (Pawson and Scott, 1997). Scaffolds are also emerging as a promising platform for engineering synthetic signalling modules. Molecular redesign of scaffolds has been used to alter the repertoire of scaffold binding partners, thereby redirecting signal flow (Park *et al*, 2003) and altering signal dynamics (Bashor *et al*, 2008).

In addition to the molecular design of the scaffold, the quantitative performance of scaffold-based modules will depend on the expression level of the scaffold and its binding partners. Computational models predict that scaffolds may not always promote signal propagation (Levchenko *et al*, 2000). When scaffold concentration exceeds an optimal level, enzymes and substrates are predicted to bind to distinct scaffolds rather than onto a single backbone, thereby weakening signal transmission via combinatorial inhibition.

These model predictions, however, are based on idealized mathematical representations of scaffold-based signalling. In contrast, scaffold-mediated signalling *in vivo* is often far more intricate as exemplified for the prototypical scaffold Ste5 in yeast cells (Figure 1A). The Ste5 scaffold undergoes dimerization, which may contribute to apparent cooperativity (Ferrell, 2000), translocates between different subcellular compartments, which regulates its proteosomal degradation (Garrenton *et al*, 2009), is regulated by Fus3-mediated negative feedback (Bhattacharyya *et al*, 2006) and binds competitively to multiple proteins (Fus3 and Kss1) with different affinities (Kusari *et al*, 2004). This complex array of mechanisms conceals precisely how real scaffolds such as Ste5 quantitatively contribute to signal transmission *in vivo*.

In fact, previous studies have shown that Ste5 overexpression augments signal throughput (Kranz *et al*, 1994; Garrenton

Molecular Systems Biology 5; Article number 323; doi:10.1038/msb.2009.82
Citation: *Molecular Systems Biology* 5:323

www.molecularsystemsbiology.com

molecular systems biology

PERSPECTIVE

Shifts in growth strategies reflect tradeoffs in cellular economics

Douwe Molenaar[1,3,4,5,*], Rogier van Berlo[2,4], Dick de Ridder[2,4] and Bas Teusink[1,3,4,5]

[1] Centre for Integrative Bioinformatics (IBIVU), Vrije Universiteit Amsterdam, Amsterdam, The Netherlands,
[2] Information and Communication Theory Group, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands,
[3] Top Institute Food and Nutrition, Wageningen, The Netherlands,
[4] Kluyver Centre for Genomics of Industrial Fermentation, Delft, The Netherlands and
[5] Netherlands Consortium for Systems Biology, Amsterdam, The Netherlands
* Corresponding author. Centre for Integrative Bioinformatics (IBIVU), Vrije Universiteit Amsterdam, De Boelelaan 1085, Amsterdam 1081 HV, The Netherlands. Tel.: +31 205 987738; Fax: +31 205 987229; E-mail: douwe.molenaar@falw.vu.nl

Received 23.4.09; accepted 5.10.09

The growth rate-dependent regulation of cell size, ribosomal content, and metabolic efficiency follows a common pattern in unicellular organisms: with increasing growth rates, cell size and ribosomal content increase and a shift to energetically inefficient metabolism takes place. The latter two phenomena are also observed in fast growing tumour cells and cell lines. These patterns suggest a fundamental principle of design. In biology such designs can often be understood as the result of the optimization of fitness. Here we show that in basic models of self-replicating systems these patterns are the consequence of maximizing the growth rate. Whereas most models of cellular growth consider a part of physiology, for instance only metabolism, the approach presented here integrates several subsystems to a complete self-replicating system. Such models can yield fundamentally different optimal strategies. In particular, it is shown how the shift in metabolic efficiency originates from a tradeoff between investments in enzyme synthesis and metabolic yields for alternative catabolic pathways. The models elucidate how the optimization of growth by natural selection shapes growth strategies.
Molecular Systems Biology 5: 323; published online 3 November 2009; doi:10.1038/msb.2009.82
Subject Categories: metabolic and regulatory networks; cellular metabolism
Keywords: growth; metabolic efficiency; overflow metabolism; ribosome content; Warburg effect

Introduction

Growth is a fundamental property of life and, as we will reason below, a property that is highly optimized in unicellular organisms. Given this goal and given the knowledge of physical and chemical limitations of living matter, it should be possible to understand the design of certain growth strategies in terms of growth rate maximization. Indeed, the idea that maximization of growth is a design objective has been successfully applied to predict experimentally observed metabolic strategies from metabolic models. Nevertheless, as we will see, there are limitations to these approaches since they fail to predict a number of commonly observed metabolic strategies, or they do so only under specific assumptions. We believe the cause for this failure to be the fact that these models consider only a subsystem, namely metabolism, and ignore important aspects such as costs related to synthesis of proteins and structural components. In this paper we will elaborate the hypothesis that growth strategies are the result of tradeoffs in the economy of the cell, in which growth rate maximization of the entire system is the objective, and where the only limitations are those set by the laws of physics and chemistry.

Growth and fitness in unicellular organisms

In contrast to the situation for multicellular organisms, growth of biomass and reproduction are tightly coupled processes in unicellular organisms, since every cell division yields a new individual. This also applies to cells originating from multicellular organisms that have compromised growth control mechanisms, like tumour cells and cell lines. The rate of growth is an important factor in the long-term reproductive success of such cells. Mutants, which are able to regulate their cellular composition and metabolism in response to environmental conditions in a way that increases their growth rate under the given conditions, will outnumber their siblings. This is clearly the case under conditions of constant nutrient supply, but numerical dominance also provides an important fitness advantage under fluctuating nutrient supply, because it improves the odds for the mutant genotype to survive starvation and colonize new resources.

Clearly, an organism cannot be optimally adapted to all possible environmental conditions. This is the reason why mutants with higher growth rates than the parental strain can be obtained in the laboratory using a chemostat or by repeated dilution in batch cultures (Helling *et al*, 1987; Mikkola and Kurland, 1992; Vasi *et al*, 1994; Ibarra *et al*, 2002; Dekel and Alon, 2005). However, the fact that these mutants are obtained within a few hundred to thousand generations, a brief period on an evolutionary scale, demonstrates that the selective pressure on growth rate is high under these circumstances.

Molecular Systems Biology 5; Article number 321; doi:10.1038/msb.2009.80
Citation: *Molecular Systems Biology* 5:321

www.molecularsystemsbiology.com

molecular systems biology

Edgetic perturbation models of human inherited disorders

Quan Zhong[1,2,4], Nicolas Simonis[1,2,4], Qian-Ru Li[1,2,4], Benoit Charloteaux[1,2,3,4], Fabien Heuze[1,2,3,4], Niels Klitgord[1,2,4], Stanley Tam[1,2], Haiyuan Yu[1,2], Kavitha Venkatesan[1,2], Danny Mou[1,2], Venus Swearingen[1,2], Muhammed A Yildirim[1,2], Han Yan[1,2], Amélie Dricot[1,2], David Szeto[1,2], Chenwei Lin[1,2], Tong Hao[1,2], Changyu Fan[1,2], Stuart Milstein[1,2], Denis Dupuy[1,2], Robert Brasseur[3], David E Hill[1,2], Michael E Cusick[1,2] and Marc Vidal[1,2,*]

[1] Center for Cancer Systems Biology (CCSB) and Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, USA, [2] Department of Genetics, Harvard Medical School, Boston, MA, USA and [3] Centre de Biophysique Moléculaire Numérique, Faculté Universitaire des Sciences Agronomiques de Gembloux, Gembloux, Wallonia, Belgium
[4] These authors contributed equally to this work
* Corresponding author. Center for Cancer Systems Biology, Department of Cancer Biology, 44 Binney Street, Smith 858, Boston, MA 02115, USA. Tel.: +1 617 632 5180; Fax: +1 617 632 5739; E-mail: marc_vidal@dfci.harvard.edu

Received 5.8.09; accepted 2.10.09

Cellular functions are mediated through complex systems of macromolecules and metabolites linked through biochemical and physical interactions, represented in interactome models as 'nodes' and 'edges', respectively. Better understanding of genotype-to-phenotype relationships in human disease will require modeling of how disease-causing mutations affect systems or interactome properties. Here we investigate how perturbations of interactome networks may differ between complete loss of gene products ('node removal') and interaction-specific or edge-specific ('edgetic') alterations. Global computational analyses of ~50 000 known causative mutations in human Mendelian disorders revealed clear separations of mutations probably corresponding to those of node removal versus edgetic perturbations. Experimental characterization of mutant alleles in various disorders identified diverse edgetic interaction profiles of mutant proteins, which correlated with distinct structural properties of disease proteins and disease mechanisms. Edgetic perturbations seem to confer distinct functional consequences from node removal because a large fraction of cases in which a single gene is linked to multiple disorders can be modeled by distinguishing edgetic network perturbations. Edgetic network perturbation models might improve both the understanding of dissemination of disease alleles in human populations and the development of molecular therapeutic strategies.
Molecular Systems Biology 5: 321; published online 3 November 2009; doi:10.1038/msb.2009.80
Subject Categories: molecular biology of disease
Keywords: binary protein interaction; genotype-to-phenotype relationships; human Mendelian disorders; network perturbation

Introduction

Decades of research into human Mendelian disorders has led to the discovery of a massive amount of disease-associated allelic variations. Most disease-causing mutations are thought to confer radical changes to proteins (Wang and Moult, 2001; Botstein and Risch, 2003; Yue *et al*, 2005; Subramanian and Kumar, 2006). Consequently, genotype-to-phenotype relationships in human genetic disorders are often modeled as: 'mutation in gene *X* leads to loss of gene product X, which leads to disease A'. A single 'gene-loss' model seems pertinent for many diseases (Botstein and Risch, 2003). However, this model cannot fully reconcile with the increasingly appreciated prevalence of complex genotype-to-phenotype associations for even 'simple' Mendelian disorders (Goh *et al*, 2007), particularly in which: (i) a single gene can be associated with multiple disorders (allelic heterogeneity), (ii) a single disorder can be caused by mutations in any one of several genes (locus heterogeneity), (iii) only a subset of individuals carrying a mutation are affected by the disease (incomplete penetrance), or (iv) not all individuals with a given mutation are affected equally (variable expressivity). More complex models to

LETTERS

Imaging chromophores with undetectable fluorescence by stimulated emission microscopy

Wei Min[1]*, Sijia Lu[1]*, Shasha Chong[1], Rahul Roy[1], Gary R. Holtom[1] & X. Sunney Xie[1]

Fluorescence, that is, spontaneous emission, is generally more sensitive than absorption measurement, and is widely used in optical imaging[1,2]. However, many chromophores, such as haemoglobin and cytochromes, absorb but have undetectable fluorescence because the spontaneous emission is dominated by their fast non-radiative decay[3]. Yet the detection of their absorption is difficult under a microscope. Here we use stimulated emission, which competes effectively with the nonradiative decay, to make the chromophores detectable, and report a new contrast mechanism for optical microscopy. In a pump–probe experiment, on photoexcitation by a pump pulse, the sample is stimulated down to the ground state by a time-delayed probe pulse, the intensity of which is concurrently increased. We extract the miniscule intensity increase with shot-noise-limited sensitivity by using a lock-in amplifier and intensity modulation of the pump beam at a high megahertz frequency. The signal is generated only at the laser foci owing to the nonlinear dependence on the input intensities, providing intrinsic three-dimensional optical sectioning capability. In contrast, conventional one-beam absorption measurement exhibits low sensitivity, lack of three-dimensional sectioning capability, and complication by linear scattering of heterogeneous samples. We demonstrate a variety of applications of stimulated emission microscopy, such as visualizing chromoproteins, non-fluorescent variants of the green fluorescent protein, monitoring *lacZ* gene expression with a chromogenic reporter, mapping transdermal drug distributions without histological sectioning, and label-free microvascular imaging based on endogenous contrast of haemoglobin. For all these applications, sensitivity is orders of magnitude higher than for spontaneous emission or absorption contrast, permitting non-fluorescent reporters for molecular imaging.

The phenomenon of stimulated emission was first described by Einstein in 1917 (ref. 4). An atom or molecule in its excited state can be stimulated down to the ground state by an incident light field, resulting in the creation of a new coherent photon identical to those in the original incident field. This process only occurs when the frequency of the incident field matches the transition energy. Stimulated emission was later used as a fundamental principle for light amplification in the laser[5]. The depopulation aspect of stimulated emission has been used for population dumping from excited states[6], super-resolution fluorescence microscopy[7], and fluorescence lifetime imaging[8]. Here we use the light-amplification aspect of stimulated emission as a contrast mechanism for highly sensitive imaging of chromophores that have undetectable fluorescence.

Such chromophores have very short-lived excited states with much faster non-radiative decay rates than their spontaneous emission rates. As a result, their feeble fluorescence is overwhelmed by backgrounds, such as stray light, solvent Raman scattering, and detector dark counts. Our solution to this problem is to conduct a dual-beam experiment to interrogate the short-lived excited state by stimulated emission, which can compete with the non-radiative decay under a strong enough stimulating field (Fig. 1a). The resulting 'amplification' of the stimulation beam can then be detected in the presence of the background signals.

Considering the optical excitation at frequency ω_{01} (Fig. 1a), the absorption cross-section $\sigma_{abs[0\rightarrow1]}$ is about 10^{-16} cm^2 for a single chromophore at room temperature[2,9]. As shown in Fig. 1b, under a tightly focused laser beam with a beam waist area of S ($\sim10^{-9}$ cm^2), the integrated intensity attenuation of the excitation beam $\Delta I_E/I_E$ is proportional to the ratio between $\sigma_{abs[0\rightarrow1]}$ and S:

$$\Delta I_E/I_E = -N_0\sigma_{abs[0\rightarrow1]}/S \quad (1)$$

where N_0 is the number of molecules in the ground state. For a single chromophore, $\Delta I_E/I_E$ is of the order of 10^{-7}. Such small attenuation cannot be detected by conventional absorption microscopy. We note that single-molecule absorption was previously achieved in cryogenic temperatures using a frequency-modulation method[10], which is, however, difficult to implement at room temperature because of the broad molecular absorption linewidth. Moreover, absorption measurement is often complicated by scattering when imaging biological specimens. Instead of detecting direct absorption, here we detect the stimulated emission followed by the excitation of the molecule.

According to Einstein[4], the molecular cross-section σ_{stim} for stimulated emission is comparable to σ_{abs}, because of microscopic reversibility. Unlike the absorption that results in attenuation, the stimulation beam will experience an intensity gain after interacting with the molecules:

$$\Delta I_S/I_S = +N_2\sigma_{stim[2\rightarrow3]}/S \quad (2)$$

where N_2 is the number of excited molecules transiently probed by the stimulation pulses (Fig. 1b). For a single chromophore, $\Delta I_S/I_S$ is also $\sim10^{-7}$. Without special techniques, such a small signal would again be buried in the noise ($\sim$1%) of the stimulation beam.

