doi:10.1038/nature08187

LETTERS

Programming cells by multiplex genome engineering and accelerated evolution

Harris H. Wang[1,2,3]*, Farren J. Isaacs[1]*, Peter A. Carr[4,5], Zachary Z. Sun[6], George Xu[6], Craig R. Forest[7] & George M. Church[1]

The breadth of genomic diversity found among organisms in nature allows populations to adapt to diverse environments[1,2]. However, genomic diversity is difficult to generate in the laboratory and new phenotypes do not easily arise on practical timescales[3]. Although *in vitro* and directed evolution methods[4–9] have created genetic variants with usefully altered phenotypes, these methods are limited to laborious and serial manipulation of single genes and are not used for parallel and continuous directed evolution of gene networks or genomes. Here, we describe multiplex automated genome engineering (MAGE) for large-scale programming and evolution of cells. MAGE simultaneously targets many locations on the chromosome for modification in a single cell or across a population of cells, thus producing combinatorial genomic diversity. Because the process is cyclical and scalable, we constructed prototype devices that automate the MAGE technology to facilitate rapid and continuous generation of a diverse set of genetic changes (mismatches, insertions, deletions). We applied MAGE to optimize the 1-deoxy-D-xylulose-5-phosphate (DXP) biosynthesis pathway in *Escherichia coli* to overproduce the industrially important isoprenoid lycopene. Twenty-four genetic components in the DXP pathway were modified simultaneously using a complex pool of synthetic DNA, creating over 4.3 billion combinatorial genomic variants per day. We isolated variants with more than fivefold increase in lycopene production within 3 days, a significant improvement over existing metabolic engineering techniques. Our multiplex approach embraces engineering in the context of evolution by expediting the design and evolution of organisms with new and improved properties.

With the advent of next-generation fluorescent DNA sequencing[10], our ability to sequence genomes has greatly outpaced our ability to modify genomes. Existing cloning-based technologies are confined to serial and inefficient introduction of single DNA constructs into cells, requiring laborious and outdated genetic engineering techniques. Whereas *in vivo* methods such as recombination-based genetic engineering (recombineering) have enabled efficient modification of single genetic targets using single-stranded DNA (ssDNA)[11–14], no such attempts have been made to modify genomes on a large and parallel scale. MAGE provides a highly efficient, inexpensive and automated solution to simultaneously modify many genomic locations (for example, genes, regulatory regions) across different length scales, from the nucleotide to the genome level (Fig. 1).

Efficiency of the MAGE process was characterized using a modified *E. coli* strain (EcNR2). Mediated by the bacteriophage λ-Red ssDNA-binding protein β, allelic replacement is achieved in EcNR2 by directing ssDNA or oligonucleotides (oligos) to the lagging strand of the replication fork during DNA replication[14]. We optimized a number of parameters (see Supplementary Information, Supplementary Fig. 2 and Supplementary Table 1) to maximize efficiency of oligo-mediated allelic replacement. To generate sequence diversity in any region of the chromosome by allelic replacement, a pool of targeting oligos is repeatedly introduced into a cell. Under optimized conditions, we can successfully introduce new genetic modifications in >30% of the cell population (Supplementary Fig. 2d) every 2–2.5 h.

Oligo-mediated allelic replacement is capable of introducing a variety of genetic modifications at high efficiency. The efficiency of generating a mismatch or insertion modification is correlated to the amount of homologous sequence between the oligo and its chromosomal target (Fig. 2a, b); the efficiency of producing a deletion modification is correlated to the size of the deletion (Fig. 2c). Figure 2d shows that the predicted two-state hybridization free energy ΔG (ref. 15) between the oligo and target chromosomal sequence is a predictor of the allelic replacement efficiency. Thus, in a pool of oligos with degenerate sequences, oligos with more homology to the target will be incorporated in the chromosome at a higher frequency than those with less homology. This feature of MAGE enables tunable generation of divergent sequences along favourable evolutionary paths.

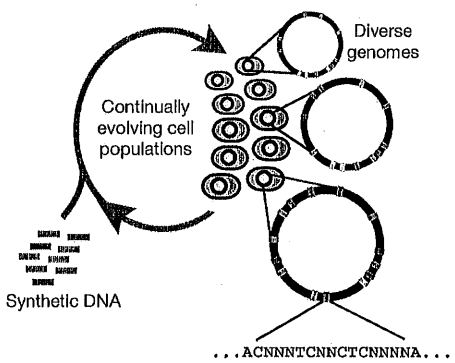


Figure 1 | Multiplex automated genome engineering enables the rapid and continuous generation of sequence diversity at many targeted chromosomal locations across a large population of cells through the repeated introduction of synthetic DNA. Each cell contains a different set of mutations, producing a heterogeneous population of rich diversity (denoted by distinct chromosomes in different cells). Degenerate oligo pools that target specific genomic positions enable the generation of a diverse set of sequences at each chromosomal location.

[1]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [2]Program in Biophysics, Harvard University, Cambridge, Massachusetts 02138, USA. [3]Program in Medical Engineering Medical Physics, Harvard-MIT Division of Health Sciences and Technology, [4]The Center for Bits and Atoms, [5]Media Lab, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [6]Harvard College, Cambridge, Massachusetts 02138, USA. [7]George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA.
*These authors contributed equally to this work.

Cell

A Synthetic Genetic Edge Detection Program

Jeffrey J. Tabor,[1] Howard M. Salis,[1] Zachary Booth Simpson,[2,3] Aaron A. Chevalier,[2,3] Anselm Levskaya,[1] Edward M. Marcotte,[2,3,4] Christopher A. Voigt,[1,*] and Andrew D. Ellington[2,3,4]
[1]Department of Pharmaceutical Chemistry, School of Pharmacy, University of California San Francisco, San Francisco, CA 94158, USA
[2]Center for Systems and Synthetic Biology
[3]Institute for Cellular and Molecular Biology
[4]Department of Chemistry and Biochemistry
University of Texas, Austin, TX 78712, USA
*Correspondence: cavoigt@picasso.ucsf.edu
DOI 10.1016/j.cell.2009.04.048

SUMMARY

Edge detection is a signal processing algorithm common in artificial intelligence and image recognition programs. We have constructed a genetically encoded edge detection algorithm that programs an isogenic community of *E. coli* to sense an image of light, communicate to identify the light-dark edges, and visually present the result of the computation. The algorithm is implemented using multiple genetic circuits. An engineered light sensor enables cells to distinguish between light and dark regions. In the dark, cells produce a diffusible chemical signal that diffuses into light regions. Genetic logic gates are used so that only cells that sense light and the diffusible signal produce a positive output. A mathematical model constructed from first principles and parameterized with experimental measurements of the component circuits predicts the performance of the complete program. Quantitatively accurate models will facilitate the engineering of more complex biological behaviors and inform bottom-up studies of natural genetic regulatory networks.

INTRODUCTION

Living cells can be programmed with genetic parts, such as promoters, transcription factors and metabolic genes (Andrianantoandro et al., 2006; Benner and Sismour, 2005; Canton et al., 2008; Endy, 2005; Haseltine and Arnold, 2007). These parts can be combined to construct genetic versions of electronic circuits, including switches (Atkinson et al., 2003; Gardner et al., 2000; Kramer and Fussenegger, 2005; Kramer et al., 2004), logic (Anderson et al., 2007; Guet et al., 2002; Rackham and Chin, 2005), memory (Ajo-Franklin et al., 2007; Gardner et al., 2000; Ham et al., 2006), pulse generators (Basu et al., 2004), and oscillators (Atkinson et al., 2003; Elowitz and Leibler, 2000; Fung et al., 2005; Stricker et al., 2008; Tigges et al., 2009). The current challenge is to assemble multiple genetic circuits into larger programs for the engineering of more sophisticated behaviors (Purnick and Weiss, 2009).

The characterization of transfer functions or the quantitative relationship between circuit input(s) and output(s) (Bintu et al., 2005a; Tabor et al., 2009; Voigt, 2006; Weiss et al., 1999) will aid the development of accurate mathematical models (Ajo-Franklin et al., 2007; Guido et al., 2006) which will allow complex genetic programs to be examined in silico prior to physical construction. Predictive models for the design of genetic programs will drive applications in biotechnology and aid bottom-up studies of natural regulatory systems.

Edge detection is a well-studied computational problem used to determine the boundaries of objects within an image (Suel et al., 2000). This process reduces the information content in a complex image and is used in applications ranging from retinal preprocessing (Maturana and Frenk, 1963) to the analysis of microarray data (Kim et al., 2001). For a digital black and white image, a typical algorithm operates by scanning for a white pixel and then comparing the intensity of that pixel to its eight neighboring pixels. If any of the neighbors is black, the algorithm classifies those pixels as being part of an edge. The serial nature of this search process results in a computation time that increases linearly with the number of pixels in the image. We aimed to implement a parallel edge detection algorithm wherein each bacterium within a spatially distributed population functions as an independent signal processor. In this design, each bacterium (up to 10^9 individuals for a 90 mm Petri dish image) processes a small amount of local information simultaneously, and the population cooperates to find the edges.

RESULTS

A Genetic Program for Edge Detection

The genetic edge detection algorithm programs a lawn of bacteria to identify the light-dark boundaries within a projected image of light (Figures 1A and 1B). To accomplish this, each bacterium in the population executes the following pseudocode (Figure 1C): IF NOT light, produce signal. IF signal AND NOT (NOT light), produce pigment.

The "produce signal" and "produce pigment" functions make the cell generate a diffusible communication signal and a black pigment, respectively. The conversion of this pseudocode into a molecular genetic system is shown in Figure 1D.

When cells sense that they are in the dark, they produce the diffusible signal (Figure 1B). Cells that sense light do not make the signal, but are allowed to respond to it. Thus only those cells that are in the light but proximal to dark areas activate the output which results in the enzymatic production of a black pigment. The biological edge detection algorithm requires: (1) a dark

Cell

A Systems-Level Analysis of Perfect Adaptation in Yeast Osmoregulation

Dale Muzzey,[1,4,5] Carlos A. Gómez-Uribe,[1,2,5] Jerome T. Mettetal,[1] and Alexander van Oudenaarden[1,3,*]
[1]Department of Physics
[2]Harvard-MIT Division of Health Sciences and Technology
[3]Department of Biology
Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[4]Harvard University Graduate Biophysics Program, Harvard Medical School, Boston, MA 02115, USA
[5]These authors contributed equally to this work
*Correspondence: avano@mit.edu
DOI 10.1016/j.cell.2009.04.047

SUMMARY

Negative feedback can serve many different cellular functions, including noise reduction in transcriptional networks and the creation of circadian oscillations. However, only one special type of negative feedback ("integral feedback") ensures perfect adaptation, where steady-state output is independent of steady-state input. Here we quantitatively measure single-cell dynamics in the *Saccharomyces cerevisiae* hyperosmotic shock network, which regulates membrane turgor pressure. Importantly, we find that the nuclear enrichment of the MAP kinase Hog1 perfectly adapts to changes in external osmolarity, a feature robust to signaling fidelity and operating with very low noise. By monitoring multiple system quantities (e.g., cell volume, Hog1, glycerol) and using varied input waveforms (e.g., steps and ramps), we assess in a minimally invasive manner the network location of the mechanism responsible for perfect adaptation. We conclude that the system contains only one effective integrating mechanism, which requires Hog1 kinase activity and regulates glycerol synthesis but not leakage.

INTRODUCTION

Positive and negative feedback loops are ubiquitous regulatory features of biological systems in which the system output reinforces or opposes the system input, respectively. Quantitative models are increasingly being used to study the function and dynamic properties of complicated, feedback-laden biological systems. These models can be broadly classified by the extent to which they represent specific molecular details of the network. At one extreme are the exhaustive models that dynamically track quantities of virtually all biomolecules in a system, often using differential equations based on either known or assumed reaction stoichiometries and rates. At the other end of the modeling spectrum is the minimalist approach, which aims to fit and predict a system's input-output dynamics with only a few key parameters, each potentially the distillation of a large group of reactions.