To overcome this noise problem in detecting stimulated emission, we implemented a high-frequency (>1 MHz) phase-sensitive detection technique. In so doing, the laser intensity fluctuation, which occurs primarily at low frequency (kilohertz to direct current), can be circumvented, as has been previously applied in other spectroscopic[11] and recently stimulated Raman scattering microscopy[12] and two-photon absorption microscopy[13]. In the scheme shown in Fig. 1b and c, the intensity of the excitation beam is modulated at 5 MHz, and this creates a modulation of the stimulated emission signal at the same frequency, because only when the excitation beam is present can the gain of the stimulation beam occur. Such an induced modulation signal can then be sensitively extracted by a lock-in amplifier referenced

[1]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA.
*These authors contributed equally to this work.

LETTERS

Direct RNA sequencing

Fatih Ozsolak[1], Adam R. Platt[1], Dan R. Jones[1], Jeffrey G. Reifenberger[1], Lauryn E. Sass[1], Peter McInerney[1], John F. Thompson[1], Jayson Bowers[1], Mirna Jarosz[1] & Patrice M. Milos[1]

Our understanding of human biology and disease is ultimately dependent on a complete understanding of the genome and its functions. The recent application of microarray and sequencing technologies to transcriptomics has changed the simplistic view of transcriptomes to a more complicated view of genome-wide transcription where a large fraction of transcripts emanates from unannotated parts of genomes[1–7], and underlined our limited knowledge of the dynamic state of transcription. Most of this broad body of knowledge was obtained indirectly because current transcriptome analysis methods typically require RNA to be converted to complementary DNA (cDNA) before measurements, even though the cDNA synthesis step introduces multiple biases and artefacts that interfere with both the proper characterization and quantification of transcripts[8–18]. Furthermore, cDNA synthesis is not particularly suitable for the analysis of short, degraded and/or small quantity RNA samples. Here we report direct single molecule RNA sequencing without prior conversion of RNA to cDNA. We applied this technology to sequence femtomole quantities of poly(A)$^+$ *Saccharomyces cerevisiae* RNA using a surface coated with poly(dT) oligonucleotides to capture the RNAs at their natural poly(A) tails and initiate sequencing by synthesis. We observed transcript 3′ end heterogeneity and polyadenylated small nucleolar RNAs. This study provides a path to high-throughput and low-cost direct RNA sequencing and achieving the ultimate goal of a comprehensive and bias-free understanding of transcriptomes.

cDNA-based transcriptome analysis approaches being used today exhibit several shortcomings that prevent us from understanding the real nature of transcriptomes and ultimately genome biology. Some of these limitations are: (1) the tendency of various reverse transcriptases (RT) to generate spurious second-strand cDNA due to their DNA-dependent DNA polymerase activities[9,10,18]; (2) the generation of artefactual cDNAs due to template switching[8,13,16,17] or contaminating DNA and primer-independent cDNA synthesis[11,12]; and (3) the error-prone[15,19] and inefficient nature of RTs yielding low quantities of cDNA. Furthermore, most RNA analysis technologies require the synthesis of not just the first strand cDNA but also a second strand cDNA that are both subjected to further ligation/amplification steps, introducing yet more biases. These limitations pose problems for the determination of RNA strandedness[14,20], the identification of chimaeric transcripts, quantification of RNA species, and the analysis of low quantity (<1 nanogram) or short RNA species, such as those obtained from formalin-fixed, paraffin-embedded tissue samples. Because almost all transcript analysis technologies in use today suffer from the limitations briefly summarized above, there is an ever-growing

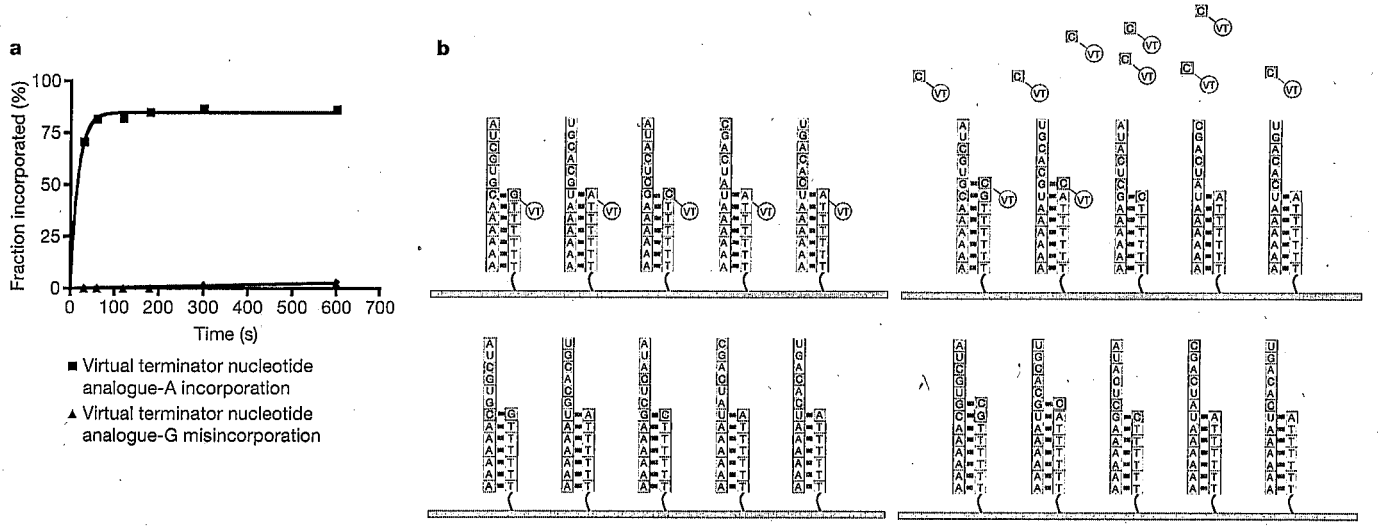

Figure 1 | DRS chemistry and sequencing steps. a, Under optimized conditions, polymerase exhibits fast correct nucleotide incorporation (VT-A) and slow misincorporation (VT-G) kinetics. **b**, DRS procedure. Top left: polyadenylated and 3′-blocked RNA is captured on surfaces coated with dT(50) oligonucleotide. A 'fill' step is performed with natural dTTP, and a 'lock' step with fluorescently labelled VT-A, -C and -G nucleotides. These steps correct for any misalignments that may be present in poly(A/T) duplexes, and ensure that the sequencing starts in the template rather than the poly(A) tail. Imaging is performed to locate the template positions. Bottom left: chemical cleavage of the dye-nucleotide linker is performed to prepare the templates for nucleotide incorporation. Top right: incubation with one VT nucleotide and polymerase is performed, followed by imaging to locate the templates that incorporated the nucleotide. Bottom right: chemical cleavage of the dye allows the surface and RNA templates to be ready for the next nucleotide addition cycle.

[1]Helicos BioSciences Corporation, One Kendall Square, Cambridge, Massachusetts 02139, USA.

ARTICLES

Genome evolution and adaptation in a long-term experiment with *Escherichia coli*

Jeffrey E. Barrick[1]*, Dong Su Yu[2,3]*, Sung Ho Yoon[2], Haeyoung Jeong[2], Tae Kwang Oh[2,4], Dominique Schneider[5], Richard E. Lenski[1] & Jihyun F. Kim[2,6]

The relationship between rates of genomic evolution and organismal adaptation remains uncertain, despite considerable interest. The feasibility of obtaining genome sequences from experimentally evolving populations offers the opportunity to investigate this relationship with new precision. Here we sequence genomes sampled through 40,000 generations from a laboratory population of *Escherichia coli*. Although adaptation decelerated sharply, genomic evolution was nearly constant for 20,000 generations. Such clock-like regularity is usually viewed as the signature of neutral evolution, but several lines of evidence indicate that almost all of these mutations were beneficial. This same population later evolved an elevated mutation rate and accumulated hundreds of additional mutations dominated by a neutral signature. Thus, the coupling between genomic and adaptive evolution is complex and can be counterintuitive even in a constant environment. In particular, beneficial substitutions were surprisingly uniform over time, whereas neutral substitutions were highly variable.

Adaptation has often been viewed as a gradual process. Darwin[1] wrote that "We see nothing of these slow changes in progress, until the hand of time has marked the long lapse of ages…". Theoretical work in quantitative genetics supported this view by showing that gradual adaptation would result from constant selection on many mutations of small effect[2]. However, an alternative model of evolution on rugged fitness landscapes challenged this perspective[3] and, later, empirical evidence was found for alternating periods of rapid phenotypic evolution and stasis in some lineages[4,5]. The causes of variation in the rate of adaptation remain controversial and are probably diverse. They may include changes in the environment, in circumstances promoting or impeding gene flow, and in opportunities for refinement following the origin of key innovations or the invasion of new habitats, among other factors[6–11].

Genomic changes underlie evolutionary adaptation, but mutations—even those substituted (fixed) in evolving populations—are not necessarily beneficial. Variation in the rate of genomic evolution is also subject to many influences and complications. On the one hand, theory predicts that neutral mutations should accumulate by drift at a uniform rate, albeit stochastically, provided the mutation rate is constant[12]. On the other hand, rates of substitution of beneficial and deleterious mutations depend on selection, and hence the environment, as well as on population size and structure[13,14]. Moreover, the relative proportions of substitutions that are neutral, deleterious and beneficial are usually difficult to infer given imperfect knowledge of any organism's genetics and ecology, in the past as well as in the present.

Experiments with tractable model organisms evolving in controlled laboratory environments minimize many of these complications and uncertainties[15,16]. Moreover, new methods have made it feasible to sequence complete genomes from evolution experiments with bacteria[17–20]. To date, such analyses have focused on finding the mutations responsible for particular adaptations. However, the application of comparative genome sequencing to experimental evolution studies also offers the opportunity to address major conceptual issues, including whether the dynamics of genomic and adaptive evolution are coupled very tightly or only loosely[10,12,13,21,22].

Genome dynamics and adaptation

To examine the tempo and mode of genomic evolution, we sequenced the genomes of *E. coli* clones sampled at generations 2,000, 5,000, 10,000, 15,000, 20,000 and 40,000 from an asexual population that evolved with glucose as a limiting nutrient for almost 20 years as part of a long-term experiment. The complete sequence of the ancestral strain served as a reference for identifying mutations in the evolved clones, which we refer to by their generation abbreviations 2K, 5K, 10K, 15K, 20K and 40K.

Figure 1 shows all mutations identified in the evolved clones through 20,000 generations. The 45 mutations in the 20K clone include 29 single-nucleotide polymorphisms (SNPs) and 16 deletions, insertions and other polymorphisms (DIPs). Figure 2 shows that the number of mutational differences between the ancestral and evolved genomes accumulated in a near-linear fashion over this period. Any deviation from linearity was not statistically significant based on randomization tests.

The near-linearity of the trajectory for genomic evolution is rather surprising, given that such constancy is widely taken as a signature of neutral evolution[12], whereas the fitness trajectory for this population[23] shows profound adaptation that is strongly nonlinear. In particular, the rate of fitness improvement decelerates over time (Fig. 2), which indicates that the rate of appearance of new beneficial mutations is declining, their average benefit is becoming smaller, or both. These effects, in turn, should cause the rate of genomic evolution to decelerate.

To understand this point, consider a simple model of the substitution of beneficial mutations in a clonal population of haploid organisms. A beneficial mutation has an initial frequency of $1/N$,

[1]Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan 48824, USA. [2]Industrial Biotechnology and Bioenergy Research Center, Korea Research Institute of Bioscience and Biotechnology, Yuseong, Daejeon 305-806, Korea. [3]Department of Computer Science and Engineering, Chungnam National University, Yuseong, Daejeon 305-764, Korea. [4]21C Frontier Microbial Genomics and Applications Center, Yuseong, Daejeon 305-806, Korea. [5]Institut Jean Roget, Laboratoire Adaptation et Pathogénie des Microorganismes, CNRS UMR 5163, Université Joseph Fourier, Grenoble 1, BP 170, F-38042 Grenoble cedex 9, France. [6]Functional Genomics Program, School of Science, University of Science and Technology, Yuseong, Daejeon 305-333, Korea.
*These authors contributed equally to this work.

LETTERS

Population context determines cell-to-cell variability in endocytosis and virus infection

Berend Snijder[1,2], Raphael Sacher[1,2], Pauli Rämö[1], Eva-Maria Damm[1], Prisca Liberali[1] & Lucas Pelkmans[1]

Single-cell heterogeneity in cell populations arises from a combination of intrinsic and extrinsic factors[1–3]. This heterogeneity has been measured for gene transcription, phosphorylation, cell morphology and drug perturbations, and used to explain various aspects of cellular physiology[4–6]. In all cases, however, the causes of heterogeneity were not studied. Here we analyse, for the first time, the heterogeneous patterns of related cellular activities, namely virus infection, endocytosis and membrane lipid composition in adherent human cells. We reveal correlations with specific cellular states that are defined by the population context of a cell, and we derive probabilistic models that can explain and predict most cellular heterogeneity of these activities, solely on the basis of each cell's population context. We find that accounting for population-determined heterogeneity is essential for interpreting differences between the activity levels of cell populations. Finally, we reveal that synergy between two molecular components, focal adhesion kinase and the sphingolipid GM1, enhances the population-determined pattern of simian virus 40 (SV40) infection. Our findings provide an explanation for the origin of heterogeneity patterns of cellular activities in adherent cell populations.

Virus infection can be considered the end result of the concerted action of many cellular processes, including endocytosis and the establishment of membrane lipid composition[7,8]. It is well accepted that these activities display heterogeneity within a population of cells, but the underlying causes of this heterogeneity are not known. We can therefore not predict how such activities behave in a population of cells[9].

To study the heterogeneity of these activities, we examined large populations of single monoclonal adherent human cells after three days of growth. These were infected by rotavirus, dengue virus, mouse hepatitis virus (MHV) or SV40. MHV uses clathrin-mediated endocytosis (CME) for infectious entry into the cell[10], and SV40 binds to the sphingolipid GM1 for host cell attachment and entry[11,12]. To analyse CME, we allowed cells to internalize fluorescent transferrin[13]. To determine the amount of GM1, which is enriched in lipid rafts on the cell surface[14], we exposed cells to fluorescent cholera toxin B[15].

We observed that the cell-to-cell variation or heterogeneity of these activities displays specific patterns in both cancer- (HeLa and A431) (Fig. 1A) and non-cancer-derived (MCF10A) cells (see below). The patterns could not have been the result of physical constraints on the accessibility of virus particles or fluorescent probes to particular cells, because different types of pattern were observed.

A potential source of single-cell heterogeneity is the deterministic interplay between the phenotypic state of a cell and its activities[2,16], mediated by sensors and signal transduction networks. Growing isogenic adherent mammalian cells create cell islets and regions that are sparsely or densely populated, to which cells adapt their size, shape and rate of proliferation[17,18] (Supplementary Movie 1), and display several discrete subpopulations[19]. Together, this implies that patterns of varying cellular states can arise from the adaptation of individual human cells to the particular 'population context' in which they reside. The coupling between cell size and timing of meiosis in yeast[20], between cell size and the determination of phage infection outcome in *Escherichia coli*[21], and the self-assembly of prokaryotic cells into complex colony patterns[22], indicate that such mechanisms also operate in unicellular organisms. We therefore hypothesized that the heterogeneity patterns that we observed are caused by regulatory mechanisms that couple the phenotypic state of a cell to intracellular processes on which virus infection depends, such as endocytosis and the lipid composition of the cell surface.

To reveal the presence of such mechanisms, we used three data-driven modelling approaches on several single-cell measurements obtained by computerized image analysis and supervised machine learning from multiple large populations of cells. (See Fig. 1B for an overview of the methods used, and Supplementary Methods for detailed information.) We first measured for each cell the size of the population to which it belonged, its local cell density, its position on a cell islet edge, its cell size, its mitotic state and its apoptotic state (microenvironment and cell state). By delineating how these parameters interact with each other, we quantified how the characteristics of a cell population are determined. This showed that each parameter, in particular population size, local cell density, position on a cell islet edge, and cell size, represents a population-determined property of an individual cell. The same interactions were observed in cancer cells (Supplementary Fig. 1), normal diploid cells (see below) and primary cells (not shown), despite their different morphologies and growth rates.

For each cell, we next determined whether it was infected, or we quantified its CME activity or cell-surface level of GM1 (cellular activities). We then used the population-determined properties as predictors to model the probability of infection, the activity of CME and the amount of GM1 on the cell surface. In each instance, models could be derived that had good fits (Fig. 1C and Supplementary Fig. 2). The model parameters demonstrated the existence of extensive and specific regulatory mechanisms between population context and virus infection, endocytosis and membrane lipid composition (Fig. 1D). Importantly, the models were able to predict accurately the heterogeneity patterns of cellular activities, solely based on a quantitative assessment of the population context and the state of individual cells (Fig. 1E).

An interesting question that arises is to what extent these population-determined effects contribute to the total variation observed in a population. Strikingly, variance analysis (Supplementary Methods) demonstrated that variation determined by population context is a major component of the total variation observed. For rotavirus, SV40 and dengue virus infection, our measures of the local environment

[1]Institute of Molecular Systems Biology, ETH Zurich (Swiss Federal Institute of Technology), Wolfgang Pauli-Strasse 16, CH-8093 Zurich, Switzerland. [2]Zurich PhD Program in Molecular Life Sciences, Zurich, Switzerland.

Vol 461|15 October 2009|doi:10.1038/nature08472

LETTERS

Cheater-resistance is not futile

Anupama Khare[1], Lorenzo A. Santorelli[1,2], Joan E. Strassmann[2], David C. Queller[2], Adam Kuspa[1,2,3] & Gad Shaulsky[1,2]

Cooperative social systems are susceptible to cheating by individuals that reap the benefits of cooperation without incurring the costs[1]. There are various theoretical mechanisms for the repression of cheating[2] and many have been tested experimentally. One possibility that has not been tested rigorously is the evolution of mutations that confer resistance to cheating. Here we show that the presence of a cheater in a population of randomly mutated social amoebae can select for cheater-resistance. Furthermore, we show that this cheater-resistance can be a noble strategy because the resister strain does not necessarily exploit other strains. Thus, the evolution of resisters may be instrumental in preserving cooperative behaviour in the face of cheating.

Dictyostelium cells propagate as unicellular amoebae in the soil. Upon starvation, they aggregate into multicellular structures and differentiate into viable spores and dead stalk cells[3]. Stalk-cell differentiation supports spore maturation and dispersal, but this altruistic behaviour can be exploited by cheaters that make more than their fair share of spores in chimaeric fruiting bodies[4]. The genetic potential for cheating is high[5] and cheaters abound in nature[4], but cheating behaviour can be restrained by various mechanisms, such as intrinsic lower fitness of the cheater[6], pleiotropy of the cheater gene[7], high genetic relatedness in natural populations[8], and kin discrimination[9,10].

Evolution of cheating-resistance is another mechanism that could restrict the spread of a specific cheater allele in the population and it could be manifested in several ways. One way is the evolution of other cheaters, but such a mechanism could lead to an arms-race of cheating strategies that would contribute to the rapid demise of cooperation. We therefore tested whether selection for cheating-resistance in *Dictyostelium* could yield mutants that resist cheating while remaining cooperative. Our strategy was to mix a population of mutated cells derived from the wild-type AX4 with a cheater strain, allow them to grow and develop into chimaeric fruiting bodies, and select spores. Under these conditions, the cheater would exploit most of the cells in the mutant population, thus increasing the proportion of any cheater-resistant mutants. We chose a strong cheater mutant, LAS5 (ref. 5) (which we renamed cheater C, *chtC*) as the selector. This mutant strain has a plasmid insertion in the *DDB_G0290959* gene, which is predicted to encode a protein with a signal peptide anchor and a transmembrane domain at the amino terminus. We generated a population of 1,000 strains in the wild-type background, each containing one insertional mutation that also conferred resistance to the antibiotic blasticidin S (BSR). We mixed *chtC* (which is sensitive to blasticidin S) with the mutant population at a ratio of 4:1 and allowed the cells to develop into chimaeric fruiting bodies. We germinated the spores and allowed the amoebae to grow in the presence of blasticidin S, thus eliminating the *chtC* cells. We then mixed the enriched mutant population with fresh *chtC* cells and developed them again, to maintain similar levels of selection in six such cycles of selection (Fig. 1a). We predicted that most of the mutant strains in the population would be cheated upon by *chtC*, eventually leading to their disappearance, and to the enrichment of any cheater-resistant strains that were present in the pool. We carried out this selection on several independent pools of mutants.

To test the population dynamics during the screen, we assessed population complexity of the mutant pool by Southern blot analysis of total genomic DNA, with a probe against the BSR marker. The initial population contained numerous strains, each giving a different banding pattern, which resulted in a smeared signal on the Southern blot (Fig. 1b). As the selection proceeded, the levels of various mutants fluctuated, as shown by the appearance and disappearance of bands on the blot. However, by the fourth cycle of selection, the population became enriched predominantly with a single mutant, suggesting that the presence of the cheater selected for a specific strain, which we predicted to be cheater-resistant. We isolated the enriched mutant and found an insertion in *DDB_G0271758*, a gene that has not been annotated in *Dictyostelium discoideum* and has no annotated homologues in other organisms. We then interrogated the above Southern blot with a probe against *DDB_G0271758* and found that the wild-type allele disappeared from the population, whereas the mutant allele became abundant as the selection progressed (Fig. 1c). We also used quantitative PCR (qPCR) to quantify the progressive enrichment of the mutant allele during the selective process and observed a 100-fold increase in the allele abundance by the sixth selection cycle (Fig. 1d).

To verify that the insertion in *DDB_G0271758* caused the observed phenotype, we regenerated the insertion in fresh wild-type cells, and named the resulting mutant and the mutated gene *rccA* (resister of *chtC* A). Both growth and development of the *rccA* mutant were indistinguishable from the wild-type AX4, suggesting that the *rccA* mutation did not confer obvious fitness costs. We then carried out direct cheating assays. We mixed either wild-type or *rccA* mutant cells at a 1:1 ratio with *chtC* and determined the proportion of spores formed by each strain. When mixed with *chtC* cells, the *rccA* mutant formed almost 50% of the spores. This proportion was significantly higher than the number of spores formed by AX4 in a mix with *chtC* cells (Fig. 2a), but not significantly lower than the hypothesized value of 50% (one-sample one-tailed *t*-test, $P = 0.28$). This observation supports the hypothesis that the *rccA* mutant can resist cheating by *chtC*. We also tested whether *rccA* was a specific resister of *chtC* or a resister of other cheaters as well, by performing similar mixes with LAS1, another facultative cheater that was isolated from the screen that yielded *chtC*[5]. We found that *rccA* did not resist cheating by LAS1 (Fig. 2a), indicating that *rccA* is unable to resist all cheaters. Therefore, it is likely that the ability to resist cheating will depend on the cheating mechanism of the specific cheater.

Cheating-resistance might be due to counter-cheating, such that the presence of a specific cheater would select for equivalent or even stronger cheaters. An alternative would be that cheater-resisters would be neutral, or noble, so they would not cheat on strains victimized by the original cheater. To distinguish between these

[1]Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA. [2]Department of Ecology and Evolutionary Biology, Rice University, Houston, Texas 77005, USA. [3]Department of Biochemistry and Molecular Biology, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA.

ARTICLES

Co-translational mRNA decay in *Saccharomyces cerevisiae*

Wenqian Hu[1]*, Thomas J. Sweet[1]*, Sangpen Chamnongpol[2], Kristian E. Baker[1] & Jeff Coller[1]

The rates of RNA decay and transcription determine the steady-state levels of all messenger RNA and both can be subject to regulation. Although the details of transcriptional regulation are becoming increasingly understood, the mechanism(s) controlling mRNA decay remain unclear. In yeast, a major pathway of mRNA decay begins with deadenylation followed by decapping and 5′–3′ exonuclease digestion. Importantly, it is hypothesized that ribosomes must be removed from mRNA before transcripts are destroyed. Contrary to this prediction, here we show that decay takes place while mRNAs are associated with actively translating ribosomes. The data indicate that dissociation of ribosomes from mRNA is not a prerequisite for decay and we suggest that the 5′–3′ polarity of mRNA degradation has evolved to ensure that the last translocating ribosome can complete translation.

In eukaryotic cells, mRNA is predominately degraded by two alternative pathways that are both initiated by shortening of the 3′ polyadenosine tail (deadenylation). After deadenylation, either the 5′ 7mGpppN cap is removed (decapping) and the message is digested exonucleolytically 5′-3′ or the transcript is destroyed 3′-5′ by the cytoplasmic exosome[1]. The two mechanisms of mRNA decay together determine basal mRNA levels, thereby significantly contributing to overall gene expression.

Translation is postulated to be a key determinant in controlling mRNA decapping[1]. The translational initiation complex eIF-4F occupies the cap during translation, which suggests that its binding must be antagonized and translational repression must ensue before decapping can occur[1–4]. This hypothesis is supported by several observations. First, translational initiation rate is inversely proportional to decapping rate[3]. Second, the decapping regulators Dhh1p and Pat1p are translational repressors and their role in promoting mRNA decapping is partly a function of this activity[5,6]. Third, mRNA decapping can occur at an unquantified level in ribosome-free cellular foci, termed P-bodies[2]. Collectively, a two-step model for mRNA decay has been proposed where ribosome dissociation is a necessary first step before mRNA decapping[1–4].

Deadenylated mRNA remains on polyribosomes

The aforementioned model for mRNA decay predicts that after deadenylation but before decapping a ribosome-free state exists[1–4]. We reasoned that in a decapping-defective cell (*dcp2Δ*), deadenylated RNA would accumulate in this ribosome-free state. We used sucrose density gradients to survey mRNA ribosome association in wild-type and decapping-defective cells (*dcp2Δ*). Greater than 90% of total cellular mRNA is analysed by this method (data not shown), and ribosome-free ribonucleoprotein (RNP) structures can be clearly separated from polyribosomes (Supplementary Fig. 2c). As predicted, inhibition of decapping did result in accumulation of deadenylated mRNA (Supplementary Fig. 2a, b, f); however, the mRNAs continued to sediment deep into a sucrose gradient even when deadenylated (Supplementary Fig. 2d, g, h). In fact, the sedimentation profiles of several mRNAs in *dcp2Δ* cells were indistinguishable from those in wild-type cells (Supplementary Fig. 2d, g, h). The rapid sedimentation of these RNAs could occur either because they were sequestered in heavy particles (perhaps P-bodies)[1,2] or because they were associated with ribosomes. The fact that sedimentation correlated with the length of the open reading frame (ORF) (Supplementary Fig. 2d, g, h) strongly suggested that the mRNAs were ribosome associated (see below).

Decapped mRNAs are found on polyribosomes

Because deadenylated mRNAs are the substrates for decapping[3] we also assessed the sedimentation profiles of decapped RNAs. This was done in cells defective for the 5′–3′ exonuclease (*xrn1Δ*). In these cells a stable decapped decay intermediate shortened by two nucleotides accumulates (indicated by '−cap'; Fig. 1a) and can be detected by using quantitative primer extension analysis (Supplementary Fig. 10)[7–9]. Interestingly, the decapped intermediate showed the same sedimentation profile as the deadenylated RNA (Fig. 1a versus Supplementary Fig. 2); most (83–95%) decapped mRNA being present in polyribosomes (Fig. 1a, d). To determine whether the decay intermediate was associated with ribosomes, we took four approaches. First, introduction of a premature termination codon that shortened the ORF of *PGK1* by 393 codons resulted in a dramatic shift to significantly lighter fractions (Fig. 1b, c). Second, introduction of a stem-loop to limit translation[5] caused a shift towards the top of the gradient both for capped and uncapped mRNAs (Fig. 1c). Third, treatment with EDTA (known to dissociate ribosomes) shifted the sedimentation to the top of the gradient (Fig. 1c). Finally, we showed that decapped mRNAs were associated with ribosomes by ribosome immunoprecipitation[10] (Supplementary Fig. 3).

To investigate ribosome-associated decapping further and to exclude the possibility that decapping had occurred before initiation of protein synthesis, we took a transcriptional-pulse chase approach using the *PGK1* mRNA reporter[9]. Using a circularization-based PCR with reverse transcription (cRT–PCR)[11] analysis, we noted that decapped RNA started to appear around 60 min after initiation of transcription (Fig. 2a–c). Separation of cell lysate into non-translating and polyribosome-associated fractions indicated that when decapping is initiated at 60 min, most decapped mRNA was polyribosome associated (Fig. 2d). To exclude further the possibility that association of

[1]Center for RNA Molecular Biology, Case Western Reserve University, Cleveland, Ohio 44106, USA. [2]Affymetrix Inc., 26111 Miles Road, Cleveland, Ohio 44128, USA.
*These authors contributed equally to this work.