Examples of exhaustive and minimalist modeling approaches illustrate their unique advantages and disadvantages. For instance, an exhaustive model of EGF-receptor regulation (Schoeberl et al., 2002) can impressively predict the dynamics of 94 specific network elements, but it requires nearly 100 parameters—some of which are not easily measured biochemically—and could suffer from the omission of important reactions not yet biologically identified. By contrast, minimalist models frequently lack such potentially desirable reaction- and network-specific details, yet they excel at providing intuitive and general insights into the dynamic properties of recurrent system architectures. For instance, two elegant studies of bacterial chemotaxis—a system said to "perfectly adapt" because abrupt changes in the amount of ligand only transiently affect the tumbling frequency, whereas steady-state tumbling is notably independent of the ligand concentration—highlighted a general feature of all perfectly adapting systems (Barkai and Leibler, 1997; Yi et al., 2000). Specifically, it was shown that a negative feedback loop implementing "integral feedback" is both necessary and sufficient for robust perfect adaptation in any biological system (Yi et al., 2000). Mathematically, a dynamic variable (e.g., x or *[cyclin-B]*) is an "integrator" if its rate of change is independent of the variable itself (e.g., if the dx/dt and $d[\text{cyclin-B}]/dt$ equations contain no terms involving x or *[cyclin-B]*, respectively), and integral feedback describes a negative feedback loop that contains at least one integrator. Biologically, a biomolecule acts as an integrator if its rate equation is not a function of the biomolecule concentration itself; such a situation arises if, say, the synthesis and degradation reactions are saturated (Supplemental Data available online). By providing specific mechanistic constraints that apply to any perfectly adapting system, these two studies underscored the function and significance of perfect adaptation in homeostatic regulation and demonstrated the power of the minimalist modeling approach.

Both exhaustive and minimalist modeling tactics have been successfully applied to the osmosensing network in the budding yeast *Saccharomyces cerevisiae* (Klipp et al., 2005; Mettetal et al., 2008). The core of this network is a highly conserved mitogen-activated protein kinase (MAPK) cascade, one of several such cascades in yeast that regulate processes ranging from mating to invasive growth while being remarkably robust to crosstalk despite their many shared components (Hohmann, 2002; Schwartz and Madhani, 2004). Yeast cells maintain an

Developmental Cell
Article

Effects of Age on Meiosis in Budding Yeast

Monica Boselli,[1,2] Jeremy Rock,[1] Elçin Ünal,[1] Stuart S. Levine,[3] and Angelika Amon[1,*]
[1]David H. Koch Institute for Integrative Cancer Research and Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2]Department of Biomolecular and Genetics Sciences, University of Milan, Via Celoria 26, Milan 20133, Italy
[3]BioMicroCenter, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
*Correspondence: angelika@mit.edu
DOI 10.1016/j.devcel.2009.05.013

SUMMARY

In humans, the frequency with which meiotic chromosome mis-segregation occurs increases with age. Whether age-dependent meiotic defects occur in other organisms is unknown. Here, we examine the effects of replicative aging on meiosis in budding yeast. We find that aged mother cells show a decreased ability to initiate the meiotic program and fail to express the meiotic inducer *IME1*. The few aged mother cells that do enter meiosis complete this developmental program but exhibit defects in meiotic chromosome segregation and spore formation. Furthermore, we find that mutations that extend replicative life span also extend the sexual reproductive life span. Our results indicate that in budding yeast, the ability to initiate and complete the meiotic program as well as the fidelity of meiotic chromosome segregation decrease with cellular age and are controlled by the same pathways that govern aging of asexually reproducing yeast cells.

INTRODUCTION

During vegetative growth, budding yeast reproduces by asymmetric division, giving rise to a daughter cell, which first emerges from the mother as an outgrowth, known as a bud. The number of daughter cells a mother cell generates before its death is known as its replicative life span and is relatively uniform for a given strain (Sinclair et al., 1998a, 1998b).

With the progression of replicative life, mitotic yeast cells show morphological and physiological changes, such as an increase in cell size, genome instability, sterility, extension in cell cycle duration, nucleolar fragmentation, accumulation of ROS (reactive oxygen species), and loss of membrane turgescence. Several factors contribute to these changes in the cell physiology associated with senescence: damaged proteins, damaged mitochondria, redistribution of the silencing regulator proteins Sir2, Sir3, and Sir4, and the accumulation of ERCs (extrachromosomal ribosomal DNA circles). ERCs are formed through homologous recombination and are self-replicating, but they lack centromeres and therefore remain in the mother cell during cell division. Interestingly, damaged proteins or damaged mitochondria are also preferentially retained in the mother cell during mitosis. These observations have led to the suggestion that it is the accumulation of ERCs and damaged proteins and organelles that lead to cell death in aged cells (Steinkraus et al., 2008).

Several genes have been implicated in regulating replicative life span in yeast. The histone deacetylase Sir2 modulates aging in virtually every organism. In budding yeast, deletion of *SIR2* severely shortens life span, whereas the presence of an additional copy of this gene increases the life span by approximately 50% (Kaeberlein et al., 1999). *SIR2* is thought to affect aging by preventing ERC formation through its role in repressing mitotic recombination within the rDNA locus. The rDNA localized Fob1 protein also controls life span in yeast. Fob1 prevents movement of DNA polymerases against the direction of rRNA transcription, thereby reducing recombination within the rDNA (Defossez et al., 1999).

Nutrient signaling also regulates life span. The TOR (target of rapamycin) kinases are highly conserved and promote cell growth in response to favorable nutrient conditions and growth factor signals. Budding yeast contains two TOR kinases, Tor1 and Tor2. *TOR2* is essential, but deletion of *TOR1* has been shown to increase replicative life span by approximately 20% (Kaeberlein et al., 2005). It has been proposed that it is through *TOR1* that caloric restriction delays aging and leads to an increase in life span in budding yeast. The protein kinase Sch9 is phosphorylated by Tor1 and thought to convey some of Tor1's growth-promoting functions (Urban et al., 2007). Consistent with the roles of caloric restriction and the TOR pathway in the regulation of replicative life span, cells lacking *SCH9* live longer (Kaeberlein et al., 2005).

The effects of age on entry into and progression through meiosis are largely unexplored. In humans, meiotic chromosome segregation errors increase with maternal age (reviewed in Hassold and Hunt, 2001). Approximately 80% of these segregation errors occur during meiosis I, and 20% result from meiosis II nondisjunction (Sherman et al., 2005). Studies on chromosome 21 nondisjunction show that only 6%–10% of all trisomy 21 cases are due to errors in spermatogenesis, but meiosis I and meiosis II errors contribute equally to these male germline nondisjunction events (Sherman et al., 2005). Additionally, there is also evidence to suggest that sperm quality decreases with age (Malaspina et al., 2001; Wyrobek et al., 2006). A gradual increase in DNA damage or a reduced ability to protect germ line stem cells from free radicals has been suggested to be the basis for this decrease in sperm quality (Zhu et al., 2007). However, how replicative age affects the meiotic divisions has not been studied in detail in any organism. The ability to isolate aged yeast cells (Smeal et al., 1996) and to induce them to undergo meiosis (Honigberg and Purnapatre, 2003) enabled us to address this question.

Leading Edge
Review

Cell

Evolution of Transcriptional Regulatory Circuits in Bacteria

J. Christian Perez[1,2] and Eduardo A. Groisman[1,*]
[1]Department of Molecular Microbiology, Howard Hughes Medical Institute, Washington University School of Medicine, Campus Box 8230, 660 S. Euclid Avenue, St. Louis, MO 63110, USA
[2]Present address: Department of Microbiology and Immunology, University of California, San Francisco, 600 16th Street, Genentech Hall, Room N374, San Francisco, CA 94143-2200, USA
*Correspondence:groisman@borcim.wustl.edu
DOI 10.1016/j.cell.2009.07.002

Related organisms typically respond to a given cue by altering the level or activity of orthologous transcription factors, which, paradoxically, often regulate expression of distinct gene sets. Although promoter rewiring of shared genes is primarily responsible for regulatory differences among related eukaryotic species, in bacteria, species-specific genes are often controlled by ancestral transcription factors, and regulatory circuit evolution has been further shaped by horizontal gene transfer. Modifications in transcription factors and in promoter structure also contribute to divergence in bacterial regulatory circuits.

Introduction

Free-living organisms typically respond to a change in their surroundings or in cellular components by modifying the expression of multiple genes. In addition to sensors that detect chemical or physical cues and signaling molecules that transduce these stimuli within a cell, the responses to such changes often rely on DNA-binding proteins that interact with specific DNA sequences in promoters to activate or repress gene transcription. Thus, the regulatory circuit defined by the "wiring" between regulatory proteins and target genes determines the repertoire of gene products that an organism synthesizes upon encountering a particular signal or experiencing a developmental cue.

Related species usually rely upon orthologous regulatory systems to orchestrate responses to a given signal. In certain circumstances, the elicited responses are largely similar across species, indicative that orthologous regulatory systems control common cellular functions across species even if the species occupy different niches. In other circumstances, the responses are distinct, either in qualitative or quantitative terms, suggesting that the regulatory systems adopted by individual species are suited to particular habitats and lifestyles.

The different responses that orthologous regulatory systems can elicit when experiencing a given signal indicate that transcription circuits experience modifications in the interactions between regulators and their targets. These modifications may result in abilities that enable organisms to occupy new niches, thus contributing to the phenotypic diversity that exists among related species (McAdams et al., 2004). Yet, the observed rewiring of regulatory circuits does not necessarily result from adaptive processes, even when it causes significant changes in gene expression outputs (Lynch, 2007).

Until very recently, the knowledge of transcription regulatory circuits was limited to a few model organisms belonging to phylogenetically distant groups and sharing relatively few genes. This prevented the comparative analyses of orthologous regulatory circuitries across closely related organisms. Thus, the extent of modifications undergone by regulatory circuitries remained largely unknown. However, the availability of an increasing number of genome sequences, the use of genome-wide computational and experimental approaches to uncover entire sets of regulatory interactions in multiple species, and the engineering of organisms harboring the regulatory architecture from a related species have made it possible to study empirically the patterns of evolution of regulatory circuits. These studies have revealed that differences in regulatory circuitry can play a significant role in the morphological and developmental evolution in animals (Carroll, 2005, 2008; Davidson, 2006), are responsible for the distinct expression of antibiotic resistance determinants in bacteria (Kato et al., 2007; Winfield and Groisman, 2004; Winfield et al., 2005), and may direct the colonization of new niches in unicellular eukaryotic organisms (Borneman et al., 2007; Tuch et al., 2008a). Therefore, tinkering with transcription factors, promoter sequences, and circuit architecture, which are often referred to as "the regulatory genome" (Carroll, 2005, 2008; Davidson, 2006), has given rise to a variety of traits both in bacteria and eukaryotes.

The investigation of bacterial regulatory circuits has focused on a relative small number of extant species (consider that most bacterial species cannot be cultured in the laboratory). Yet, these investigations suggest that the evolution of regulatory circuits in bacteria proceeds in a different manner from what has been described thus far in eukaryotes (Carroll, 2005, 2008; Davidson, 2006; Tuch et al., 2008b). The reasons for the differences are the following: First, unlike closely related eukaryotic organisms, closely related bacterial species exhibit significant differences in gene content due to the pervasiveness of horizontal gene transfer. This means that the spectrum of targets controlled by orthologous transcription factors can be quite different among related bacteria. Moreover, it can cre-

Leading Edge
Minireview

Cell

Getting to First Base in Proteasome Assembly

Henrike C. Besche,[1] Andreas Peth,[1] and Alfred L. Goldberg[1,*]
[1]Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA
*Correspondence: alfred_goldberg@hms.harvard.edu
DOI 10.1016/j.cell.2009.06.035

Assembly of complex structures such as the eukaryotic 26S proteasome requires intricate mechanisms that ensure precise subunit arrangements. Recent studies have shed light on the pathway for ordered assembly of the base of the 19S regulatory particle of the 26S proteasome by identifying new precursor complexes and four dedicated chaperones involved in its assembly.