Vol 462|5 November 2009|doi:10.1038/nature08504

LETTERS

Experimental evolution of bet hedging

Hubertus J. E. Beaumont[1,2]†, Jenna Gallie[1], Christian Kost[1]†, Gayle C. Ferguson[1] & Paul B. Rainey[1]

Bet hedging—stochastic switching between phenotypic states[1–3]—is a canonical example of an evolutionary adaptation that facilitates persistence in the face of fluctuating environmental conditions. Although bet hedging is found in organisms ranging from bacteria to humans[4–10], direct evidence for an adaptive origin of this behaviour is lacking[11]. Here we report the *de novo* evolution of bet hedging in experimental bacterial populations. Bacteria were subjected to an environment that continually favoured new phenotypic states. Initially, our regime drove the successive evolution of novel phenotypes by mutation and selection; however, in two (of 12) replicates this trend was broken by the evolution of bet-hedging genotypes that persisted because of rapid stochastic phenotype switching. Genome re-sequencing of one of these switching types revealed nine mutations that distinguished it from the ancestor. The final mutation was both necessary and sufficient for rapid phenotype switching; nonetheless, the evolution of bet hedging was contingent upon earlier mutations that altered the relative fitness effect of the final mutation. These findings capture the adaptive evolution of bet hedging in the simplest of organisms, and suggest that risk-spreading strategies may have been among the earliest evolutionary solutions to life in fluctuating environments.

Life exists in ever-changing environments, but surviving under fluctuating conditions poses challenges. One solution is the evolution of mechanisms that allow modulation of phenotype in response to specific environmental cues. An alternative solution is stochastic phenotype switching, a strategy based on bet hedging, rather than direct environmental sensing[4–10].

The general prediction from theory is that fluctuating selection generated by unpredictable environments can favour the evolution of bet hedging[1–4,12–14]. Under such conditions, a strategy that generates random variation in fitness-related traits among individuals within a population can enhance long-term fitness by increasing the likelihood that a subset of individuals expresses a phenotype that will be adaptive in a future environment[3,15,16]. However, the outcome of adaptive evolution under fluctuating selection is also shaped by factors such as the frequency of environmental change[12], the capacity of a given population to respond to fluctuations by mutation and selection (that is, evolvability), the presence of suitable environmental cues[13] and the cost–benefit balance of different strategies[12].

Here we report the *de novo* evolution of bet hedging in experimental bacterial populations. Our populations experienced repeated bouts of selection in two contrasting environments; they also experienced fluctuating selection wrought by imposition of an exclusion rule and population bottleneck. Applied at the point of transfer between environments, the exclusion rule assigned a fitness of zero to the type that was common in the current environment; imposition of the bottleneck meant that only a single phenotypically distinct type was selected from among the survivors to found the next bout of selection. The exclusion rule imposed strong selection for phenotypic innovation, whereas the bottleneck negated the cost of bet hedging—that is, the generation of types maladapted to the prevailing conditions—by eliminating competition with conspecifics. A natural analogue of this mode of fluctuating selection is imposed by the host immune system on invading microorganisms. Indeed, many pathogens have evolved bet-hedging strategies based on stochastic antigen switching[8].

When the ancestral genotype of *Pseudomonas fluorescens* SBW25 is grown in static broth microcosms, it rapidly diversifies into a range of niche specialist genotypes by mutation and selection, which each form distinct colonies on agar plates[17]. In contrast, diversification is constrained in shaken microcosms, which favour genotypes with an ancestral colony type[18]. We exploited the pleiotropic correlation between niche specialization and colony morphology to realize our selection regime. The expected evolutionary response was repeated evolution, fixation and extinction of genotypes with novel colony morphologies.

Twelve replicate selection lines were founded with the ancestral genotype and subjected to 16 rounds of alternating selection in static and shaken microcosms. During each round, populations were propagated by serial dilution until the emergence of cells that formed colonies with a heritable morphology different from that of their immediate ancestor. Selection for different colonies was open-ended rather than for an a priori defined morphology. Once detected, cells derived from a single individual of this new type were transferred to the opposing environment for the next round of selection (Fig. 1a and Methods). Thus, we imposed selection for a high growth rate in static and shaken microcosms, and simultaneously fluctuating selection for colony innovation. Each lineage was associated with a cognate control line that was under stabilizing selection for the ancestral colony morphology, but otherwise treated identically (see Methods).

This selection regime indeed drove the repeated evolution of new colony morphologies. However, here we concentrate on the evolution of genotypes capable of rapid, stochastic colony-morphology switching (hereafter referred to as colony switching), which emerged in two of 12 replicates ($1B^4$ and $6B^4$, Supplementary Note 1). Each of these switching genotypes formed distinct translucent and opaque colonies (Fig. 1b). Colony switching persisted in both lines for seven additional rounds of selection, after which the experiment was ended. None of the control lines gave rise to colony switching.

To obtain insight into the mechanisms by which switching types persisted, we re-imposed our selection regime on the genotype $1B^4$ (Fig. 1b), but this time ignored the colony variants it generated through colony switching. Without exception, $1B^4$ was driven to extinction by the evolution of new genotypes with novel colony morphologies, some of which did not switch (Supplementary Note 2). This shows that persistence was attributable to bet hedging—that is, the capacity to generate at high-frequency colonies that were phenotypically distinct from those selected in the previous round—rather than to an intrinsic growth rate advantage over non-switching types. This finding also draws attention to the importance of the bottleneck: in its absence,

[1]New Zealand Institute for Advanced Study and Allan Wilson Centre for Molecular Ecology & Evolution, Massey University, Private Bag 102904, North Shore Mail Centre, North Shore City 0745, Auckland, New Zealand. [2]Institute of Biology Leiden, Leiden University, PO Box 9505, 2300 RA Leiden, The Netherlands. †Present addresses: Institute of Biology Leiden, Leiden University, PO Box 9505, 2300 RA Leiden, The Netherlands (H.J.E.B.); Department of Bioorganic Chemistry, Max Planck Institute for Chemical Ecology, 07745 Jena, Germany (C.K.).

nature biotechnology

The challenges of sequencing by synthesis

Carl W Fuller[1], Lyle R Middendorf[2], Steven A Benner[3], George M Church[4], Timothy Harris[5], Xiaohua Huang[6], Stevan B Jovanovich[7], John R Nelson[8], Jeffery A Schloss[9], David C Schwartz[10] & Dmitri V Vezenov[11]

DNA sequencing-by-synthesis (SBS) technology, using a polymerase or ligase enzyme as its core biochemistry, has already been incorporated in several second-generation DNA sequencing systems with significant performance. Notwithstanding the substantial success of these SBS platforms, challenges continue to limit the ability to reduce the cost of sequencing a human genome to $100,000 or less. Achieving dramatically reduced cost with enhanced throughput and quality will require the seamless integration of scientific and technological effort across disciplines within biochemistry, chemistry, physics and engineering. The challenges include sample preparation, surface chemistry, fluorescent labels, optimizing the enzyme-substrate system, optics, instrumentation, understanding tradeoffs of throughput versus accuracy, and read-length/phasing limitations. By framing these challenges in a manner accessible to a broad community of scientists and engineers, we hope to solicit input from the broader research community on means of accelerating the advancement of genome sequencing technology.

Attaining the Human Genome Project goal of sequencing the human genome and rapidly and publicly disseminating the data was a milestone in human biomedical research that was enabled by scientific, technical and cultural innovation. Central to the project's success was the development of robust, automated methods and technologies to identify the linear sequence of nucleotides. Recognizing the opportunities to use dramatically expanded sequencing technology in the subsequent phase of genomics research, in 2004 the National Human Genome Research Institute (NHGRI) of the National Institutes of Health (NIH) initiated a funding program with a goal of reducing the cost of genome sequencing to ~$1,000 in 10 years, with an intermediate goal of $100,000 by the end of 2009 (ref. 1; http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-04-003.html). Numerous grant awards have been made in this program (http://www.genome.gov/10000368), which has stimulated a strong record of publications and patents and the successful commercialization of several second-generation sequencing platforms now in active use worldwide (e.g., https://www.roche-applied-science.com/sis/sequencing/index.jsp, http://www.illumina.com/pages.ilmn?ID=204, https://products.appliedbiosystems.com/ab/en/US/adirect/ab?cmd=catNavigate2&catID=604416, http://www.polonator.org/ and http://www.helicosbio.com/) with others in the wings (http://www.pacificbiosciences.com/, http://visigenbio.com/, http://www.intelligentbiosystems.com/ and http://www.completegenomicsinc.com/technology). At annual grantee meetings, open discussions of advances and challenges have stimulated collaboration and considerably accelerated research.

Several peer-reviewed articles have identified the strengths and limitations of commercial SBS platforms from a user's perspective[2–13] (for reviews, see refs. 5,14–16) and an assessment of the challenges facing nanopore sequencing has recently been published[17]. Here, we present current views of some of the investigators who received the technology development grants mentioned above on the underlying limitations and challenges of next-generation sequencing, with the goal of informing and engaging the broader research and engineering communities. Successful engagement requires the cross-pollination of ideas from experts across various disciplines to develop and identify solutions, as few scientists and engineers, including all but the most sophisticated users and active developers of technology, understand the full complexities of seamlessly integrating instrumentation, reagents and protocols necessary to promote scientific discovery.

SBS platforms

A common SBS strategy is to use DNA polymerase (Fig. 1) or ligase enzymes to extend many DNA strands in parallel. Nucleotides or short oligonucleotides are provided either one at a time or modified with identifying tags so that the base type of the incorporated nucleotide or oligonucleotide can be determined as extension proceeds.

SBS strategies may be categorized as either single molecule–based (involving the sequencing of a single molecule) or ensemble based (involving the sequencing of multiple indentical copies of a DNA molecule, typically amplified together on isolated surfaces or beads). They may be real-time (that is, with a free-running DNA polymerase given all nucleotides required) or synchronous-controlled (that is, using a priori temporal information to facilitate the identification process in a 'stop-and-go' iterative fashion). This can be achieved by using nucleotide substrates that are reversibly blocked or by simply adding only a single kind of nucleotide (e.g., dATP) at a time.

[1]GE Healthcare Life Sciences, Piscataway, New Jersey, USA. [2]LI-COR Biosciences, Lincoln, Nebraska, USA. [3]Foundation for Applied Molecular Evolution, Alachua, Florida, USA. [4]Harvard Medical School, Boston, Massachusetts, USA. [5]Howard Hughes Medical Institute, Janelia Farm Research Campus, Ashburn, Virginia, USA. [6]University of California, San Diego, La Jolla, Californi, USA. [7]Microchip Biotechnologies Inc., Dublin, California, USA. [8]General Electric Global Research Center, Niskayuna, New York, USA. [9]National Human Genome Research Institute, NIH, Bethesda, Maryland, USA. [10]University of Wisconsin-Madison, Madison, Wisconsin, USA. [11]Lehigh University, Bethlehem, Pennsylvania, USA. Correspondence should be addressed to C.W.F. (carl.fuller@alumni.upenn.edu).

Published online 6 November 2009; doi:10.1038/nbt.1585

nature biotechnology

Automated design of synthetic ribosome binding sites to control protein expression

Howard M Salis[1], Ethan A Mirsky[2] & Christopher A Voigt[1]

Microbial engineering often requires fine control over protein expression—for example, to connect genetic circuits[1–7] or control flux through a metabolic pathway[8–13]. To circumvent the need for trial and error optimization, we developed a predictive method for designing synthetic ribosome binding sites, enabling a rational control over the protein expression level. Experimental validation of >100 predictions in *Escherichia coli* showed that the method is accurate to within a factor of 2.3 over a range of 100,000-fold. The design method also correctly predicted that reusing identical ribosome binding site sequences in different genetic contexts can result in different protein expression levels. We demonstrate the method's utility by rationally optimizing protein expression to connect a genetic sensor to a synthetic circuit. The proposed forward engineering approach should accelerate the construction and systematic optimization of large genetic systems.

Trial-and-error mutation to optimize an engineered genetic circuit or metabolic pathway becomes prohibitively inefficient as the system's size and complexity grows. To address this problem, we developed a predictive design method that interconverts between the DNA sequence of a key genetic element—ribosome binding sites—and their function inside a genetic system (controlling the translation initiation rate and the protein expression level). The design method's capabilities enable the systematic optimization of genetic systems, which will be increasingly valuable as it becomes possible to synthesize larger pieces of DNA[14], including whole genomes[15].

In bacteria, ribosome binding sites (RBSs) and other regulatory RNA sequences are effective control elements for translation initiation[16–19], and thereby protein expression. Previous studies have generated libraries of RBS sequences with the goal of optimizing the function of a genetic system[1,7,18]. However, library size increases combinatorially with the number of proteins in the engineered system—for example, randomly mutating four nucleotides of an RBS generates a library of 256 sequences, thus requiring 256^3, or 16.7 million, sequences for three proteins and 256^6, or 2.8×10^{14}, sequences for six proteins.

In contrast to a library-based approach, we combined a biophysical model of translation initiation with an optimization algorithm to predict the sequence of a synthetic RBS sequence that provides a target translation initiation rate on a proportional scale. The model builds on previous work that characterized the free energies of key molecular interactions involved in translation initiation[20,21] and on measurements of the sequence-dependent energetic changes that occur during RNA folding and hybridization[22–26].

Bacterial translation consists of four phases: initiation, elongation, termination and ribosome turnover (**Fig. 1a**)[27]. In most cases, translation initiation is the rate-limiting step. Its rate is determined by multiple molecular interactions, including the hybridization of the 16S rRNA to the RBS sequence, the binding of $tRNA^{fMET}$ to the start codon, the distance between the 16S rRNA binding site and the start codon (called spacing) and the presence of RNA secondary structures that occlude either the 16S rRNA binding site or the standby site[20,21,28–31].

Our equilibrium statistical thermodynamic model quantifies the strengths of the molecular interactions between an mRNA transcript and the 30S ribosome complex—which includes the 16S rRNA and the $tRNA^{fMET}$—to predict the resulting translation initiation rate, r (equation (1), derived in **Supplementary Methods**).

$$r \propto \exp(-\beta \Delta G_{tot}) \quad (1)$$

The model describes the system as having two states separated by a reversible transition (**Fig. 1b**). The initial state is the folded mRNA transcript and the free 30S complex. The final state is the assembled 30S pre-initiation complex bound on an mRNA transcript. The difference in Gibbs free energy between these two states (ΔG_{tot}) depends on the mRNA sequence surrounding a specified start codon. ΔG_{tot} is more negative when attractive interactions between ribosome and mRNA are present, and ΔG_{tot} is more positive when mutually exclusive secondary structures are present. β is the apparent Boltzmann constant for the system, which converts thermodynamic free energies to temperature differences. Importantly, equation (1) describes the differences in translation initiation rate that result from differences in mRNA sequence. The amount of expressed protein is proportional to the translation initiation rate where the proportionality factor accounts for any ribosome-mRNA molecular interactions that are independent of mRNA sequence and any translation-independent parameters, such as the DNA copy number, the promoter's transcription rate, the mRNA stability and the protein dilution rate (**Supplementary Fig. 1**).

[1]Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, California, USA. [2]Graduate Group in Biophysics, University of California San Francisco, San Francisco, California, USA. Correspondence should be addressed to C.A.V. (cavoigt@picasso.ucsf.edu).

Received 30 July; accepted 8 September; published online 4 October 2009; doi:10.1038/nbt.1568

nature biotechnology

Induction of protein-protein interactions in live cells using light

Masayuki Yazawa[1], Amir M Sadaghiani[1], Brian Hsueh[1,2] & Ricardo E Dolmetsch[1]

Protein-protein interactions are essential for many cellular processes. We have developed a technology called light-activated dimerization (LAD) to artificially induce protein hetero- and homodimerization in live cells using light. Using the FKF1 and GIGANTEA (GI) proteins of *Arabidopsis thaliana*, we have generated protein tags whose interaction is controlled by blue light. We demonstrated the utility of this system with LAD constructs that can recruit the small G-protein Rac1 to the plasma membrane and induce the local formation of lamellipodia in response to focal illumination. We also generated a light-activated transcription factor by fusing domains of GI and FKF1 to the DNA binding domain of Gal4 and the transactivation domain of VP16, respectively, showing that this technology is easily adapted to other systems. These studies set the stage for the development of light-regulated signaling molecules for controlling receptor activation, synapse formation and other signaling events in organisms.

The ability to experimentally control protein activity is essential for deciphering many cellular, developmental and physiological processes. Small organic molecules have been developed to regulate protein activity by controlling protein-protein interactions, but these have relatively poor temporal and spatial resolution and suffer from a lack of specificity. Light-inducible systems are attractive for applications in which high spatial and temporal specificity is needed. They are also less likely to cause non-specific effects. Several technologies have been developed recently that use light-sensitive proteins to control electrical activity[1], cyclic AMP production[2] and activation of G protein–coupled receptors[3,4] in mammalian cells. Despite the utility of these approaches, they are not easily adapted to control signal transduction cascades that depend on protein-protein interactions.

To develop a method to control protein interactions using light, we took advantage of the light-dependent binding of FKF1 to GI, two proteins that control flowering in *Arabidopsis thaliana*[5]. FKF1 contains a 'light, oxygen or voltage' (LOV) domain that detects light using flavin mononucleotide (FMN), a derivative of riboflavin that is abundant in all eukaryotic cells[5–7] (**Fig. 1a**). Illumination with 450 nm light induces formation of a covalent bond between FMN and cysteine 91 of FKF1[8], which allows FKF1 to bind to the nuclear protein GI. The cysteinyl-flavin bond is subsequently hydrolyzed returning the LOV domain to its resting state and dissolving the FKF1-GI complex.

We first investigated the expression and localization of FKF1 and GI in mammalian cells by labeling the proteins with hemagglutinin (HA) and FLAG tags and introducing them into NIH 3T3 cells. Detection of the proteins with anti-HA and FLAG antibodies revealed that FKF1 and GI were abundantly expressed in the nucleus of cells reflecting their endogenous role in the signaling cascade associated with flowering[9] (**Supplementary Fig. 1**). Because nuclear localization is undesirable in a general-purpose protein interaction system, we used the PSORT II prediction program[10,11] to identify NLS sequences in both proteins (FKF1, amino acids (aa) 13–16 and GI aa 607–610; **Supplementary Fig. 1a**). Disruption of these sequences by substituting threonines for the lysines at positions 15 of FKF1 (K15T) and 607 of GI (K607T) substantially reduced the concentration of both proteins in the nucleus (**Supplementary Fig. 1**).

To determine whether FKF1 and GI interact with each other in mammalian cells in response to blue light, we first labeled GI with the fluorescent protein mCherry and targeted this protein to the cell membrane using the C-terminal farnesylation motif from K-Ras called the CAAX box (**Fig. 1b**)[12]. We then labeled FKF1 with yellow fluorescent protein (YFP), introduced it into NIH3T3 cells together with GI-mCherry-CAAX, and measured the localization of both proteins using epifluorescence time-lapse microscopy. At rest GI-mCherry-CAAX was localized on the cell membrane whereas FKF1-YFP was cytoplasmic. Within 5 min of illumination with 450 nm light, however, FKF1-YFP translocated to the membrane of about 30% of the cells (**Fig. 1c,d**). In contrast, cells expressing FKF1-YFP without GI-mCherry-CAAX did not show any increase in membrane fluorescence, suggesting that FKF1-YFP membrane recruitment depended on binding to GI-mCherry-CAAX (**Fig. 1d** and **Supplementary Fig. 2a**). These results indicate that FKF1 binds to GI in mammalian cells after a short exposure to light. FKF1-YFP remained at the membrane for at least 1.5 h after exposure to weak blue light (**Fig. 1d**), consistent with studies demonstrating that the cysteinyl-flavin adduct of FKF1 has a long half-life at room temperature[8].

The failure to detect light-inducible recruitment of FKF1 to the membrane in 70% of the cells suggested that this process is relatively inefficient. Close examination of the cells that did not show FKF1-YFP translocation to the membrane revealed that these cells expressed FKF1-YFP and GI-mCherry-CAAX at high levels and had membranous FKF1-YFP before illumination (**Fig. 1e**, left panels). Measurement of the expression of FKF1-YFP and GI-mCherry-CAAX and of the amount of FKF1-YFP at the membrane

[1]Department of Neurobiology, Stanford University School of Medicine, Stanford, California, USA. [2]Present address: Department of Chemistry, Princeton University, Princeton, New Jersey, USA. Correspondence should be addressed to R.E.D. (ricardo.dolmetsch@stanford.edu).

Received 16 March; accepted 9 September; published online 4 October 2009; doi:10.1038/nbt.1569

nature biotechnology

Microdroplet-based PCR enrichment for large-scale targeted sequencing

Ryan Tewhey[1,2], Jason B Warner[3], Masakazu Nakano[1,4], Brian Libby[3], Martina Medkova[3], Patricia H David[3], Steve K Kotsopoulos[3], Michael L Samuels[3], J Brian Hutchison[3], Jonathan W Larson[3], Eric J Topol[1], Michael P Weiner[3,4], Olivier Harismendy[1,4], Jeff Olson[3], Darren R Link[3] & Kelly A Frazer[1,4]

Targeted enrichment of specific loci of the human genome is a promising approach to enable sequencing-based studies of genetic variation in large populations. Here we describe an enrichment approach based on microdroplet PCR, which enables 1.5 million amplifications in parallel. We sequenced six samples enriched by microdroplet or traditional singleplex PCR using primers targeting 435 exons of 47 genes. Both methods generated similarly high-quality data: 84% of the uniquely mapping reads fell within the targeted sequences; coverage was uniform across ~90% of targeted bases; sequence variants were called with >99% accuracy; and reproducibility between samples was high ($r^2 = 0.9$). We scaled the microdroplet PCR to 3,976 amplicons totaling 1.49 Mb of sequence, sequenced the resulting sample with both Illumina GAII and Roche 454, and obtained data with equally high specificity and sensitivity. Our results demonstrate that microdroplet technology is well suited for processing DNA for massively parallel enrichment of specific subsets of the human genome for targeted sequencing.

Technical advances in sequencing methods and instruments are rapidly transforming our ability to study both common and rare genetic variants in the human genome. Indeed, several human genomes have already been sequenced in their entirety[1–4]. However, for the time being, the cost for sequencing whole human genomes is prohibitive for addressing research questions in a large cohort of individuals.

Three approaches are currently being used for enrichment of target sequences of interest. The first approach is traditional singleplex PCR, which has been used for hundreds of samples to examine large kilobase-sized contiguous intervals[5] or the exons of hundreds of genes[6,7]. Although traditional PCR enriches target sequences with high specificity and sensitivity, it is difficult to scale the method to match the throughput of current sequencing instruments.

The second approach is based on multiplex amplification of thousands of target sequences in a single tube by array-synthesized padlock 'molecular inversion' probes[8–10]. Molecular inversion probes allow for a highly efficient multiplex reaction owing to the tethering of primer pairs by a DNA linker. However, published results show that although the captures are highly specific and represent upwards of 90% of the targets, there is >100-fold range in coverage of the targeted sequences, and 34–42% of sequence capacity is consumed by either sequencing of primer sequences or the molecular inversion probes' linker backbone[9] (**Supplementary Discussion**).

The third approach, based on hybridization with long oligonucleotides that are either matrix-bound or in solution, captures and pulls down the target sequences[11–14]. The hybridization-based methods have good capture rates, uniform coverage of target sequences and good reproducibility. However, the methods are known to be biased to repetitive elements, which can result in a high proportion of reads that map nonuniquely. Additionally, sequences that are highly homologous to other sequences in the genome cannot be individually targeted.

We have developed an approach, involving microdroplet-based technology, which takes advantage of the high specificity and sensitivity of PCR and allows for massively parallel singleplex amplification of complex target sequences. The discrete encapsulation of microdroplet PCR reactions prevents possible primer pair interactions, allowing for highly efficient simultaneous amplification of up to 4,000 targeted sequences and greatly reduces the amount of reagents required.

RESULTS

Microdroplet PCR Workflow

Microdroplet technology is particularly well suited for processing DNA for massively parallel amplification of sequencing targets. It involves preparation of 1.5 million separate PCR reactions from 20 µl of template solution containing just 7.5 µg of genomic DNA. Microdroplet PCR requires the following steps (**Fig. 1**): merging picoliter volume droplets of fragmented genomic DNA template with premade primer pair droplets (primer library), pooled thermal cycling of the resulting PCR reactions (droplet PCR), and destabilizing the droplets to release the PCR product (break emulsion) for purification and sequencing.

[1]Scripps Genomic Medicine, Scripps Translational Science Institute, The Scripps Research Institute, La Jolla, California, USA. [2]Division of Biological Sciences, University of California at San Diego, La Jolla, California, USA. [3]RainDance Technologies, Lexington, Massachusetts, USA. [4]Present addresses: Moores UCSD Cancer Center, La Jolla, California, USA (M.N.) and Affomix, Branford, Connecticut, USA (M.P.W.). Correspondence should be addressed to K.A.F. (kfrazer@ucsd.edu) or D.R.L. (dlink@raindancetech.com).

Received 15 June; accepted 1 October; published online 1 November 2009; doi:10.1038/nbt.1583

nature biotechnology

RESOURCE

The transcription unit architecture of the *Escherichia coli* genome

Byung-Kwan Cho[1], Karsten Zengler[1], Yu Qiu[1], Young Seoub Park[1], Eric M Knight[1,3], Christian L Barrett[1], Yuan Gao[2] & Bernhard Ø Palsson[1]

Bacterial genomes are organized by structural and functional elements, including promoters, transcription start and termination sites, open reading frames, regulatory noncoding regions, untranslated regions and transcription units. Here, we iteratively integrate high-throughput, genome-wide measurements of RNA polymerase binding locations and mRNA transcript abundance, 5′ sequences and translation into proteins to determine the organizational structure of the *Escherichia coli* K-12 MG1655 genome. Integration of the organizational elements provides an experimentally annotated transcription unit architecture, including alternative transcription start sites, 5′ untranslated region, boundaries and open reading frames of each transcription unit. A total of 4,661 transcription units were identified, representing an increase of >530% over current knowledge. This comprehensive transcription unit architecture allows for the elucidation of condition-specific uses of alternative sigma factors at the genome scale. Furthermore, the transcription unit architecture provides a foundation on which to construct genome-scale transcriptional and translational regulatory networks.

Over the last decade, considerable progress has been made in determining whole genome sequences of bacteria and in describing their transcriptomes and proteomes[1–5]. Despite these advances, however, the in-depth organizational structure of bacterial genomes has not been fully elucidated. Understanding the organizational structure of bacterial genomes is of fundamental importance as it dictates the flow of genetic information at the systems level. Bacterial genomes are highly organized into various structural and functional elements. These organizational elements include, but are not limited to, promoters, transcription start sites (TSSs), open reading frames (ORFs), regulatory noncoding regions, untranslated regions (UTRs) and transcription units (transcription units). A bacterial transcription unit is defined as having one or more ORFs that are transcribed from one promoter into a single mRNA.

With the publication of the first full genome sequence in the mid-1990s, it became possible, in principle, to identify all the gene products involved in complex biological processes in a single organism[6]. In practice, almost 15 years later, this has proved difficult to accomplish using sequence information alone. Multiple simultaneous genome-scale measurements are therefore needed to identify all gene products and, more generally, to determine their cellular locations and their interactions with the genome (e.g., transcription factor binding to regulatory sequences). Establishing the organizational structure of a genome is a challenging task. In-depth analyses of the transcriptomes and proteomes of multiple prokaryotic organisms indicate that the information content and structure of a genome is much more complex than previously thought[4,7–9] and that the process of revealing the role of cellular components in transcription and translation on a genome scale has just begun[2,10–14]. Here we describe a four-step systems approach that iteratively integrates multiple genome-scale measurements on the basis of genetic information flow to identify the organizational elements and map them onto the genome sequence (**Figs. 1** and **2a**)[15]. We have applied the approach to the *E. coli* K-12 MG1655 genome to generate a detailed description of its transcription unit architecture.

RESULTS

Determination of RNA polymerase–binding regions

The first step of the flow of genetic information is its transfer into mRNA by transcription. Although mRNA synthesis is highly regulated by external signals, it always requires binding of RNA polymerase (RNAP) to a promoter region. We therefore integrated RNAP-binding regions and mRNA transcript abundance to determine segments of contiguous transcription originating from promoter regions. To identify RNAP-binding regions at the genome scale, we used a ChIP-chip method for *E. coli* K-12 MG1655 grown in the presence or absence of rifampicin (a small-molecular inhibitor that traps RNAP at promoters) under multiple growth conditions[14,16]. Using an antibody specific to the RNAP β-subunit, we obtained RNAP-associated DNA fragments that were then fluorescently labeled and hybridized to a high-density oligonucleotide tiling microarray representing the entire *E. coli* genome[14]. Rifampicin treatment generated a genome-wide static map of RNAP-binding regions compared to a dynamic map of RNAP-binding regions without rifampicin treatment (**Fig. 2a**)[14]. From the static map, we identified a total of 1,511 and 1,444 RNAP-binding regions on the forward and reverse strand,

[1]Department of Bioengineering, University of California, San Diego, La Jolla, California, USA. [2]Center for the Study of Biological Complexity and Department of Computer Science, Virginia Commonwealth University, Richmond, Virginia, USA. [3]Present address: Center for Systems Biology, University of Iceland, Vatnsmyravegar, Reykjavik, Iceland. Correspondence should be addressed to B.Ø.P. (bpalsson@ucsd.edu).