The formation of biological structures—from multimeric enzymes to large molecular machines—relies primarily on the assembly of protein subunits with complementary surfaces. When a structure is composed of identical or a few subunits, self-assembly can drive the formation of large structures such as polyhedral viruses. However, when a structure consists of many diverse components that must associate in a definite order and can have potentially detrimental nonproductive interactions, specific assembly factors are required for its formation. These factors act as process-specific molecular chaperones that prevent incorrect subunit associations. Recent studies (Funakoshi et al., 2009; Kaneko et al., 2009; Park et al., 2009; Roelofs et al., 2009; Saeki et al., 2009) have revealed an ordered pathway to assemble the 19S component of the 26S proteasome, the primary site for protein degradation in eukaryotic cells. In this assembly pathway, four chaperones ensure efficient formation of this complex structure.

The 26S proteasome is a 2.5 MDa proteolytic machine composed of 33 distinct subunits that are highly conserved among eukaryotes (Figure 1). Its primary function is to rapidly degrade proteins marked for destruction by ubiquitination. Consequently, the proteasome has many essential homeostatic functions, including protecting against the accumulation of misfolded polypeptides and controlling diverse processes through

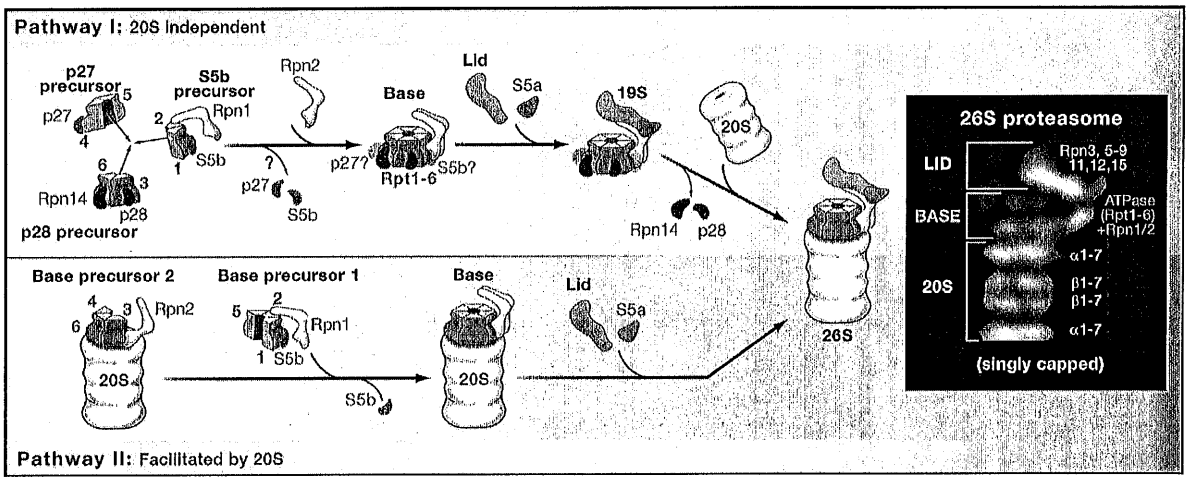


Figure 1. Pathways of Proteasome Assembly
The eukaryotic 26S proteasome consists of base, lid, and 20S particles (inset, electron micrographs of structures and subunit composition). The 20S particle is composed of 14 different subunits, and its assembly requires five cofactors (not shown). The proteasome base contains six homologous ATPase subunits (Rpt1–6) that assemble into a hexameric ring with Rpn1 and Rpn2. Base assembly is mediated by four chaperones called p27, S5b, p28, and Rpn14 in mammals and Nas2, Hsm3, Nas6 and Rpn14 in yeast (the mammalian nomenclature is used in the figure). Each chaperone interacts with one or two Rpt subunits (indicated as 1–6) and guides their assembly into the base. The exact sequence of events during assembly is unclear. In one possible pathway (pathway I) favored by Funakoshi et al. (2009), Kaneko et al. (2009), and Murata et al. (2009), chaperone-bound subcomplexes coalesce to form a hexameric base that includes Rpn1, Rpn2, Rpn14, and p28. In this complex, the S5b precursor likely releases S5b, and Rpn2 displaces p27. The lid assembles independently by an unknown mechanism and completes the 19S along with S5a (Rpn10 in yeast). The 20S particle then displaces p28 and Rpn14 to form the active 26S proteasome. In an alternative pathway (pathway II) favored by Park et al. (2009) and Roelofs et al. (2009), the 20S particle serves as a template on which Rpn2, Rpt4, Rpt6, and Rpt3 (base precursor 2) assemble with the assistance of p28 and Rpn14 to form a subcomplex. This subcomplex is then joined by the S5b precursor and Rpt5 (base precursor 1) as well as the lid and S5a to form the 26S proteasome.

Cell

Theory

Intrinsic Protein Disorder and Interaction Promiscuity Are Widely Associated with Dosage Sensitivity

Tanya Vavouri,[1] Jennifer I. Semple,[1] Rosa Garcia-Verdugo,[1] and Ben Lehner[1,2,*]
[1]EMBL-CRG Systems Biology Unit
[2]ICREA
Centre for Genomic Regulation, UPF, Dr. Aiguader 88, Barcelona 08003, Spain
*Correspondence: ben.lehner@crg.es
DOI 10.1016/j.cell.2009.04.029

SUMMARY

Why are genes harmful when they are overexpressed? By testing possible causes of overexpression phenotypes in yeast, we identify intrinsic protein disorder as an important determinant of dosage sensitivity. Disordered regions are prone to make promiscuous molecular interactions when their concentration is increased, and we demonstrate that this is the likely cause of pathology when genes are overexpressed. We validate our findings in two animals, *Drosophila melanogaster* and *Caenorhabditis elegans*. In mice and humans the same properties are strongly associated with dosage-sensitive oncogenes, such that mass-action-driven molecular interactions may be a frequent cause of cancer. Dosage-sensitive genes are tightly regulated at the transcriptional, RNA, and protein levels, which may serve to prevent harmful increases in protein concentration under physiological conditions. Mass-action-driven interaction promiscuity is a single theoretical framework that can be used to understand, predict, and possibly treat the effects of increased gene expression in evolution and disease.

INTRODUCTION

Most of the genetic variation between any two individuals or species consists of regulatory or copy number variants that alter gene expression rather than coding sequence (Stranger et al., 2007). Despite the importance of altered gene expression to disease and evolution, it is not understood why only certain genes are pathological when their expression is increased (are dosage sensitive), and what the molecular mechanisms are that drive these phenotypic changes (Semple et al., 2008). Indeed there are no known molecular mechanisms that are predictive of dosage sensitivity across the genome of an organism (Gelperin et al., 2005; Semple et al., 2008; Sopko et al., 2006). As a result, it is currently very difficult to understand the consequences of increased gene expression in either disease or evolution.

In yeast, ~80% of genes can be constitutively overexpressed without any severe detrimental effect on growth (Gelperin et al., 2005; Sopko et al., 2006). In contrast a subset of genes are harmful when overexpressed. These dosage-sensitive genes are enriched for diverse and multiple functions (Gelperin et al., 2005; Sopko et al., 2006), and they do not significantly overlap the set of genes that are harmful when their expression is decreased (Deutschbauer et al., 2005; Semple et al., 2008). Unlike essential genes, dosage-sensitive genes are not enriched among the subunits of protein complexes (Sopko et al., 2006). Moreover, whereas the loss-of-function phenotype of one subunit of a protein complex is highly predictive of the loss-of-function phenotype of the other subunits (Fraser and Plotkin, 2007; Hart et al., 2007), this is not true for overexpression phenotypes (Semple et al., 2008). Indeed, in the majority of cases examined overexpression causes phenotypic effects that are different from underexpression (Niu et al., 2008; Sopko et al., 2006). It has also been shown that dosage-sensitive genes are only very weakly enriched for cell-cycle-regulated genes (Sopko et al., 2006), so forced expression of periodically expressed genes cannot be a major cause of phenotypic change. In short, it is not understood why cells function robustly following the overexpression of most genes but are very sensitive to increases in the levels of a subset of genes. It is also not clear what the most important molecular mechanisms are that cause gain-of-function phenotypes following gene overexpression.

To resolve this, we systematically tested possible causes of dosage sensitivity in yeast. We find that the intrinsic disorder content of a protein is an important determinant of dosage sensitivity. These disordered regions are prone to make promiscuous molecular interactions when their concentration is increased, and we present evidence that this is a frequent cause of dosage sensitivity. We confirm our findings in two animals, *Drosophila melanogaster* and *Caenorhabditis elegans*, and we show that the properties of dosage-sensitive genes detected in model organisms are also strongly associated with dosage-sensitive oncogenes in mice and humans. Finally, we show that dosage-sensitive genes are tightly regulated at the transcriptional, RNA, and protein levels, and we argue that this control acts to prevent potentially harmful increases in protein concentration under physiological conditions. The interaction promiscuity theory yields predictions for future experimental studies and

Self-Organization of the *Escherichia coli* Chemotaxis Network Imaged with Super-Resolution Light Microscopy

Derek Greenfield[1,2☯], Ann L. McEvoy[1☯], Hari Shroff[3], Gavin E. Crooks[2], Ned S. Wingreen[4], Eric Betzig[3], Jan Liphardt[1,2,5*]

1 Biophysics Graduate Group, University of California Berkeley, Berkeley, California, United States of America, **2** Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, United States of America, **3** Howard Hughes Medical Institute, Janelia Farm Research Campus, Ashburn, Virginia, United States of America, **4** Department of Molecular Biology, Princeton University, Princeton, New Jersey, United States of America, **5** Department of Physics, University of California Berkeley, Berkeley, California, United States of America

Abstract

The *Escherichia coli* chemotaxis network is a model system for biological signal processing. In *E. coli*, transmembrane receptors responsible for signal transduction assemble into large clusters containing several thousand proteins. These sensory clusters have been observed at cell poles and future division sites. Despite extensive study, it remains unclear how chemotaxis clusters form, what controls cluster size and density, and how the cellular location of clusters is robustly maintained in growing and dividing cells. Here, we use photoactivated localization microscopy (PALM) to map the cellular locations of three proteins central to bacterial chemotaxis (the Tar receptor, CheY, and CheW) with a precision of 15 nm. We find that cluster sizes are approximately exponentially distributed, with no characteristic cluster size. One-third of Tar receptors are part of smaller lateral clusters and not of the large polar clusters. Analysis of the relative cellular locations of 1.1 million individual proteins (from 326 cells) suggests that clusters form via stochastic self-assembly. The super-resolution PALM maps of *E. coli* receptors support the notion that stochastic self-assembly can create and maintain approximately periodic structures in biological membranes, without direct cytoskeletal involvement or active transport.