Received 20 April; accepted 1 October; published online 1 November 2009; doi:10.1038/nbt.1582

nature genetics

A new statistic and its power to infer membership in a genome-wide association study using genotype frequencies

Kevin B Jacobs[1–3], Meredith Yeager[1,2], Sholom Wacholder[2], David Craig[4], Peter Kraft[5], David J Hunter[5], Justin Paschal[6], Teri A Manolio[7], Margaret Tucker[2], Robert N Hoover[2], Gilles D Thomas[2], Stephen J Chanock[2,8] & Nilanjan Chatterjee[2,8]

Aggregate results from genome-wide association studies (GWAS)[1–3], such as genotype frequencies for cases and controls, were until recently often made available on public websites[4,5] because they were thought to disclose negligible information concerning an individual's participation in a study. Homer *et al.*[6] recently suggested that a method for forensic detection of an individual's contribution to an admixed DNA sample could be applied to aggregate GWAS data. Using a likelihood-based statistical framework, we developed an improved statistic that uses genotype frequencies and individual genotypes to infer whether a specific individual or any close relatives participated in the GWAS and, if so, what the participant's phenotype status is. Our statistic compares the logarithm of genotype frequencies, in contrast to that of Homer *et al.*[6], which is based on differences in either SNP probe intensity or allele frequencies. We derive the theoretical power of our test statistics and explore the empirical performance in scenarios with varying numbers of randomly chosen or top-associated SNPs.

A recent publication by Homer *et al.*[6] reported a method used to determine whether a given individual contributed a trace amount (<0.1%) of DNA to a pool of DNA from 200 individuals based on allelic probe intensities from the same microarray genotyping technology commonly employed in GWAS. Homer's test statistic, denoted here as T_{allele}, required three sets of probe data: (i) from a pool to which the individual may or may not have contributed DNA; (ii) from a reference pool of DNA sampled from members of the same genetic population; and (iii) from DNA obtained from a single individual to be tested for membership in the first pool. The primary applications in the study by Homer *et al.* were directed toward specific forensic challenges, such as determining whether a DNA sample contributed to a biospecimen of mixed origin. The authors also discussed potential applications of their method to GWAS allele-frequency data.

We now report an extension of this with a new test statistic, T_{geno}, that can be used to detect whether an individual contributed DNA to a GWAS, and, if so, to determine the individual's case or control status. Conceptually, T_{geno} substitutes frequencies from two groups (for example, cases and controls) and genotype states of the individual to be tested in place of the pools of admixed DNA and measures of allelic probe intensities as used by Homer *et al.*[6] T_{geno} is a new likelihood-ratio statistic that can detect whether an individual contributed DNA to neither, one or both groups given only genotypes for the individual to be tested and genotype frequencies from each group. Our statistic is similar to T_{allele}, except that T_{geno} contrasts the logarithm of the frequencies of each observed genotype of the tested individual between two populations (as opposed to allelic probe intensity or allele frequency as done for T_{allele}).

The predictive power of T_{geno} was derived approximately (see Online Methods) and evaluated by simulation under a range of scenarios consistent with the size and scope of current GWAS designs. In each scenario, we assumed two independent groups of individuals, denoted 'test' and 'reference', drawn from a homogeneous population with genotypes from independent biallelic SNP markers in Hardy-Weinberg equilibrium with fixed minor allele frequency (MAF). We varied the size of each group, the number and MAF of independent loci and the genotype error rate. Power was computed by simulation using random sampling of test and reference groups as well as individuals to be tested and by comparison of T_{geno} to its theoretical null distribution using a two-sided test that assumed an individual could be a member of either group.

Table 1 shows simulation results of the sensitivity and specificity of T_{geno} for significance levels ranging from 0.05 to 10^{-6}. We observed that the power to detect membership in either group increased as

[1]Core Genotyping Facility, Advanced Technology Program, SAIC-Frederick, Inc., National Cancer Institute at Frederick, Frederick, Maryland, USA. [2]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA. [3]BioInformed LLC, Gaithersburg, Maryland, USA. [4]Neurogenomics Division, The Translational Genomics Research Institute, Phoenix, Arizona, USA. [5]Program in Molecular and Genetic Epidemiology, Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. [6]National Center for Biotechnology Information, National Library of Medicine, and [7]Office of Population Genomics, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA. [8]These authors contributed equally to this work. Correspondence should be addressed to K.B.J. (jacobske@mail.nih.gov).

Received 22 April; accepted 4 September; published online 4 October 2009; doi:10.1038/ng.455

High-resolution, long-term characterization of bacterial motility using optical tweezers

Taejin L Min[1,2], Patrick J Mears[1,2], Lon M Chubiz[3], Christopher V Rao[3], Ido Golding[1,2,4] & Yann R Chemla[1,2,4]

We present a single-cell motility assay, which allows the quantification of bacterial swimming in a well-controlled environment, for durations of up to an hour and with a temporal resolution greater than the flagellar rotation rates of ~100 Hz. The assay is based on an instrument combining optical tweezers, light and fluorescence microscopy, and a microfluidic chamber. Using this device we characterized the long-term statistics of the run-tumble time series in individual *Escherichia coli* cells. We also quantified higher-order features of bacterial swimming, such as changes in velocity and reversals of swimming direction.

Many microorganisms move around by swimming in liquid medium and can modulate their swimming behavior to move up gradients of chemicals, temperature or light. In liquid environments, *Escherichia coli* swims in a random pattern composed of 'runs', in which the cell maintains an approximately constant direction, and 'tumbles', in which the cell stops and randomly changes direction[1]. Runs and tumbles are generated by different states of the motors that rotate the bacterial flagella. Each cell has several flagellar motors that can rotate either clockwise or counterclockwise. When the motors turn counterclockwise, the flagella rotate together in a bundle and push the cell forward. When one or more of the motors turn clockwise, some flagella may break from the bundle and cause the cell to tumble and randomize its orientation. During chemotaxis, *E. coli* biases its 'random walk' based on temporal changes in chemical concentration. When the bacterium moves up a gradient of attractant, it detects an increase in attractant concentration and reduces its probability of tumbling. The result is that the cell tends to continue going up the gradient.

The modulation of bacterial swimming serves as a model system for the way a living cell processes signals from its environment and changes its behavior based on those signals[1,2]. Standard methods for assaying bacterial swimming and chemotaxis typically fall into two categories. The first consists of observing freely swimming cells, typically in a flow-cell setup. Chemoeffector variation is created in space or time[3–5], and the change in swimming behavior is then examined[6,7]. The second type of assay uses cells that are tethered to a surface, usually a microscope slide, so that the rotation of an individual flagellar motor can be followed[8,9].

These approaches have enabled the acquisition of large amounts of data that have yielded important insights into bacterial swimming and its modulation. However, both assays are limited in their ability to quantify whole-cell swimming (**Supplementary Note 1**). Here we describe an optical trap–based assay to investigate cell motility. This assay allowed us to quantify bacterial swimming in a well-controlled environment for durations up to 1 h and at data acquisition rates that are faster than the ~100 Hz flagellar rotation rates. We thus characterized the long-term statistics of the run-tumble time series in individual cells. Moreover, we characterized higher-order features of bacterial swimming, such as changes in velocity and reversals of swimming direction.

RESULTS

Experimental setup

Our single-cell motility assay involves a custom-made instrument combining optical tweezers, light and fluorescence microscopy and a microfluidic chamber (**Fig. 1a**). The optical tweezers consist of two traps generated by two orthogonally polarized beams from a single 1,064-nm diode-pumped solid-state laser[10]. The separation between the two traps is controlled by a piezo-actuated mirror stage. A custom flow-cell (**Supplementary Fig. 1** and Online Methods) serves as the experimental trap chamber and can be displaced relative to the two traps in all directions by a three-axis translational stage. For measurements of bacterial motility, we filled the chambers with a tryptone broth–based 'trapping medium', though other buffers are also appropriate (Online Methods). We injected bacteria into a top 'antechamber' and flowed them through a narrow inlet into the bottom channel, where they were captured by the traps. Trapping a rod-shaped bacterium by each end with two optical traps[11] allowed us to orient the cell at will in the plane of the chamber (**Fig. 1b**). We visualized trapped bacteria either by brightfield or epifluorescence microscopy (**Fig. 1c** and Online Methods).

Despite immobilization by the optical traps, cells displayed motile behavior, evinced by flagellar bundle rotation and counter-rotation ('rolling') of the cell body[12]. This behavior was detected directly and sensitively by the optical traps themselves, by imaging light from both orthogonally polarized trapping beams onto two separate position-sensitive photodetectors. Consistent with

[1]Department of Physics, [2]Center for the Physics of Living Cells, [3]Department of Chemical and Biomolecular Engineering and [4]Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA. Correspondence should be addressed to I.G. (igolding@illinois.edu).

RECEIVED 15 MAY; ACCEPTED 8 SEPTEMBER; PUBLISHED ONLINE 4 OCTOBER 2009; DOI:10.1038/NMETH.1380

Compensatory Evolution of Gene Regulation in Response to Stress by *Escherichia coli* Lacking RpoS

Daniel M. Stoebel[1], Karsten Hokamp[2], Michael S. Last[3], Charles J. Dorman[1]*

1 Department of Microbiology and Moyne Institute of Preventive Medicine, School of Genetics and Microbiology, Trinity College, Dublin, Ireland, **2** Smurfit Institute of Genetics, School of Genetics and Microbiology, Trinity College, Dublin, Ireland, **3** UC Toxic Substance Research and Teaching Program, University of California Davis, Davis, California, United States of America

Abstract

The RpoS sigma factor protein of *Escherichia coli* RNA polymerase is the master transcriptional regulator of physiological responses to a variety of stresses. This stress response comes at the expense of scavenging for scarce resources, causing a trade-off between stress tolerance and nutrient acquisition. This trade-off favors non-functional *rpoS* alleles in nutrient-poor environments. We used experimental evolution to explore how natural selection modifies the regulatory network of strains lacking RpoS when they evolve in an osmotically stressful environment. We found that strains lacking RpoS adapt less variably, in terms of both fitness increase and changes in patterns of transcription, than strains with functional RpoS. This phenotypic uniformity was caused by the same adaptive mutation in every independent population: the insertion of IS*10* into the promoter of the *otsBA* operon. OtsA and OtsB are required to synthesize the osmoprotectant trehalose, and transcription of *otsBA* requires RpoS in the wild-type genetic background. The evolved IS*10* insertion rewires expression of *otsBA* from RpoS-dependent to RpoS-independent, allowing for partial restoration of wild-type response to osmotic stress. Our results show that the regulatory networks of bacteria can evolve new structures in ways that are both rapid and repeatable.

Citation: Stoebel DM, Hokamp K, Last MS, Dorman CJ (2009) Compensatory Evolution of Gene Regulation in Response to Stress by *Escherichia coli* Lacking RpoS. PLoS Genet 5(10): e1000671. doi:10.1371/journal.pgen.1000671

Editor: David S. Guttman, University of Toronto, Canada

Received February 16, 2009; **Accepted** September 2, 2009; **Published** October 2, 2009

Funding: This work was supported by grant 080054/Z/06/Z from the Wellcome Trust and grant 07/IN1/B918 from Science Foundation Ireland. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: cjdorman@tcd.ie

Introduction

Bacterial adaptation to environmental stress involves, in part, a modification of transcription patterns, with downstream impacts on physiological function. In *Escherichia coli*, the RNA polymerase sigma factor RpoS is a global regulator that coordinates the expression of up to 10% of the genome when the bacterium enters stationary phase or experiences stresses such as starvation, acidity or increased osmolarity [1]. Despite the importance of this protein in many environments, a functional RpoS seems to lower the ability of *E. coli* to scavenge for scarce nutrients [2,3]. This cost is hypothesized to occur because there is a limiting amount of core RNA polymerase subunits in the cell, meaning that transcription of stress responsive, RpoS-dependent promoters will decrease the transcription from RpoS-independent promoters involved in nutrient acquisition and utilization [2–4].

The hypothesis that the nature of the RpoS regulatory network creates an inherent conflict between stress protection and nutritional competence (SPANC) [3,4] provides a basis for predicting how natural selection acts on the global regulatory networks of *E. coli*. The SPANC hypothesis predicts that natural selection will modify the network in favor of nutritional ability at the expense of stress resistance in some environments by decreasing or abolishing RpoS function. Just this type of selection against RpoS activity has been documented in laboratory studies [2,5,6]. In addition, strains with low- or null-activity *rpoS* alleles have been found in natural populations of *E. coli* and *Salmonella enterica* [3,7].

While strains without functional RpoS are favored in some environments, those same strains may do less well in other, more stressful environments where they may be less fit due to an inability to respond to new challenges. While *rpoS* strains could adapt by recovering or increasing their RpoS function, the mutations that abolished RpoS function may be very unlikely or impossible to reverse. An alternative mechanism involves the selection of mutations that modify the regulatory network to compensate for the loss of RpoS. These compensatory mutations would then increase the fitness of the bacterium in this new, more stressful environment. We sought to understand this type of adaptation by observing the patterns of increased fitness seen in evolving bacterial lines, and by elucidating the molecular basis of the adaptation.

Results from previous experimental studies suggest that compensation for deleterious mutations is a general phenomenon [8–11]. Less is known, however, about the variability of the process of compensation. Will strains that lack RpoS adapt more or less variably to a stressful environment than strains with a fully functional regulatory network? Will this involve larger or smaller increases in fitness? At a molecular level, mutations affecting other global regulators of transcription [1], or local changes at a promoter [12,13] may permit transcription in the absence of RpoS, but we were interested in discovering which options actually are favored by natural selection. Would only a few key genes be involved in adaptation, or would adaptation involve changes in large parts of the transcriptome? Here we used

Fast Growth Increases the Selective Advantage of a Mutation Arising Recurrently during Evolution under Metal Limitation

Hsin-Hung Chou, Julia Berthet¤, Christopher J. Marx*

Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, United States of America

Abstract

Understanding the evolution of biological systems requires untangling the molecular mechanisms that connect genetic and environmental variations to their physiological consequences. Metal limitation across many environments, ranging from pathogens in the human body to phytoplankton in the oceans, imposes strong selection for improved metal acquisition systems. In this study, we uncovered the genetic and physiological basis of adaptation to metal limitation using experimental populations of *Methylobacterium extorquens* AM1 evolved in metal-deficient growth media. We identified a transposition mutation arising recurrently in 30 of 32 independent populations that utilized methanol as a carbon source, but not in any of the 8 that utilized only succinate. These parallel insertion events increased expression of a novel transporter system that enhanced cobalt uptake. Such ability ensured the production of vitamin B_{12}, a cobalt-containing cofactor, to sustain two vitamin B_{12}–dependent enzymatic reactions essential to methanol, but not succinate, metabolism. Interestingly, this mutation provided higher selective advantages under genetic backgrounds or incubation temperatures that permit faster growth, indicating growth-rate–dependent epistatic and genotype-by-environment interactions. Our results link beneficial mutations emerging in a metal-limiting environment to their physiological basis in carbon metabolism, suggest that certain molecular features may promote the emergence of parallel mutations, and indicate that the selective advantages of some mutations depend generically upon changes in growth rate that can stem from either genetic or environmental influences.