Citation: Greenfield D, McEvoy AL, Shroff H, Crooks GE, Wingreen NS, et al. (2009) Self-Organization of the *Escherichia coli* Chemotaxis Network Imaged with Super-Resolution Light Microscopy. PLoS Biol 7(6): e1000137. doi:10.1371/journal.pbio.1000137

Academic Editor: Howard C. Berg, Harvard University, United States of America

Received October 27, 2008; **Accepted** May 14, 2009; **Published** June 23, 2009

Funding: ALM thanks the NSF for graduate research support. This work was supported by the Sloan and Searle Foundations (JL), the DOE Office of Science, Energy Biosciences Program (JL), and National Institutes of Health grants R01 GM77856 (JL), R01 GM084716 (JL), and R01 GMO73186 (NSW). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors declare competing financial interests. EB and Harald Hess (Janelia Farm) have licensed the PALM technology to Carl Zeiss Microimaging, GmbH.

Abbreviations: DIC, differential interference contrast; epi, epifluorescence; PALM, photoactivated localization microscopy; TIR, total internal reflection.

* E-mail: Liphardt@berkeley.edu

☯ These authors contributed equally to this work.

Introduction

Efficient biological signal processing often requires complex spatial organization of the signaling machinery. Understanding how this spatial organization is generated, maintained, and repaired inside cells is a fundamental theme of biology. A well-understood signaling network with complex spatial organization is the bacterial chemotaxis system, which directs the movement of cells towards or away from sugars, amino acids, and many other soluble molecules [1]. In *Escherichia coli*, five types of transmembrane chemoreceptors form trimers of dimers [2,3], which cluster into large complexes containing tens of thousands of proteins [4–7]. Receptor clustering enables cooperative interactions between receptors [8–11], contributing to a bacterium's ability to sense nanomolar concentrations of chemicals and small fractional changes in chemical concentrations over a wide range [12–14]. Chemotaxis clusters are stabilized by the adaptor protein CheW and the histidine kinase CheA, which bind receptors in a ternary complex. CheA transduces signals from membrane receptors to the cytoplasmic response regulator CheY, which diffuses to flagellar motors and modulates their direction of rotation (Figure 1A; for review see [5]).

A variety of imaging studies have advanced our understanding of how the spatial organization of the chemotaxis network arises and contributes to function [15]. Time-lapse fluorescence microscopy suggests that receptors are inserted randomly into the lateral membrane via the general protein translocation machinery and then diffuse to existing clusters [16]. Immunoelectron and fluorescence microscopy have shown that receptor clusters are found at the cell poles [4] and future division sites [17].

Despite much research, the fundamental mechanisms responsible for positioning chemotaxis clusters at specific sites in the membrane remain unclear [15]. Perhaps cells possess intracellular structures that anchor clusters to periodic sites along cell length [17]. However, fluorescence microscopy of cells overexpressing all chemotaxis proteins showed that the number of clusters per cell saturates well below the number of proposed cluster anchoring sites. Furthermore, the distance between chemotaxis clusters varies broadly within cells [18]. Based on those observations, Thiem and Sourjik [18] proposed that cluster nucleation and growth is a stochastic self-assembly process in which receptors freely diffuse in the membrane and then join existing clusters or nucleate new clusters. In their model, clusters nucleate anywhere in the

Regulated Fluctuations in Nanog Expression Mediate Cell Fate Decisions in Embryonic Stem Cells

Tibor Kalmar[1☯], Chea Lim[1☯], Penelope Hayward[1☯], Silvia Muñoz-Descalzo[1], Jennifer Nichols[2], Jordi Garcia-Ojalvo[3], Alfonso Martinez Arias[1]*

1 Department of Genetics, University of Cambridge, Cambridge, United Kingdom, **2** Wellcome Trust Centre for Stem Cell Research, University of Cambridge, Cambridge, United Kingdom, **3** Departament de Fisica i Enginyeria Nuclear, Universitat Politecnica de Catalunya, Colom 11, Terrassa, Spain

Abstract

There is evidence that pluripotency of mouse embryonic stem (ES) cells is associated with the activity of a network of transcription factors with Sox2, Oct4, and Nanog at the core. Using fluorescent reporters for the expression of Nanog, we observed that a population of ES cells is best described by a dynamic distribution of Nanog expression characterized by two peaks defined by high (HN) and low (LN) Nanog expression. Typically, the LN state is 5%–20% of the total population, depending on the culture conditions. Modelling of the activity of Nanog reveals that a simple network of Oct4/Sox2 and Nanog activity can account for the observed distribution and its properties as long as the transcriptional activity is tuned by transcriptional noise. The model also predicts that the LN state is unstable, something that is born out experimentally. While in this state, cells can differentiate. We suggest that transcriptional fluctuations in Nanog expression are an essential element of the pluripotent state and that the function of Sox2, Oct4, and Nanog is to act as a network that promotes and maintains transcriptional noise to interfere with the differentiation signals.

Citation: Kalmar T, Lim C, Hayward P, Muñoz-Descalzo S, Nichols J, et al. (2009) Regulated Fluctuations in Nanog Expression Mediate Cell Fate Decisions in Embryonic Stem Cells. PLoS Biol 7(7): e1000149. doi:10.1371/journal.pbio.1000149

Academic Editor: Margaret A. Goodell, Baylor College of Medicine, United States of America

Received August 22, 2008; **Accepted** May 28, 2009; **Published** July 7, 2009

Funding: This work has been supported by The Wellcome Trust (AMA, PH, JN, SM), BBSRC (TK, AMA), MEC Spain, project FIS2006-11452 and the I3 program, (JGO) and the Yousef Jameel scholarship fund (CL). 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.

Abbreviations: EB, embryoid body; EC, embryonal carcinoma; ES, embryonic stem; GRN, gene regulatory network; HN, high Nanog expression; LN, low Nanog expression; SON, Sox2, Oct4, and Nanog.

* E-mail: ama11@hermes.cam.ac.uk

☯ These authors contributed equally to this work.

Introduction

Embryonic stem (ES) cells are cultured pluripotent cell populations derived from the epiblast of mammalian embryos, which can be induced to differentiate into a variety of cell types under controlled conditions [1–4]. Most studies with ES cells have been performed on mouse cells, and their pluripotent nature has been demonstrated by their ability to contribute to all tissues of a developing embryo in chimeras [5]. Although derived from the epiblast, the state of the ES cells is similar to that of the inner-cell mass (ICM), and this is reflected in their patterns of differentiation in embryos and in culture [5–8]. These observations suggest that ES cells might represent a good experimental system to explore the molecular basis that underlies the establishment and maintenance of different cell fates and their transitions during development.

Traditionally the maintenance of pluripotency in culture requires Leukemia Inhibiting Factor (LIF) [9] and serum or BMP4 [10], as well as the activity of a small, gene regulatory network (GRN) with three core transcription factors: Sox2, Oct4, and Nanog [11–13]. Oct4, a homeobox-containing factor, and Sox2, an HMG box protein, bind together at many sites in the genome, including their own promoters and that of Nanog [14–16]. Loss-of-function studies show that Oct4 and Sox2 act together to promote self-renewal of ES cells by preventing differentiation [17–20]. The levels of Oct4 are particularly critical for the state of a cell: whereas loss of Oct4 results in the loss of pluripotency and differentiation into trophoectoderm (TE), excess Oct4 activity results in differentiation into primitive endoderm (PE)-like cells [21–23]. However, these two factors are not sufficient to maintain the pluripotent state, as they cannot act in the absence of LIF. In contrast, when the divergent-homeobox–containing protein Nanog is overexpressed in ES cells, it is sufficient for sustaining pluripotency in the absence of LIF [13,24], and there is a correlation between its levels and the degree of pluripotency of a cell [25,26]. Although there are other transcription factors associated with ES cells [27–29], a large number of studies support the notion that the trio Sox2, Oct4, and Nanog (SON) is at the heart of a GRN that generates and maintains the pluripotent state. This has been underlined in a number of recent experiments in which these factors play an essential role in the induction of pluripotent stem cells from differentiated cells [30–32]. Despite the effectiveness of these transcription factors in promoting and maintaining pluripotency, their mode of action remains unclear.

A number of studies have explored the possibility that there are genes downstream of the SON network that implement the pluripotent state [29,33–36] and, furthermore, that chromatin modifications play a role in the maintenance of this state (reviewed

TraR, a Homolog of a RNAP Secondary Channel Interactor, Modulates Transcription

Matthew D. Blankschien[1], Katarzyna Potrykus[2], Elicia Grace[1,3], Abha Choudhary[1], Daniel Vinella[2], Michael Cashel[2], Christophe Herman[1,3]*

1 Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, United States of America, **2** Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland, United States of America, **3** Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, Texas, United States of America

Abstract

Recent structural and biochemical studies have identified a novel control mechanism of gene expression mediated through the secondary channel of RNA Polymerase (RNAP) during transcription initiation. Specifically, the small nucleotide ppGpp, along with DksA, a RNAP secondary channel interacting factor, modifies the kinetics of transcription initiation, resulting in, among other events, down-regulation of ribosomal RNA synthesis and up-regulation of several amino acid biosynthetic and transport genes during nutritional stress. Until now, this mode of regulation of RNAP was primarily associated with ppGpp. Here, we identify TraR, a DksA homolog that mimics ppGpp/DksA effects on RNAP. First, expression of TraR compensates for *dksA* transcriptional repression and activation activities in vivo. Second, mutagenesis of a conserved amino acid of TraR known to be critical for DksA function abolishes its activity, implying both structural and functional similarity to DksA. Third, unlike DksA, TraR does not require ppGpp for repression of the *rrnB* P1 promoter in vivo and in vitro or activation of amino acid biosynthesis/transport genes in vivo. Implications for DksA/ppGpp mechanism and roles of TraR in horizontal gene transfer and virulence are discussed.

Citation: Blankschien MD, Potrykus K, Grace E, Choudhary A, Vinella D, et al. (2009) TraR, a Homolog of a RNAP Secondary Channel Interactor, Modulates Transcription. PLoS Genet 5(1): e1000345. doi:10.1371/journal.pgen.1000345

Editor: William F. Burkholder, Stanford University, United States of America

Received June 19, 2008; **Accepted** December 17, 2008; **Published** January 16, 2009

Funding: This work is funded in part by the NICHD intramural program of the NIH and HFSPO grant RGY0060/2006.

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

* E-mail: herman@bcm.edu

Introduction

The ability to respond to changes in nutritional environment is a universal need inherent in all cells and is characterized by rapid global changes in gene expression. Regulation of transcription initiation is a central way to control gene expression and is largely achieved through the use of DNA-binding proteins (activators and repressors) restricted to distinct promoters through recognition of specific DNA elements. Study of the nutritional response in *Escherichia coli* has detailed a novel mechanism of modulating transcription initiation, both positively and negatively, through the use of a single small nucleotide effector, guanosine tetraphosphate (ppGpp), that interacts with RNA polymerase [1]. In *E. coli*, the accumulation of ppGpp causes rapid effects on transcription; ppGpp binds to RNA polymerase, provoking an alteration in transcription kinetics that is proposed to result from a reduction in open complex stability [2]. Such effects include, but are not limited to, upregulation of amino acid biosynthesis and transport genes, as well as genes involved in stasis/stress survival, and downregulation of translational components such as rRNA and tRNA genes [3].

Recently, an additional factor, DksA, has been shown to potentiate the action of ppGpp on RNAP both in vitro and in vivo [4–6]. The loss of either ppGpp or DksA results in similar, though not identical, phenotypes including the downregulation of several amino acid biosynthetic pathways, and the inability to negatively regulate ribosomal RNA transcription [7,8]. Separate from mediating the stringent response, DksA has roles in other processes including chromosome segregation, DNA repair, protein folding, bacterial motility, virulence, and the expression of type 1 fimbriae [7–14]. The crystal structure of DksA has been determined and shows that the 151 amino acid-long protein folds into three distinct structural domains: an N-terminal region containing two α-helices (coiled coil), a globular domain with a C4 Zn^{+2} finger motif, and a short C-terminal helix [15]. DksA is structurally analogous to GreA and GreB, transcriptional anti-pausing/fidelity factors that are homologs of the eukaryotic TFIIS [16–19]. The Gre factors bind RNAP and protrude their coiled coils deep into the secondary channel toward the active site [20,21]. DksA also binds RNAP, and it has been suggested that, similar to the Gre factors, DksA could also interact with the secondary channel. This is supported by a growing body of evidence indicating that the DksA and Gre proteins compete in vivo for the same substrate, the secondary channel of RNAP [22,23]. A proposed mechanism of action for DksA positions the coiled coil region deep within the RNAP secondary channel near the active site. At the tip of the coiled coil region, two invariant aspartic acid residues, Asp71 and Asp74, are thought to coordinate the ppGpp bound Mg^{+2} ion to effectively position ppGpp near the active site, and allow it to exert its transcriptional modulation effects [15]. Mutation of these two conserved aspartic acid residues abolishes DksA's ability to modulate transcription with ppGpp [15]. The proposed DksA mechanism remains highly speculative, and it is unknown exactly how ppGpp and DksA influence each other or how their binding alters RNAP activity. Furthermore, detailed mutational analysis of

OPEN ACCESS Freely available online

PLoS GENETICS

Review

Bacterial Toxin–Antitoxin Systems: More Than Selfish Entities?

Laurence Van Melderen*, Manuel Saavedra De Bast

Laboratoire de Génétique et Physiologie Bactérienne, IBMM, Faculté des Sciences, Université Libre de Bruxelles, Gosselies, Belgium

Abstract: Bacterial toxin–antitoxin (TA) systems are diverse and widespread in the prokaryotic kingdom. They are composed of closely linked genes encoding a stable toxin that can harm the host cell and its cognate labile antitoxin, which protects the host from the toxin's deleterious effect. TA systems are thought to invade bacterial genomes through horizontal gene transfer. Some TA systems might behave as selfish elements and favour their own maintenance at the expense of their host. As a consequence, they may contribute to the maintenance of plasmids or genomic islands, such as super-integrons, by post-segregational killing of the cell that loses these genes and so suffers the stable toxin's destructive effect. The function of the chromosomally encoded TA systems is less clear and still open to debate. This Review discusses current hypotheses regarding the biological roles of these evolutionarily successful small operons. We consider the various selective forces that could drive the maintenance of TA systems in bacterial genomes.

Introduction

Although bacteria have long been known to exchange genetic information through horizontal gene transfer, the impact of this dynamic process on genome evolution was fully appreciated only recently using comparative genomics (reviewed in [1]). Bacterial chromosomes are composed of genes that have quite different evolutionary origins (reviewed in [2]). The set of genes that is preferentially transmitted vertically over long evolutionary time scales composes the core genome. Core genes are relatively well conserved among different monophyletic groups and encode the cellular core functions. These core genes are interspersed with groups of genes that have been acquired from other prokaryotic genomes by horizontal transmission. These genomic islands mostly originate from integration events of mobile genetic elements, such as insertion sequences, transposons, phages, and plasmids. They might, therefore, be found in phylogenetically distant species and are not conserved among different isolates belonging to the same bacterial species. This set of genes constitutes the flexible genome.

Both gene influx and efflux processes are important in shaping bacterial-genome content. A vast majority of horizontally transferred genes are quickly lost after integration [3], although some remain interspersed in the genome (reviewed in [2]). Bacterial toxin-antitoxin (TA) systems appear to be subjected to this flux. Indeed, these small gene systems are found in plasmids as well as in chromosomes, and they are thought to be part of the flexible genome [4]. Although their role, when they are located in plasmid, is fairly clear, the involvement in physiological processes of the TA systems' chromosomally encoded counterparts is still open to debate.

Here we discuss current hypotheses regarding the biological roles of chromosomally encoded TA systems and consider the various selective forces that could drive the maintenance of TA systems in bacterial genomes.

Diversity and Abundance of Bacterial TA Systems

Bacterial TA systems are of two different types depending on the nature of the antitoxin; the toxin always being a protein. The antitoxin of type I systems is a small RNA (antisense or adjacent and divergent to the toxin gene) showing complementarity to the toxin mRNA (for recent reviews on type I systems, see [5,6]). Type I antitoxins regulate toxin expression by inhibiting the toxin's translation. The toxins of type I systems are small, hydrophobic proteins that cause damage in bacterial cell membranes. In type II systems, the antitoxin is a small, unstable protein that sequesters the toxin through proteic complex formation (for a recent review on type II systems, see [7]). Much more information is available for type II systems, especially in terms of their biological roles. We will focus on the type II systems and use the term TA systems for brevity.

Type II TA systems are organised in operons, with the upstream gene usually encoding the antitoxin protein. The expression of the two genes is regulated at the level of transcription by the antitoxin–toxin complex. Nine families of toxins have been defined so far based on amino sequence homology [4]. Their targets and the cellular processes that are affected by their activities are shown in Table 1.

Comprehensive genome analyses have highlighted the diversity in the distribution of TA systems [4,8,9]. Some genomes such as that of *Nitrosomonas europeae*, *Sinorhizobium meliloti*, and *Mycobacterium bovis* contain more than 50 putative TA systems. Some others contain no or very few (less than three) putative TA systems, such as *Rickettsia prowazeki*, *Campylobacter jejuni*, or *Bacillus subtilis*. No correlation between the number of TA systems, the lifestyle, the membership of a phylum, or the growth rate (as it was proposed [4]) could be drawn [9]. Another level of diversity in distribution of TA systems among bacteria is added when comparing the occurrence of TA systems between different isolates of the same

Citation: Van Melderen L, Saavedra De Bast M (2009) Bacterial Toxin–Antitoxin Systems: More Than Selfish Entities? PLoS Genet 5(3): e1000437. doi:10.1371/journal.pgen.1000437

Editor: Susan M. Rosenberg, Baylor College of Medicine, United States of America

Published March 27, 2009

Funding: The authors received no specific funding for this article.

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

* E-mail: lvmelder@ulb.ac.be

JOURNAL OF BACTERIOLOGY, July 2009, p. 4487–4491 Vol. 191, No. 14
0021-9193/09/$08.00+0 doi:10.1128/JB.00056-09

Dominant Negative Autoregulation Limits Steady-State Repression Levels in Gene Networks▽†

Szabolcs Semsey,[1]* Sandeep Krishna,[2] János Erdőssy,[1] Péter Horváth,[1] László Orosz,[1] Kim Sneppen,[2] and Sankar Adhya[3]

Department of Genetics, Eötvös Lóránd University, H-1117 Budapest, Hungary[1]; Center for Models of Life, Niels Bohr Institute, Copenhagen, Denmark[2]; and Laboratory of Molecular Biology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892-4264[3]

Received 16 January 2009/Accepted 3 May 2009

Many transcription factors repress transcription of their own genes. Negative autoregulation has been shown to reduce cell-cell variation in regulatory protein levels and speed up the response time in gene networks. In this work we examined transcription regulation of the *galS* gene and the function of its product, the GalS protein. We observed a unique operator preference of the GalS protein characterized by dominant negative autoregulation. We show that this pattern of regulation limits the repression level of the target genes in steady states. We suggest that transcription factors with dominant negative autoregulation are designed for regulating gene expression during environmental transitions.

A large class of transcription factors (TFs) responds to environmental or intracellularly synthesized signals and changes transcription of a gene set accordingly. In many cases TFs enhance (positive autoregulation) or inhibit (negative autoregulation) their own synthesis (8, 16). Autoregulation of TFs plays an important role in genetic networks. For example, positively autoregulated TFs are key elements of switches and memory devices, while negative autoregulation of a TF can provide stability and speed up response time (1, 2, 6, 9, 11, 15).

In this work we study regulation of the *galS* gene and the function of its product, the Gal isorepressor (GalS) protein in *Escherichia coli*. GalS inhibits transcription of the galactose (*gal*) regulon promoters in *E. coli* by binding to 16-bp operator sequences in the *cis*-regulatory regions of the $P1_{galE}$, P_{mglB}, P_{galP}, P_{galS}, and P_{galR} promoters (14, 18). DNA binding by GalS is inhibited in the presence of the sugar D-galactose (19). Transcription of the *galS* gene is highly dependent on the cyclic AMP (cAMP) receptor protein (CRP), which in the presence of cAMP activates the P_{galS} promoter (17). Negative autoregulation of GalS has been demonstrated both in vivo and in vitro (14, 17). There are two operator sites in the *galS* gene, one upstream of the promoter (*galS* O_E) and a second in the coding sequence (*galS* O_I); however, the role of *galS* O_I is unclear (10, 17) (Fig. 1). We measure transcription levels of the *gal* regulon promoters in an in vitro system in the presence of GalS and use mathematical modeling for analysis of results.

We report that negative autoregulation dominates GalS function. The significance of dominant negative autoregulation in genetic networks is discussed.

MATERIALS AND METHODS

Proteins. Hexahistidine-tagged GalS was purified as described earlier (13). CRP was purified as described by Ryu et al. (12). σ^{70} RNA polymerase was purchased from USB.

In vitro transcription. Transcription reactions were performed as described previously (4). The reaction mixture (50 μl) contained 20 mM Tris acetate, pH 7.8, 10 mM magnesium acetate, 200 mM potassium glutamate, and 2 nM supercoiled pRPGSM plasmid DNA template. Plasmid pRPGSM was maintained in Stbl2 cells (Invitrogen) to avoid instability of directly repeated sequences. CRP was used at 50 nM, and cAMP was used at 100 μM, when present. RNA polymerase (20 nM) was added before the reaction mixtures were incubated at 37°C for 5 min. Transcription was started by the addition of 1.0 mM ATP, 0.1 mM GTP, 0.1 mM CTP, 0.01 mM UTP, and 5 μCi of [α-^{32}P]UTP (3,000 Ci/mmol). Reactions were terminated after 10 min by the addition of an equal volume of transcription loading buffer (0.025% bromophenol blue, 0.025% xylene cyanol, 0.01 M EDTA, and 90% deionized formamide). After samples were heated at 90°C for 3 min, they were loaded onto 7% polyacrylamide-urea DNA sequencing gels. RNA bands were quantified using an ImageQuant PhosphorImager (Molecular Dynamics, CA).

Mathematical model of TF autoregulation. We model regulation by a transcription factor that responds to a small-molecule signal and represses transcription of a gene set including its own gene. In this model both small-molecule binding and DNA binding by the transcription factor are reversible and occur on a timescale much faster than production or degradation of the TF(5, 7). Binding of the small-molecule inducer inhibits TF binding to DNA. Therefore, the total intracellular TF concentration (TF_T) is the sum of concentrations of three forms of TF: (i) bound to DNA (TF_D), (ii) forming a complex with the small molecule (TFi), and (iii) free, i.e., not bound to ligands (TF_F). Typically TF_D is almost always negligible compared to TF_F; therefore,

$$TF_T = TF_F + (TF_i) = TF_F + TF_F \times (i/K_i)^{h_i}$$

$$\Rightarrow TF_F = \frac{TF_T}{1 + (i/K_i)^{h_i}} \quad (1)$$

where i is the concentration of small molecule inducer, K_i is the dissociation constant for TF-inducer binding, and h_i is the associated Hill coefficient.

The equation governing the dynamics of the concentration of the autoregulated TF is

$$\frac{dTF_T}{dt} = \alpha \times \left(\frac{1}{1 + (TF_F)^{h_{self}}}\right) - TF_T \quad (2)$$

Here, time is measured in units of the average lifetime of the TF (taken to be one cell generation), and the concentration TF_T is measured in units of the self-repression dissociation constant, K_{self}. Then, α is the maximum rate of production of TF_T possible in these units (which is also the maximum possible concen-

* Corresponding author. Mailing address: Department of Genetics, Eötvös Lóránd University, Pázmány P. s. 1/C, H-1117 Budapest, Hungary. Phone: 36 (1) 3722500. Fax: 36 (1) 3722641. E-mail: semseys@yahoo.com.