Citation: Chou H-H, Berthet J, Marx CJ (2009) Fast Growth Increases the Selective Advantage of a Mutation Arising Recurrently during Evolution under Metal Limitation. PLoS Genet 5(9): e1000652. doi:10.1371/journal.pgen.1000652

Editor: Ivan Matic, Université Paris Descartes, INSERM U571, France

Received May 20, 2009; **Accepted** August 17, 2009; **Published** September 18, 2009

Funding: This project was supported by grant 1R01 GM078209 from the National Institutes of Health (www.nih.gov). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: cmarx@oeb.harvard.edu

¤ Current address: Swarthmore College, Swarthmore, Pennsylvania, United States of America

Introduction

Adaptation is a product of genetic modification and natural selection imposed by environmental challenges. A complete understanding of adaptation of biological systems thus requires identification of how selection acts upon organismal traits and mapping adaptive phenotypes to underlying genotypic changes. Experimentally testing the genotype-phenotype association and phenotypic effects of mutations is an ongoing research direction in many fields of biology [1–3]. Studies on mutations have shown that genetic interactions (epistasis) are common in biological systems [4–7] and fitness effects of beneficial mutations can vary greatly depending on environmental conditions (genotype-by-environment interactions, G×E) [8–10]. Many studies of beneficial mutations, however, stop short of elucidating the exact molecular mechanisms connecting genotypic changes to phenotypic adaptation [11–13]. The lack of this level of information has rendered prediction of fitness effects, epistasis, and G×E interactions elusive. On the other hand, much of our current knowledge of biological systems has come from studying phenotypes of deleterious gene knockouts. Such approaches have uncovered many gene functions and genetic interactions but provided little information about the quantitative response of biological networks to environmental or genetic perturbations as well as the functional significance of a gene in the context of adaptation. A complementary approach to studying the function and evolution of biological systems, therefore, is to characterize molecular mechanisms through which beneficial mutations alter physiology, and reciprocally, how physiological differences due to genetic backgrounds or environments influence the effects of beneficial mutations.

In recent years, evolution experiments using microorganisms have offered a powerful means to investigate the genetic basis of adaptation [14]. Evolution of experimental populations is often conducted using resource-limiting conditions, a challenge many organisms encounter in nature. One competitive strategy to survive under such a scenario is to enhance resource uptake through transport systems. If physiological acclimation is insufficient to alleviate resource limitation, natural selection can favor mutations that further increase uptake capacity. Phenotypes competent to import resources at low concentrations emerge frequently in microbial populations subjected to evolution under resource limitation [10,15,16]. Interestingly, beneficial mutations emerging from evolution experiments often occur repeatedly at

Analysis of Genetic Toggle Switch Systems Encoded on Plasmids

Adiel Loinger and Ofer Biham
Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel
(Received 16 December 2008; published 7 August 2009)

Genetic switch systems with mutual repression of two transcription factors, encoded on plasmids, are studied using stochastic methods. The plasmid copy number is found to strongly affect the behavior of these systems. More specifically, the average time between spontaneous switching events quickly increases with the number of plasmids. It was shown before that for a single copy encoded on the chromosome, the exclusive switch is more stable than the general switch. Here we show that when the switch is encoded on a sufficiently large number of plasmids, the situation is reversed and the general switch is more stable than the exclusive switch. These predictions can be tested experimentally using methods of synthetic biology.

DOI: 10.1103/PhysRevLett.103.068104 PACS numbers: 87.10.Mn, 87.10.Rt, 87.16.A−, 87.16.dj

Regulation processes in cells are performed by networks of interacting genes, which control each other's expression [1,2]. In recent years these networks have been studied extensively in different organisms. It was proposed that the networks exhibit a modular structure [3–5]. In particular, they include modules or motifs which may perform specific functions and appear in different parts of the network. Common examples of such motifs are the autoregulator [6] and the feed forward loop [4].

In addition to the genetic circuits found in natural organisms, in recent years it has became possible to construct synthetic circuits of a desired architecture [7,8]. These circuits are constructed from available components, namely, genes and promoters. They do not require the manipulation of the structure of proteins and other regulatory elements at the molecular level. These genes and promoters are often inserted into plasmids rather than on the chromosome. This is advantageous because plasmids are easier to manipulate and give rise to a stronger signal. Important examples of synthetic circuits, which were constructed on plasmids in *Escherichia coli*, include the genetic toggle switch [9] which exhibits bistability and the repressilator [10] which exhibits oscillations.

Measurements of gene expression in single cells have shown significant variability in populations of genetically identical cells [11–14]. These results focused the attention on the role of fluctuations and noise in gene expression processes [15–19]. The stochastic analysis is required because some of the proteins and their binding sites appear in low copy numbers. The copy number of the plasmids on which the circuit is encoded can be used as a control parameter for the noise level in the circuit. This is due to the fact that one copy of the circuit is encoded on each plasmid. Thus, increasing the number of plasmids enhances the synthesis rate of proteins and increases the copy numbers of free and bound proteins. Therefore, the plasmids copy number is an important parameter, which should be taken into account in theoretical and computational studies of such circuits. In previous studies, much attention was given to the control of the plasmids copy number itself [20,21]. Recently, it was shown that the plasmids copy number affects the behavior of genetic circuits encoded on them [22–24]. In a recent paper [25] we performed a comprehensive analysis of the repressilator. In particular we have studied the effect of the plasmids copy number of the behavior. We have found that as the number of plasmids is increased the fluctuations are reduced and the stochastic simulation results coincide with those obtained from the rate equations.

In this Letter we analyze the effect of the plasmids copy number on the genetic toggle switch system. The toggle switch consists of two genes, a and b, which negatively regulate each other by transcriptional regulation (Fig. 1). This system was constructed on plasmids using methods of synthetic biology [9]. It was found that under suitable conditions this system exhibits bistability, namely, two stable steady states. In each of these states, one of the proteins (A or B) appears in a high copy number, while the other protein is suppressed. Induced transitions between the two states were demonstrated.

In subsequent theoretical work [26–29], the conditions for bistability were elucidated. Taking into account the effect of noise and fluctuations, spontaneous transitions were found to take place between the two states. The stability of the switch can be characterized by the switch-

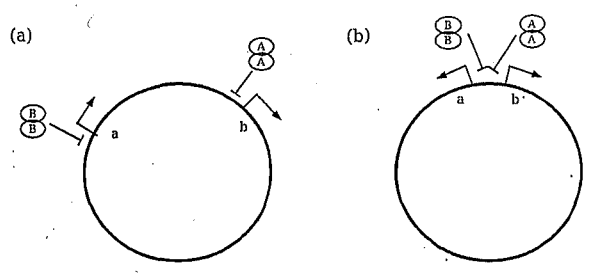

FIG. 1. Schematic illustrations of the general switch (a) and the exclusive switch (b), encoded on a plasmid. The transcriptional regulation is performed by dimers.

Gated Narrow Escape Time for Molecular Signaling

Jürgen Reingruber and David Holcman

Department of Computational Biology, Ecole Normale Supérieure, 46 rue d'Ulm 75005 Paris, France

(Received 30 July 2009; published 30 September 2009)

The mean time for a diffusing ligand to activate a target protein located on the surface of a microdomain can regulate cellular signaling. When the ligand switches between various states induced by chemical interactions or conformational changes, while target activation occurs in only one state, this activation time is affected. We investigate this dynamics using new equations for the sojourn times spent in each state. For two states, we obtain exact solutions in dimension one, and asymptotic ones confirmed by Brownian simulations in dimension 3. We find that the activation time is quite sensitive to changes of the switching rates, which can be used to modulate signaling. Interestingly, our analysis reveals that activation can be fast although the ligand spends most of the time "hidden" in the nonactivating state. Finally, we obtain a new formula for the narrow escape time in the presence of switching.

DOI: 10.1103/PhysRevLett.103.148102 PACS numbers: 87.16.Xa

Cellular chemical reactions depend on the activation of small targets by diffusing ligands. This process can be regulated by several parameters such as the geometry of the cellular microdomain, the target shape, or the state of the target and ligand upon encounter. In the past, the activation rate and survival probability of the target were studied for diffusion-influenced chemical reactions where the reactivities of the target or the ligands stochastically fluctuate in time [1–3]. Interestingly, for slow gating dynamics it was found that both processes are not equivalent, and, for example, a gated target may lead to nonexponential behavior at short times [1,4–6]. When the target has a fluctuating potential barrier, there is an optimal combination of parameters for which the mean activation time is minimal, leading to a resonantlike phenomena [7,8]. In cellular microdomains, target activation may depend on the state of the ligand. This is, for example, the case in the cytoplasm, where enzymes can switch between an inactive and active state, or in the nucleus, where a transcription factor needs to be first activated in order to bind to a specific DNA promoter [9,10]. Recently, intermittent search scenarios were introduced and extensively analyzed for a dynamics switching stochastically between fast ballistic phases and slow diffusive phases, while the target can only be found in the diffusing phase [11–13]: interestingly, switching can decrease the search time, and there are optimal search strategies that minimizes the search time.

In the absence of switching, target activation is determined by the narrow escape time (NET), which is the mean time for a Brownian ligand to find a small target in a confined environment [14–18]. The NET has been used to compute the mean and variance of chemical reactions with few molecules diffusing in microdomains [19], to estimate the probability and the arrival time of viral particles to nuclear pores [20], or to study the early steps of phototransduction in rod photoreceptor [21].

We study in this Letter the gated narrow escape time (GNET) to exit the domain Ω through a small window, if a diffusing ligand stochastically switches between two states 1 and 2 with diffusion coefficients D_1 and D_2, and can exit only in state 1 (see Fig. 1). Switching may be due to conformational changes or chemical interactions. We estimate the GNET using new equations for the sojourn times the ligand spends in the different states. We find that switching not only affects drastically the exit time and the sojourn times, but also, only for $D_2 > D_1$ the GNET can be optimized as a function of the switching rates. In addition, a ligand may exit almost as fast as possible, although it spends most of the time in state 2. Finally, we give a new formula for the GNET in dimension three, which extends the NET formula to the switching case. We also discuss briefly possible applications in cellular signaling.

A Brownian ligand diffuses in a confined domain Ω while switching between states 1 and 2 with Poissonian rate constants k_{12} and k_{21} and diffusion constants D_1 and D_2. Upon encounter, the ligand activates the target located on the small boundary portion $\partial\Omega_a$ only in state 1, while it is reflected otherwise everywhere on the boundary. To study the GNET, we consider the mean sojourn times $u_n(x, m)$ the ligand spends in state n before exiting, conditioned on starting at position x in state m. From the backward Chapman-Kolmogorov equation [22,23] we find that the $u_n(x, m)$ satisfy the coupled system of equations

$$L_m^* u_n(x, m) - \sum_{i=1}^{2} k_{mi}(u_n(x, m) - u_n(x, i)) = -\delta_{nm}, \quad (1)$$

where L_m^* is the backward Kolmogorov operator in state m, which in our case is $L_m^* = D_m\Delta$, and we have absorbing boundary conditions on $\partial\Omega_a$ for $u_n(x, 1)$, and reflecting conditions otherwise. Equation (1) separately constitutes a closed system of equations for each value n, and it is sufficient to study the equations for $n = 1$, because the solutions $u_2(x, m)$ are obtained through the linear transformation

Maximum Likelihood and the Single Receptor

Robert G. Endres[1,2,*] and Ned S. Wingreen[3,†]

[1]Division of Molecular Biosciences, Imperial College London, London SW7 2AZ, United Kingdom
[2]Centre for Integrated Systems Biology at Imperial College, Imperial College London, London SW7 2AZ, United Kingdom
[3]Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544-1014, USA

(Received 11 June 2009; published 7 October 2009)

The accuracy by which biological cells sense chemical concentration is ultimately limited by the random arrival of particles at the receptors by diffusion. This fundamental physical limit is generally considered to be the Berg-Purcell limit [Biophys. J. **20**, 193 (1977)]. Here we derive a lower limit by applying maximum likelihood to the time series of receptor occupancy. The increased accuracy stems from solely considering the unoccupied time intervals—disregarding the occupied time intervals as these do not contain any information about the external particle concentration, and only decrease the accuracy of the concentration estimate. Receptors which minimize the bound time intervals achieve the highest possible accuracy. We discuss how a cell could implement such an optimal sensing strategy by absorbing or degrading bound particles.

DOI: 10.1103/PhysRevLett.103.158101 PACS numbers: 87.10.Mn, 87.15.kp, 87.16.dj

Single cells can sense external chemical concentrations with extremely high accuracy. For instance, the chemotactic bacterium *Escherichia coli* can detect 3.2 nM of the attractant aspartate [1], which corresponds to only about 3 attractant particles in the volume of the cell. Single eukaryotic cells such as *Dictyostelium discoideum* [2] and *Saccharomyces cerevisiae* [3] (budding yeast) are well known to measure and respond to extremely shallow gradients of chemical signals [4]. These observations raise the question how close do cells operate to the fundamental physical limit of sensing accuracy set by the random arrival of particles by diffusion at the receptors? This question was addressed in a seminal work by Berg and Purcell [5], and recently reinvestigated by Bialek and Setayeshgar [6,7]. Today, it is generally accepted that the limit derived by Berg and Purcell is a fundamental physical limit which cannot be exceeded. In this Letter, we show for a single receptor how this limit can be improved (using maximum likelihood estimation), and discuss how cells could implement this improved sensing strategy in practice.

Berg and Purcell calculated the accuracy of concentration sensing by a single receptor which binds particles of concentration c_0 with rate k_+c_0 and unbinds particles with rate k_- [see Fig. 1(a)]. Specifically, they considered a binary time series of total length T composed of bound and unbound time intervals [see Fig. 1(b)]. Berg and Purcell estimated concentration directly from the fraction of time T that a particle is bound. By considering the time correlations of particles bound to the receptor, they found the variance $(\delta c)^2$ in the estimated concentration to be [5]

$$\frac{(\delta c)^2}{c_0^2} = \frac{2\bar{\tau}_b}{T\bar{p}} = \frac{1}{2Dsc_0(1-\bar{p})T}, \qquad (1)$$

where D is the diffusion coefficient, $\bar{\tau}_b$ is the true average duration of bound intervals, s describes the receptor dimension, and $\bar{p}$ is the true equilibrium probability for the receptor to be bound. The last equality in Eq. (1) is obtained using detailed balance; i.e., at equilibrium the rate of unbinding transitions $\bar{p}/\bar{\tau}_b$ must equal the rate of binding transitions $(1-\bar{p})/\bar{\tau}_u$, where $\bar{\tau}_u$ is the average duration of unbound intervals. For diffusion-limited binding, $1/\bar{\tau}_u = 4Dsc_0$, yielding the right-hand side of Eq. (1). In the following we revisit the Berg and Purcell limit on the accuracy of concentration sensing from the perspective of maximum likelihood estimation.