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

▽ Published ahead of print on 12 May 2009.

JOURNAL OF BACTERIOLOGY, July 2009, p. 4180–4185 Vol. 191, No. 13
0021-9193/09/$08.00+0 doi:10.1128/JB.01707-08

Active Transcription of rRNA Operons Condenses the Nucleoid in *Escherichia coli*: Examining the Effect of Transcription on Nucleoid Structure in the Absence of Transertion[▽]

Julio E. Cabrera,[1]†‡ Cedric Cagliero,[1]† Selwyn Quan,[2] Catherine L. Squires,[2] and Ding Jun Jin[1]*

Transcription Control Section, Gene Regulation and Chromosome Biology Laboratory, National Cancer Institute-Frederick, National Institutes of Health, Frederick, Maryland 21702,[1] and Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, Massachusetts 02111[2]

Received 6 December 2008/Accepted 5 April 2009

In *Escherichia coli* the genome must be compacted ~1,000-fold to be contained in a cellular structure termed the nucleoid. It is proposed that the structure of the nucleoid is determined by a balance of multiple compaction forces and one major expansion force. The latter is mediated by transertion, a coupling of transcription, translation, and translocation of nascent membrane proteins and/or exported proteins. In supporting this notion, it has been shown consistently that inhibition of transertion by the translation inhibitor chloramphenicol results in nucleoid condensation due to the compaction forces that remain active in the cell. Our previous study showed that during optimal growth, RNA polymerase is concentrated into transcription foci or "factories," analogous to the eukaryotic nucleolus, indicating that transcription and RNA polymerase distribution affect the nucleoid structure. However, the interpretation of the role of transcription in the structure of the nucleoid is complicated by the fact that transcription is implicated in both compacting forces and the expansion force. In this work, we used a new approach to further examine the effect of transcription, specifically from rRNA operons, on the structure of the nucleoid, when the major expansion force was eliminated. Our results showed that transcription is necessary for the chloramphenicol-induced nucleoid compaction. Further, an active transcription from multiple rRNA operons in chromosome is critical for the compaction of nucleoid induced by inhibition of translation. All together, our data demonstrated that transcription of rRNA operons is a key mechanism affecting genome compaction and nucleoid structure.

An *Escherichia coli* cell is small, measuring approximately 2 to 4 μm in length and 1 μm in diameter. The bacterial genome is 4.6 million bp, which would be approximately 1.5 mm in length if stretched fully. In a rapidly growing cell, there are multiple genome equivalents. Thus, the genome must be compressed at least 1,000-fold to fit into the cell. The bacterial chromosome forms a cellular structure named the nucleoid (25, 42). Normally the *E. coli* nucleoid shows a characteristic "flexible doublet" shape (49) and is membrane associated (2, 46). Despite great advances being made in understanding the biochemistry and molecular biology of *E. coli*, the structure of the bacterial nucleoid remains poorly defined.

Woldringh et al. proposed that the structure of the nucleoid is determined by a balance of expansion and compaction forces (44). Suggested compaction forces include (i) DNA binding proteins (9, 17), (ii) DNA supercoiling (29, 35, 38), (iii) macromolecular crowding (20, 23, 51), and (iv) entropy-driven depletion attraction (18). One of the proposed forces that significantly contributes to expansion of the nucleoid is called transertion. During this process, coupled transcription and translation of membrane proteins and/or periplasmic exported proteins pull and anchor the transcribed bacterial nucleoid onto the cytoplasmic membrane (3, 43). In addition, RNA polymerase (RNAP) is thought to be a driving force in nucleoid segregation (12). This notion is further supported by the report of an interaction between RNAP and the actin-like MreB protein during chromosome segregation (15). According to the proposition (44), inhibition of both translation and transcription would lead to the disruption of transertion. It is expected that when transertion is blocked, the nucleoid will be condensed due to the remaining compaction forces in the cell. In support of this concept, it is consistently reported that chloramphenicol, a translation inhibitor, induces nucleoid compaction in the cell (41, 50). However, there are conflicting results regarding the effect of rifampin on nucleoid structure: both rifampin-induced nucleoid expansion (6, 13, 26, 36) and nucleoid compaction (3, 40, 52, 53) have been reported. Thus, the exact role of transcription in the structure of the nucleoid remains poorly understood and understudied (28).

Imaging of RNAP inside the cell containing the chromosomal *rpoC-gfp* fusion has provided a new tool to study the effects of transcription and RNAP (re)distribution on the structure of the nucleoid (6). Our recent studies suggest that transcription, in particular that from rRNA (*rrn*) operons, is linked to the dynamic structure of the nucleoid (14). In *E. coli*, there are seven *rrn* operons, four of which are located near the origin of replication in the chromosome. Despite the fact that collectively the *rrn* operons represent only about 1% of the genome, transcription of the genes in the *rrn* operons accounts for approximately 85% of the total transcription in an *E. coli*

* Corresponding author. Mailing address: Gene Regulation and Chromosome Biology Laboratory, National Cancer Institute-Frederick, National Institutes of Health, 1050 Boyles St., Frederick, MD 21702. Phone: (301) 846-7684. Fax: (301) 846-1489. E-mail: djjin@helix.nih.gov.

† These two authors contributed equally to this work.

‡ Present address: ProteinOne, Inc., Bethesda, MD 20814.

▽ Published ahead of print on 24 April 2009.

Molecular Systems Biology 5; Article number 276; doi:10.1038/msb.2009.33
Citation: *Molecular Systems Biology* 5:276

www.molecularsystemsbiology.com

molecular systems biology

REPORT

Backup in gene regulatory networks explains differences between binding and knockout results

Anthony Gitter[1], Zehava Siegfried[2], Michael Klutstein[2], Oriol Fornes[3], Baldo Oliva[4], Itamar Simon[2] and Ziv Bar-Joseph[1,5,*]

[1] Computer Science Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA, [2] Department of Molecular Biology, Hebrew University Medical School, Jerusalem, Israel, [3] Department of Experimental Sciences and Health, Municipal Institute for Medical Research (IMIM-Hospital del Mar), Barcelona, Catalonia, Spain, [4] Department of Experimental Sciences and Health, Pompeu Fabra University, Barcelona, Catalonia, Spain and [5] Machine Learning Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA
* Corresponding author. Computer Science Department, School of Computer Science, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA. Tel.: +1 412 268 8595; Fax: +1 412 268 3431; E-mail: zivbj@cs.cmu.edu

Received 28.11.08; accepted 29.4.09

The complementarity of gene expression and protein–DNA interaction data led to several successful models of biological systems. However, recent studies in multiple species raise doubts about the relationship between these two datasets. These studies show that the overwhelming majority of genes bound by a particular transcription factor (TF) are not affected when that factor is knocked out. Here, we show that this surprising result can be partially explained by considering the broader cellular context in which TFs operate. Factors whose functions are not backed up by redundant paralogs show a fourfold increase in the agreement between their bound targets and the expression levels of those targets. In addition, we show that incorporating protein interaction networks provides physical explanations for knockout effects. New double knockout experiments support our conclusions. Our results highlight the robustness provided by redundant TFs and indicate that in the context of diverse cellular systems, binding is still largely functional.
Molecular Systems Biology 5: 276; published online 16 June 2009; doi:10.1038/msb.2009.33
Subject Categories: simulation and data analysis; chromatin and transcription
Keywords: backup mechanisms; paralogs; protein interactions

Introduction

Many successful studies in systems biology focus on integrating complementary datasets to model systems in the cell. Several computational methods have been developed and applied to combine mRNA expression data and protein–DNA interaction data (using DNA-binding motifs, ChIP-chip experiments, or both) (Lee *et al*, 2002; Liao *et al*, 2003; Beer and Tavazoie, 2004; Yeang *et al*, 2005; Ernst *et al*, 2007). These methods assume that transcript levels are largely driven by binding of transcription factors (TFs) to DNA leading to either expression or repression of the bound genes. Indeed, by some estimates close to 60% of binding sites are actively driving expression of their bound genes (Gao *et al*, 2004).

This assumption was recently challenged by several studies that compared the set of genes bound by a TF with the set of genes affected when that factor is knocked out or knocked down. One of the earliest reports of this phenomenon involved the yeast cell cycle (Horak *et al*, 2002). Using ChIP-chip experiments, researchers looked at the set of genes bound by 11 TFs and concluded that complementary knockout experiments did not affect the same set of genes. In mouse, it was reported that only 11% of those genes that were differentially expressed after glucocorticoid dexamethasone injection were also bound by the glucocorticoid receptor (Phuc Le *et al*, 2005). An estrogen-response study in human reported that 6% of E_2-induced genes were bound by ERα, and 13% of ERα-bound genes were regulated by E_2 (Kwon *et al*, 2007). A human study in which p63 was depleted led to similar conclusions (Yang *et al*, 2006).

Although the above experiments looked at only one, or few, TFs, a recent study in yeast examined the overlap for the entire set of TFs and surprisingly concluded that the overlap was even smaller than the overlaps reported above for individual factors. In a comprehensive analysis of the agreement between binding and knockout experiments, 269 budding yeast TFs

Molecular Systems Biology 5; Article number 285; doi:10.1038/msb.2009.42
Citation: *Molecular Systems Biology* 5:285

www.molecularsystemsbiology.com

molecular systems biology

Prevalence of transcription promoters within archaeal operons and coding sequences

Tie Koide[1,5,6], David J Reiss[1,5], J Christopher Bare[1], Wyming Lee Pang[1], Marc T Facciotti[1,2], Amy K Schmid[1], Min Pan[1], Bruz Marzolf[1], Phu T Van[1], Fang-Yin Lo[1], Abhishek Pratap[1], Eric W Deutsch[1], Amelia Peterson[3], Dan Martin[1,3] and Nitin S Baliga[1,4,*]

[1] Institute for Systems Biology, Seattle, WA, USA, [2] Department of Biomedical Engineering and UC Davis Genome Center, One Shields Avenue, University of California, Davis, CA, USA, [3] Divisions of Human Biology and Clinical Research, Fred Hutchinson Cancer Research Center, Seattle, WA, USA and [4] Departments of Microbiology, and Molecular and Cellular Biology, University of Washington, Seattle, WA, USA
[5] These authors contributed equally to this work
[6] Present address: Departamento de Bioquímica e Imunologia, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Brazil. E-mail: tiekoide@gmail.com
* Corresponding author. Institute for Systems Biology, Departments of Microbiology, and Molecular and Cellular Biology, University of Washington, 1441 N 34th Street, Seattle, WA 98103, USA. Tel.: +1 206 732 1266; Fax: +1 206 732 1299; E-mail: nbaliga@systemsbiology.org

Received 20.11.08; accepted 13.5.09

Despite the knowledge of complex prokaryotic-transcription mechanisms, generalized rules, such as the simplified organization of genes into operons with well-defined promoters and terminators, have had a significant role in systems analysis of regulatory logic in both bacteria and archaea. Here, we have investigated the prevalence of alternate regulatory mechanisms through genome-wide characterization of transcript structures of ~64% of all genes, including putative non-coding RNAs in *Halobacterium salinarum NRC-1*. Our integrative analysis of transcriptome dynamics and protein–DNA interaction data sets showed widespread environment-dependent modulation of operon architectures, transcription initiation and termination inside coding sequences, and extensive overlap in 3′ ends of transcripts for many convergently transcribed genes. A significant fraction of these alternate transcriptional events correlate to binding locations of 11 transcription factors and regulators (TFs) inside operons and annotated genes—events usually considered spurious or non-functional. Using experimental validation, we illustrate the prevalence of overlapping genomic signals in archaeal transcription, casting doubt on the general perception of rigid boundaries between coding sequences and regulatory elements.
Molecular Systems Biology 5: 285; published online 16 June 2009; doi:10.1038/msb.2009.42
Subject Categories: functional genomics; chromatin & transcription
Keywords: archaea; ChIP–chip; non-coding RNA; tiling array; transcription

Introduction

Systems-biology approaches have been successfully applied to construct quantitative and predictive models of biological networks (Bonneau *et al*, 2007; Faith *et al*, 2007). However, a significant amount of information is missing from these models because of incomplete parts lists (unannotated genes, non-coding RNAs (ncRNAs), poorly understood protein modifications and so on) as well as a lack of molecular detail associated with these processes. Incorporating such detail will make these models mechanistically accurate and useful for synthetic-biology approaches targeting large-scale biological-circuit re-engineering. Among the current systems-scale models most amenable for such large-scale redesign are those that describe gene-regulatory networks (GRNs).

GRN models are usually built upon transcriptome data, in which typically genes or gene modules (with similar expression patterns and shared regulatory motifs) are associated with their transcriptional regulators through linear or Bayesian models. However, although these models can be predictive (Bonneau *et al*, 2007), they often rely on approximations of the transcription process and lack finer details of dynamic environment-dependent assembly of transcription complexes at each of the numerous promoters in the genome. High-density tiling arrays can be used to define transcribed regions (David *et al*, 2006), start sites (McGrath *et al*, 2007), and protein–DNA interaction sites (Reiss *et al*, 2008), which can be used to identify some of these missing details associated with transcriptional regulation, and thereby enable us to construct systems-scale predictive models of GRNs that are also mechanistically accurate.

nature biotechnology

Quantification of the yeast transcriptome by single-molecule sequencing

Doron Lipson[1,2], Tal Raz[1,2], Alix Kieu[1], Daniel R Jones[1], Eldar Giladi[1], Edward Thayer[1], John F Thompson[1], Stan Letovsky[1], Patrice Milos[1] & Marie Causey[1]

We present single-molecule sequencing digital gene expression (smsDGE), a high-throughput, amplification-free method for accurate quantification of the full range of cellular polyadenylated RNA transcripts using a Helicos Genetic Analysis system. smsDGE involves a reverse-transcription and polyA-tailing sample preparation procedure followed by sequencing that generates a single read per transcript. We applied smsDGE to the transcriptome of *Saccharomyces cerevisiae* strain DBY746, using 6 of the available 50 channels in a single sequencing run, yielding on average 12 million aligned reads per channel. Using spiked-in RNA, accurate quantitative measurements were obtained over four orders of magnitude. High correlation was demonstrated across independent flow-cell channels, instrument runs and sample preparations. Transcript counting in smsDGE is highly efficient due to the representation of each transcript molecule by a single read. This efficiency, coupled with the high throughput enabled by the single-molecule sequencing platform, provides an alternative method for expression profiling.

Analysis of gene expression has been a primary tool in the study of cellular mechanisms. Large-scale sequencing of cDNA clones and comparisons of transcript abundance between samples have provided valuable insights into the gene content and tissue-specific and developmental expression patterns of a wide range of organisms. More recently, microarray expression profiling has provided gene expression information at relatively low cost and increased throughput[1,2]. Although microarrays are now widely used for monitoring transcript expression, hybridization-based technologies have several important limitations[2]. First, low-abundance transcripts cannot be measured accurately. Second, discovery of novel transcripts is limited. Third, direct comparison of transcripts within an individual sample is inaccurate because hybridization kinetics for individual mRNAs are sequence dependent, necessitating ratiometric comparison between paired samples.

To overcome these limitations, researchers have developed digital gene expression (DGE) technologies, such as serial analysis of gene expression (SAGE) and massively parallel signature sequencing (MPSS)[3–8], with the goal of discovering new transcripts as well as performing complete transcriptome profiling over the full dynamic range of cellular mRNA expression. In general, DGE methods use high-throughput sequencing of short cDNA fragments (tags) that are matched to a reference transcriptome to identify the corresponding gene. Individual transcript abundances are then inferred from the relative tag counts for each gene in a 'digital' manner, in contrast to the 'analog' nature of microarray intensity-based quantification. To date, most SAGE-like strategies rely on restriction digestion, adaptor ligation and additional steps. This extensive sample manipulation and the generation of tags from limited sequence contexts per transcript are likely sources of transcript quantification biases[9–13].

Recent studies have demonstrated that high-throughput short-read sequencing platforms can be used to generate high-resolution maps of complete transcriptomes by sequencing a significant fraction of the transcriptome at depth[14–16]. Because these 'RNA-Seq' methods generate variable numbers of reads from each mRNA molecule, extraction of quantitative measurements requires an assessment of coverage depth for each transcript. Although this approach yields informative transcript quantification, it is costly in terms of the sheer number of reads that are required to completely cover an entire transcriptome (several tens of millions of reads per sample). Because the number of reads generated from each transcript is dependent on its length, additional normalization steps are required[15], and quantification accuracy for shorter transcripts is lower[17].

We present a DGE technology based on single-molecule sequencing[18]. Because no amplification is employed, sample preparation does not involve adding adaptors to cDNA, thus enabling a simple procedure free of restriction digestion, ligation or amplification steps. This methodology generates strand-specific, accurate transcript counts covering the complete cellular dynamic range. Single-molecule sequencing DGE (smsDGE) is optimized for mRNA quantification rather than full transcriptome sequencing. The effectiveness of counting by smsDGE is driven by the fact that only a single read is generated from each cDNA molecule, thereby maintaining a faithful representation of transcript distribution. In contrast to RNA-Seq, quantification is independent of transcript length, and sequence read counts are directly proportional to transcript abundance. smsDGE generates sequence reads from the 3′ ends of first-strand cDNA molecules, which usually corresponds to the 5′ ends of mRNAs, depending on the completeness of the reverse transcription (**Fig. 1a**). It does not require

[1]Helicos Biosciences Corporation, Cambridge, Massachusetts, USA. [2]These authors contributed equally to this work. Correspondence should be addressed to T.R. (traz@helicosbio.com).

Received 30 March; accepted 9 June; published online 5 July 2009; doi:10.1038/nbt.1551

Molecular Systems Biology 5; Article number 287; doi:10.1038/msb.2009.45
Citation: *Molecular Systems Biology* 5:287

www.molecularsystemsbiology.com

molecular systems biology

Elucidating regulatory mechanisms downstream of a signaling pathway using informative experiments

Ewa Szczurek[1,2,3,*], Irit Gat-Viks[1,4], Jerzy Tiuryn[3] and Martin Vingron[1]

[1] Computational Molecular Biology Department, Max Planck Institute for Molecular Genetics, Berlin, Germany, [2] International Max Planck Research School for Computational Biology and Scientific Computing, Berlin, Germany and [3] Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Warsaw, Poland
[4] Present address: Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA
* Corresponding author. Computational Molecular Biology Department, Max Planck Institute for Molecular Genetics, Ihnestr. 73, 14195 Berlin, Germany.
Tel.: +49 30 8413 1261; Fax: +49 30 8413 1152; E-mail: szczurek@molgen.mpg.de

Received 15.8.08; accepted 26.5.09

Signaling cascades are triggered by environmental stimulation and propagate the signal to regulate transcription. Systematic reconstruction of the underlying regulatory mechanisms requires pathway-targeted, informative experimental data. However, practical experimental design approaches are still in their infancy. Here, we propose a framework that iterates design of experiments and identification of regulatory relationships downstream of a given pathway. The experimental design component, called MEED, aims to minimize the amount of laboratory effort required in this process. To avoid ambiguity in the identification of regulatory relationships, the choice of experiments maximizes diversity between expression profiles of genes regulated through different mechanisms. The framework takes advantage of expert knowledge about the pathways under study, formalized in a predictive logical model. By considering model-predicted dependencies between experiments, MEED is able to suggest a whole set of experiments that can be carried out simultaneously. Our framework was applied to investigate interconnected signaling pathways in yeast. In comparison with other approaches, MEED suggested the most informative experiments for unambiguous identification of transcriptional regulation in this system.
Molecular Systems Biology **5:** 287; published online 7 July 2009; doi:10.1038/msb.2009.45
Subject Categories: metabolic and regulatory networks; signal transduction
Keywords: experimental design; logical modeling; signal transduction; transcription regulation

Introduction

Revealing the mechanism of transcription regulation in the cell, the interplay of transcription factors and the way they influence their target genes, is a central problem in molecular biology. Diverse approaches have been proposed for the identification of transcriptional regulation based on high-throughput gene expression data (e.g. Akutsu *et al*, 1998; Bussemaker *et al*, 2001; Bolouri and Davidson, 2002; Gardner *et al*, 2003; Segal *et al*, 2003; Nachman *et al*, 2004; Hartemink, 2005). All these methods heavily depend on the available experiments and are prone to the problem of ambiguity in the identification of regulatory relationships. For example, it is possible that a transcription factor remains inactive in all experiments and therefore its targets cannot be revealed. Alternatively, consider two transcription factors located in distinct signaling pathways with a different role, different environmental stimulation and different target genes. In a given set of experiments, if the target genes have similar expression profiles, they will be falsely considered as co-regulated. Moreover, taking any of the two transcription factors as the common regulator of these targets will be equally supported by the experimental data, leading to ambiguous hypothesis about their transcriptional regulation. To avoid such problems, the experiments must generate enough information to draw clear conclusions about regulatory relationships.

In this study, we introduce an algorithm called MEED (model expansion experimental design). MEED is meant to guide experimentalists who focus their research on a chosen signaling pathway and are interested in the regulation of its downstream targets. We assume the researcher has initial qualitative knowledge about the studied pathway and wishes to systematically perturb the pathway components to characterize the gene expression response. Such experimental studies, in which a specific signaling system is perturbed to investigate its downstream regulation mechanisms (rather than global mapping of cellular transcription), became

PERSPECTIVE

Fifteen years of microbial genomics: meeting the challenges and fulfilling the dream

Nikos C Kyrpides

As we approach the completed sequencing of 1,000 microbial genomes, the field of microbial genomics is poised at a crossroads. The future holds great promise for far-reaching advancements in microbiology as well as in diverse, related sciences. But realizing that potential will require meeting the challenges that have accompanied the rapid development of the underlying technology and the exponential growth of data. New technologies provide unprecedented opportunities but also call for conceptual shifts. Experience gained in the first decade of genomics can guide the improved approaches now needed for the selection of genome sequencing projects and their funding, for genome publication and annotation, as well as for data analysis and access. Equipped with these new tools and policies, microbiologists will have a unique opportunity for unprecedented exploration of our microbial planet.

The dramatic advancements in sequencing technology achieved during the past decade have mediated a rapid transition from single-gene to whole-genome studies. In so doing, they also transformed what had been an almost purely experimental discipline into a predominantly theoretical and predictive one[1]. Although not the driving force *per se* behind the development of the technology, microbial genomics was in the forefront of this transition from the very beginning and paved the 'genomics way'[2]. Indicative of this leading role, more than two-thirds of the 4,800 currently reported genome projects[3] are microbial (http://genomesonline.org/), and the same percentage is observed for microbial proteins in the public archive sequence databases from genome sequencing projects[4].

During its first decade, the newly defined field of genomics adopted the fundamentally reductionist perspective that marked twentieth century biology[5]. Accordingly, genome projects were initiated almost exclusively on the basis of potential practical applications for the selected organism, often in the fields of medicine (e.g., pathogenicity or drug targets) or biotech (e.g., bioenergy, agriculture, environmental remediation or industrial production of microbial products). Indeed, just as the human genome project originally set the tone for all genomics, direct practical exploitation of microorganisms set the stage for the microbial sequencing program.