Maximum likelihood estimation is a statistical method used for fitting a mathematical model to data [8]. For a fixed set of data and an underlying parametrized model, maximum likelihood picks the values of the model parameters that make the data "more likely" than they would be for any other values of the parameters. Here, the cell's best estimate of concentration can be obtained from maximum likelihood applied to the time series $\{t_+, t_-\}$ of duration T with particle binding events at times $t_{+,i}$ and unbinding events at times $t_{-,i}$ [see Fig. 1(b)]. Following Berg and Purcell, we disregard potential rebinding of previously bound particles, assuming diffusion is sufficiently fast to remove recently unbound particles from the vicinity of the

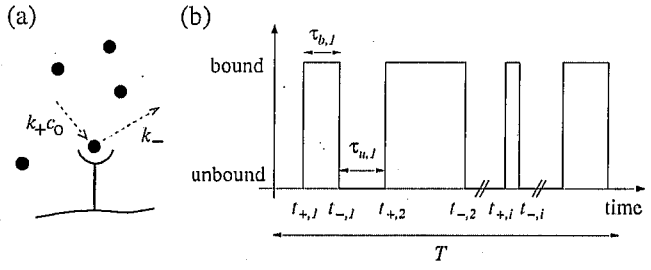

FIG. 1. Schematic of particle-receptor binding. (a) An unoccupied receptor can bind a particle with rate k_+c_0, and an occupied receptor can unbind a bound particle with rate k_-. (b) Binary time series of receptor occupancy.

Subcellular localization of a bacterial regulatory RNA

Jay H. Russell and Kenneth C. Keiler[1]

The Pennsylvania State University, Department of Biochemistry and Molecular Biology, 401 Althouse Lab, University Park, PA 16802

Edited by Robert T. Sauer, Massachusetts Institute of Technology, Cambridge, MA, and approved July 13, 2009 (received for review May 4, 2009)

Eukaryotes and bacteria regulate the activity of some proteins by localizing them to discrete subcellular structures, and eukaryotes localize some RNAs for the same purpose. To explore whether bacteria also spatially regulate RNAs, the localization of tmRNA was determined using fluorescence *in situ* hybridization. tmRNA is a small regulatory RNA that is ubiquitous in bacteria and that interacts with translating ribosomes in a reaction known as *trans*-translation. In *Caulobacter crescentus*, tmRNA was localized in a cell-cycle–dependent manner. In G_1-phase cells, tmRNA was found in regularly spaced foci indicative of a helix-like structure. After initiation of DNA replication, most of the tmRNA was degraded, and the remaining molecules were spread throughout the cytoplasm. Immunofluorescence assays showed that SmpB, a protein that binds tightly to tmRNA, was colocalized with tmRNA in the helix-like pattern. RNase R, the nuclease that degrades tmRNA, was localized in a helix-like pattern that was separate from the SmpB-tmRNA complex. These results suggest a model in which tmRNA-SmpB is localized to sequester tmRNA from RNase R, and localization might also regulate tmRNA-SmpB interactions with ribosomes.

RNase R | SmpB | tmRNA

Cells use spatial as well as temporal regulation of gene products to control normal physiological events and to respond to environmental challenges. Precise localization of proteins is used in eukaryotes and bacteria for processes such as cytoskeleton formation, cytokinesis, replication, and a wide array of signal transduction pathways (1). Eukaryotes also localize RNAs for regulatory control. For example, *Xenopus* and *Drosophila* localize mRNAs to regulate developmental timing (2, 3), and yeast localize the *ash1* mRNA to the bud tip of dividing cells to control mating type switching (4). In contrast, there have been no reports of localization of endogenous bacterial RNAs. Here, we examine the localization of a small regulatory RNA, tmRNA, in the model bacterial species *Caulobacter crescentus*.

tmRNA is a highly abundant and widespread RNA that interacts with translating ribosomes in a reaction known as *trans*-translation (5–8). The termini of tmRNA fold into a structure that mimics the acceptor stem and TΨC arm of $tRNA^{Ala}$, and tmRNA is charged with alanine by alanyl-tRNA synthetase. Instead of an anticodon arm, tmRNA contains three to four pseudoknots and a specialized ORF. During *trans*-translation, tmRNA enters translating ribosomes that are near the 3′ end of an mRNA. tmRNA displaces the mRNA, and a peptide tag encoded in the tmRNA reading frame is added to the C terminus of the nascent polypeptide. The tmRNA peptide tag is recognized by several proteases, targeting the protein for rapid degradation (9). This reaction also releases the ribosome and promotes degradation of the mRNA that was being translated. A small protein, SmpB, binds to the tRNA-like domain of tmRNA and is required for tmRNA structure, stability, and activity (10). tmRNA and SmpB have been found in every bacterial species investigated. They are essential in some bacteria, and in others they are required for virulence (11, 12), antibiotic resistance (13), symbiosis (14), and cell-cycle control (15). In *C. crescentus*, tmRNA and SmpB are required for correct timing of DNA replication and differentiation (15).

C. crescentus is a model system for studies of bacterial differentiation and cell-cycle regulation because it is easy to obtain synchronized cultures of cells in G_1 phase, called swarmer cells. Swarmer cells execute a developmental program that coordinates initiation of DNA replication with morphological differentiation into stalked cells. During differentiation, the polar flagellum is ejected and a stalk grows at the same site. Stalked cells complete S phase and divide, producing a stalked cell and a new swarmer cell. *trans*-Translation is required for correct timing of events during the developmental program. Cells lacking *ssrA* (the gene encoding tmRNA) or *smpB* do not initiate DNA replication at the correct time, and all subsequent cell-cycle–regulated events are delayed. Consistent with this phenotype, tmRNA and SmpB levels are temporally regulated as a function of the cell cycle (10, 15). tmRNA and SmpB accumulate during the G_1–S phase transition, when both molecules are synthesized and stable. Proteolysis of SmpB in early S phase allows degradation of tmRNA by the exoribonuclease RNase R, dramatically decreasing the levels of tmRNA-SmpB complex in stalked cells (10). The experiments described here show that tmRNA, SmpB, and RNase R are localized to specific sites within the cytoplasm, and suggest that spatial as well as temporal mechanisms are used to regulate *trans*-translation.

Results

tmRNA Is Localized in *C. crescentus*. Fluorescence *in situ* hybridization (FISH) has been used to observe the subcellular localization of DNA and RNA in eukaryotic cells, and several studies in bacteria have used FISH to study the localization of DNA segments (16, 17). We adapted FISH to observe RNA localization in *C. crescentus*. To probe for tmRNA, a DNA oligonucleotide probe complementary to the tmRNA reading frame sequence was synthesized and conjugated to the fluorophore Cy3 (SsrA-Cy3). Fixed *C. crescentus* cells were probed with SsrA-Cy3 to determine whether tmRNA was localized within the bacterial cytoplasm. A localized fluorescence signal was observed in 68% of cells (n = 122) (Fig. 1). These cells contained a pattern of four to six regularly spaced foci that in some cases were connected with bands or extended filaments of fluorescence signal. Of the cells, 32% had very little fluorescence signal. To ensure that the fluorescence pattern was a result of specific hybridization of SsrA-Cy3 with tmRNA, FISH experiments were repeated using cells deleted for *ssrA*, the gene encoding tmRNA. The fluorescence signal after probing Δ*ssrA* cells with SsrA-Cy3 was indistinguishable from mock FISH experiments using wild-type or Δ*ssrA* cells in which no SsrA-Cy3 was added, indicating that SsrA-Cy3 specifically recognized tmRNA. These results suggest that when tmRNA is expressed in *C. crescentus*, it is localized in a discrete pattern in the cytoplasm.

The pattern of tmRNA localization is similar to that observed for proteins that form helical structures in *C. crescentus* and other bacteria. For example, MreB in *C. crescentus* (18), *Esch-*

Author contributions: J.H.R. and K.C.K. designed research; J.H.R. performed research; J.H.R. contributed new reagents/analytic tools; J.H.R. and K.C.K. analyzed data; and J.H.R. and K.C.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: kkeiler@psu.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0904904106/DCSupplemental.

MICROBIOLOGY

Paper-supported 3D cell culture for tissue-based bioassays

Ratmir Derda[a,b], Anna Laromaine[a], Akiko Mammoto[c], Sindy K. Y. Tang[a], Tadanori Mammoto[c], Donald E. Ingber[b,c,d], and George M. Whitesides[a,b,1]

[a]Department of Chemistry and Chemical Biology, [b]Wyss Institute for Biologically Inspired Engineering, and [d]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138; and [c]Vascular Biology Program, Departments of Pathology and Surgery, Children's Hospital and Harvard Medical School, Boston, MA 02115

Contributed by George M. Whitesides, September 17, 2009 (sent for review August 19, 2009)

Fundamental investigations of human biology, and the development of therapeutics, commonly rely on 2D cell-culture systems that do not accurately recapitulate the structure, function, or physiology of living tissues. Systems for 3D cultures exist but do not replicate the spatial distributions of oxygen, metabolites, and signaling molecules found in tissues. Microfabrication can create architecturally complex scaffolds for 3D cell cultures that circumvent some of these limitations; unfortunately, these approaches require instrumentation not commonly available in biology laboratories. Here we report that stacking and destacking layers of paper impregnated with suspensions of cells in extracellular matrix hydrogel makes it possible to control oxygen and nutrient gradients in 3D and to analyze molecular and genetic responses. Stacking assembles the "tissue", whereas destacking disassembles it, and allows its analysis. Breast cancer cells cultured within stacks of layered paper recapitulate behaviors observed both in 3D tumor spheroids in vitro and in tumors in vivo: Proliferating cells in the stacks localize in an outer layer a few hundreds of microns thick, and growth-arrested, apoptotic, and necrotic cells concentrate in the hypoxic core where hypoxia-sensitive genes are overexpressed. Altering gas permeability at the ends of stacks controlled the gradient in the concentration of the O_2 and was sufficient by itself to determine the distribution of viable cells in 3D. Cell cultures in stacked, paper-supported gels offer a uniquely flexible approach to study cell responses to 3D molecular gradients and to mimic tissue- and organ-level functions.

diffusion | hypoxia | molecular gradients | multilayer constructs

A distinguishing feature of cell growth in 3D is the diffusion-limited distribution of oxygen, nutrients, metabolites, and signaling molecules, which are continuously consumed and produced by the cells (1, 2). These distributions are not mimicked in 2D monolayer cultures (3–5). Chemical gradients influence embryo development, tissue homeostasis, and pathologies ranging from cancer to infectious disease. The mass transport of oxygen and nutrients is particularly important because it limits cell proliferation in 3D cultures to distances less than a few hundred microns from convectively stirred, oxygenated medium (2, 6) or in living tissues to similar distance from capillary blood vessels (7, 8). Validating mechanisms for regulation of cell growth in 3D requires culture systems in which gradients can be designed, generated, and tested (6). Chemical gradients in tissues are often generated by cellular metabolism. Hence, controlling the distribution of oxygen and nutrients in 3D cultures is only possible if the location and properties of cells can be manipulated on a scale similar to the characteristic length that balances diffusion and metabolism (hundreds of microns) in tissue. Methods that enable generation and analysis of such distributions in vivo (e.g., tissue-specific gene expression, laser microdissection) or in vitro [e.g., studies in microfabricated devices (9–12)] exist, but are time consuming to use and require specialized engineering approaches or instrumentation not available in most biology laboratories.

In this report, we present a strategy for controlling the distribution of cultured cells in 3D by fabrication of multilaminate structures of fiber-supported hydrogels with each layer composed of chromatography paper impregnated with an extracellular matrix (ECM) hydrogel containing living cells. This technique starts by adding a hydrogel precursor, as a fluid containing suspended cells, to a paper support and gelling it in place. This simple procedure generates 3D fiber-supported gels with defined numbers and densities of cells with easily controlled lateral dimensions and with thickness defined by the paper. Importantly, a single layer (SL) of paper impregnated with the hydrogel containing cells is sufficiently strong mechanically that it can be easily manipulated and stacked into multiple-layer constructs to generate 3D cultures. These constructs control the properties of 3D cultures of cells at the $n \times 100$-μm length scale that is relevant to mass transport in vivo. We demonstrate that these constructs may be designed to exhibit desired spatial distributions of cells, and they can be easily destacked to analyze cell structure and function within physical sections of these 3D tissues structures without requiring optical or histological sectioning.

Results

Analysis of Cells Inside Paper Permeated with Hydrogels. When a suspension of cells in a hydrogel precursor (liquid) is placed in contact with dry paper, the fluid containing the cell redistributes by capillary wicking, and the cells are carried with the flow of liquid. To analyze geometry of the 3D culture generated in this process, we spotted suspension of HS-5 cells stably expressing mCherry fluorescent protein in Alexa Fluor 488-labeled Matrigel on 200 μm-thick chromatography paper (Fig. 1*A*). The use of solutions containing different concentrations of cells (1.3×10^6–4×10^7 cells/mL) made the variance of concentration of cells in 3D cultures straightforward. Spotting different volumes of cell suspension onto the paper (1–5 μL) changed the area of these 3D cultures (Fig. 1 *B–D*). We incubated the samples in growth media for four hours, fixed them, and stained them with Alexa Fluor 633-conjugated phalloidin to label intracellular F-actin. A fluorescent gel scanner showed the distribution of cells (Fig. 1*B*) and matrix (Fig. 1*C*) in the resulting constructs; as expected, the area of a 3D culture that contains cells within a paper is similar to the area of spreading of liquid hydrogel precursor (Fig. 1*E*).

To quantify the number of cells inside the paper, in the images obtained by fluorescent gel scanner we defined the background intensity as that of the paper that did not contain any cells and

Author contributions: R.D., A.L., A.M., D.E.I., and G.M.W. designed research; R.D., A.L., A.M., and T.M. performed research; S.K.Y.T. contributed new reagents/analytic tools; R.D., A.L., A.M., and T.M. analyzed data; and R.D., D.E.I., and G.M.W. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: gwhitesides@gmwgroup.harvard.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0910666106/DCSupplemental.