When the boundaries of science are shattered by the sheer force of technological innovation in the absence of a guiding vision, there usually follows a period of time before the affected scientific community realizes that something has gone awry. Such realizations are now arising. Here I present some of the underlying problems and myths that I believe substantially hinder additional growth of the field and, even more importantly, compromise the ability of biologists to use and interpret the available data. Where possible, a solution is proffered, often one whose implementation will necessitate action by the entire community, including scientific journals, sequencing centers and the funding agencies.

Genome publication and data release policy

The policy for publication of complete genomes that we witnessed for most of the first genomics decade has violated a longstanding precept in all scientific fields. Namely, the vast majority of genome papers have been submitted and accepted for publication long before the public release of the sequenced data. As a result, reviewers, unable to examine the actual data, could evaluate such papers only on faith and trust, thus undermining the peer-review process in this field. Although no genome paper has been retracted, subsequent analyses have often revealed fundamental flaws in both the derived conclusions and the data itself[6,7].

Granted, annotation is generally considered a never-ending bioinformatics adventure; closure is achieved only when all the functions and all the genes of an organism have been experimentally verified. Still, a fine degree of separation exists between what might be considered an acceptable error due to the incomplete adventure[8] and what is essentially an incorrect and misleading conclusion. Charging the reviewers with the responsibility for making this distinction calls for a strict policy from the publishing journals requiring the actual sequence data to be publicly released well in advance of publication of the genome. Moreover, simply providing the sequence files would not suffice, as most scientists cannot make much sense out of a GenBank file. Rather, the data should be provided to the community in a meaningful way that facilitates cogent analysis and evaluation. This can be accomplished only through data management systems that support comparative genome analysis. Several such systems are already freely available in the community[9].

Over time, as the substantial benefits of prepublication release of genome data have been recognized, many funding agencies and most of the large sequencing centers now adhere to the rapid data release policy set forth as the Bermuda Principles in 1996 and renewed in 2003 (http://www.genome.gov/page.cfm?pageID=10506376). Thus, over the past few years, we have witnessed an increasing number of complete genomes released in GenBank without accompanying publications[10].

Owing to the exponential increase in the number of completed genome sequences, coupled with frequent phylogenetic redundancy of

Genome Biology Program, DOE Joint Genome Institute, Walnut Creek, California, USA. Correspondence should be addressed to N.C.K. (nckyrpides@lbl.gov).

Published online 8 July 2009; doi:10.1038/nbt.1552

Lab on a Chip RSCPublishing
Journals · Lab on a Chip Advance Articles DOI: 10.1039/b909386a

Lab Chip, 2009 | DOI: 10.1039/b909386a | Communication

Beating Poisson encapsulation statistics using close-packed ordering†

Adam R. Abate [a], Chia-Hung Chen [b], Jeremy J. Agresti [a] and David A. Weitz *[a]

[a]*School of Engineering and Applied Sciences/Department of Physics, Harvard University, Cambridge, Massachusetts, USA. E-mail: weitz@seas.harvard.edu; Fax: +1 617 4953275; Tel: +1 617 496 2842*

[b]*Cavendish Laboratory, University of Cambridge, Cambridge, UK CB3 0HE*

Received 12th May 2009 , Accepted 24th July 2009

First published on the web 28th July 2009

Loading drops with discrete objects, such as particles and cells, is often necessary when performing chemical and biological assays in microfluidic devices. However, random loading techniques are inefficient, yielding a majority of empty and unusable drops. We use deformable particles that are close packed to insert a controllable number of particles into every drop. This provides a simple, flexible means of efficiently encapsulating a controllable number of particles per drop.

Drops formed in microfluidic devices are useful for chemical and biological assays.[1–6] The drops can serve as picoliter vessels within which individual reactions can be performed. With microfluidic devices, the drops can be formed, merged, and sorted at kilohertz rates.[7] This combination of speed, containment, and small volumes is very useful for many applications, such as screening libraries of unknown chemical compounds, evolving cells and enzymes, and analyzing genetic material.[8–14] All such applications require the encapsulation of cells, beads, and other discrete reagents in the drops. However, current methods to encapsulate objects into drops are very inefficient.[15] Typically, this is accomplished by diluting suspensions of the materials and encapsulating into the drops at random; the resulting Poisson statistics lead to a large number of empty drops with a much smaller number having a single particle. This can be wasteful, negating the speed and efficiency afforded by droplet microfluidics.

This inefficiency has stimulated the development of new methods that provide more efficient encapsulation.[16] For example, a laser can be used to guide particles, ensuring single particle encapsulation.[17] This method affords superb control for the encapsulation of objects over a range of sizes – from beads tens of microns in diameter to cells and cellular organelles only a few microns in diameter. However, this method is very difficult to use, requiring sophisticated optical equipment and extensive participation by the user. Moreover, it is slow, having a maximum speed of only a few hertz. Alternatively, inertial ordering can be used to passively organize particles

Lab on a Chip RSCPublishing
Journals Lab on a Chip Contents List DOI: 10.1039/b821021g

Lab Chip, 2009, **9**, 1707-1713 | DOI: 10.1039/b821021g | Paper

Simultaneous measurement of reactions in microdroplets filled by concentration gradients†

Nicolae Damean [a], Luis F. Olguin [b], Florian Hollfelder [b], Chris Abell [a] and Wilhelm T. S. Huck *[a]

[a] *Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK. E-mail: wtsh2@cam.ac.uk*

[b] *Department of Biochemistry, Tennis Court Rd, Cambridge, CB2 1QW, UK*

Received 27th November 2008 , Accepted 23rd February 2009

First published on the web 19th March 2009

This work describes a technology for performing and monitoring simultaneously several reactions confined in strings of microdroplets having identical volumes but different composition, and travelling with the same speed in parallel channels of a microfluidic device. This technology, called parallel microdroplets technology (PμD), uses an inverted optical microscope and a charge-coupled device (CCD) camera to collect images and analyze them so as to report on the reactions occurring in these microdroplets. A concentration gradient of one reactant is created in the microfluidic device. In each channel, a different concentration of this reactant is mixed with a fixed amount of a second reactant. Using planar flow-focusing methodology, these mixtures are confined in microdroplets of pL size which travel in oil as continuous medium, avoiding laminar dispersion. By analyzing the images of parallel strings of microdroplets, the time courses of several reactions with different reagent compositions are investigated simultaneously. In order to design the microfluidic device that consists in a complex network of channels having well-defined geometries and restricted positions, the theoretical concept of equivalent channels (*i.e.* channels having identical hydraulic resistance) is exploited and developed. As a demonstration of the PμD technology, an enzyme activity assay was carried out and the steady-state kinetic constants were determined.

1. Introduction

This paper describes a microsystem-based technology that compartmentalizes and measures simultaneously different reactions in pL volumes. This technology, which we call parallel microdroplets technology (PμD), analyses optically a set of reactions confined in strings of water-in-oil microdroplets that have identical volumes (5 to 60 pL), travel with identical speeds (0.25 to 6 mm s^{-1}) and contain different reactant concentrations. The technology provides a flexible method for analyzing a number of reactions simultaneously and continuously.

Microdroplets technology is emerging as a robust technology for performing individual experiments in nL–pL volumes.[1–4] This technology offers consistent compartmentalization (identical microdroplets under identical conditions), avoids cross-contamination of the samples to be analyzed by minimal manipulation of reactants and uses small quantities of reactants which are transported without dispersion. It has applications in chemistry and biology such as enzyme kinetics,[5–9] cell based-assays,[10,11] *in vitro* protein expression[12] and mass production of monodisperse particles.[13]

Despite the importance of performing and analyzing experiments simultaneously, no such

Lab on a Chip RSCPublishing
Journals Lab on a Chip Contents List DOI: 10.1039/b821678a

Lab Chip, 2009, **9**, 1714-1720 | DOI: 10.1039/b821678a | Paper

Parallel single-cell light-induced electroporation and dielectrophoretic manipulation

Justin K. Valley *, Steven Neale , Hsan-Yin Hsu , Aaron T. Ohta , Arash Jamshidi and Ming C. Wu

Berkeley Sensor and Actuator Center, Department of Electrical Engineering and Computer Science, University of California Berkeley, 497 Cory Hall, Berkeley, CA 94720, USA. E-mail: valleyj@eecs.berkeley.edu; Fax: +1 510 643 5817; Tel: +1 510 642 1023

Received 3rd December 2008 , Accepted 20th February 2009

First published on the web 13th March 2009

Electroporation is a common technique for the introduction of exogenous molecules across the, otherwise, impermeant cell membrane. Conventional techniques are limited by either low throughput or limited selectivity. Here we present a novel technique whereby we use patterned light to create virtual electrodes which can induce the parallel electroporation of single cells. This technique seamlessly integrates with optoelectronic tweezers to provide a single cell manipulation platform as well. We present evidence of parallel, single cell electroporation using this method through use of fluorescent dyes and dielectrophoretic responses. Additionally, through the use of integrated microfluidic channels, we show that cells remain viable following treatment in the device. Finally, we determine the optimal field dosage to inject propidium iodide into a HeLa cell and maintain cellular viability.

Introduction

There has been an increasing amount of interest in the past decade in creating a system capable of performing single cell based assays for a variety of applications. One interesting application involves the creation of a chip with integrated cell membrane poration functionality. The ability to introduce foreign molecules into the intra-cellular space is important in applications ranging from genetic transfection to the study of cell-to-cell signaling.[1,2]

One of the most common membrane poration methods is electroporation. Temporary permeation of the cellular membrane is achieved in electroporation by subjecting the cell to an

Lab on a Chip RSCPublishing
Journals Lab on a Chip Contents List DOI: 10.1039/b901635j

Lab Chip, 2009, **9**, 1831-1835 | DOI: 10.1039/b901635j | Communication

A chip-to-chip nanoliter microfluidic dispenser†

Jianbin Wang‡§ [a], Ying Zhou‡ [a], Haiwei Qiu¶ [a], Huang Huang [b], Changhong Sun [b], Jianzhong Xi *[b] and Yanyi Huang *[a]

[a]*Department of Advanced Materials and Nanotechnology, College of Engineering, Peking University, Beijing, 100871, China. E-mail: yanyi@pku.edu.cn*

[b]*Department of Biomedical Engineering, College of Engineering, Peking University, Beijing, 100871, China. E-mail: jzxi@pku.edu.cn*

Received 26th January 2009 , Accepted 25th March 2009

First published on the web 23rd April 2009

A high-throughput microfluidic device is developed to handle liquid dispensation in nanoliter range. The dispenser system shows no cross-contamination between the microwells, indicating its great potential in large-scale screening experiments. An array of 115 nl PCR reactions, as well as the single channel addressable chip demonstrate the high flexibility and wide applications of this novel system.

Most biochemical reactions and cell based chemical assays, especially the large-scale screening experiments and single cell studies, are performed in liquid phase with a volume range from nanoliters to microliters.[1–5] A small volume reaction not only reduces the cost for each reaction, but also contributes to the increase of the reagent concentration for an efficient reaction, and generates much less waste. These experiments are usually laborious, most of them requiring numerous times of pipetting actions to transfer the liquid samples. The accurate small-volume sample dispensation, which may be easily achieved by commercially available instruments in the range of microliters but difficult to scale down to nanoliter range, remains one of the major challenges in high-throughput reactions. On the other hand, a few alternative methods, including non-contact liquid-jet printing[6–8] and ultrasonic droplets generation,[9] have been introduced to liquid transfer. But most of these technologies deal with liquid in picoliter range. They are expensive and hard to be integrated into a highly parallel way for higher throughput fashion. Microfluidic technology shows promising potentials for accurate liquid manipulation, making itself an ideal platform for nanoliter scale reactions.[10–13] A high-throughput miniaturized reaction array

IMAGING AND VISUALIZATION

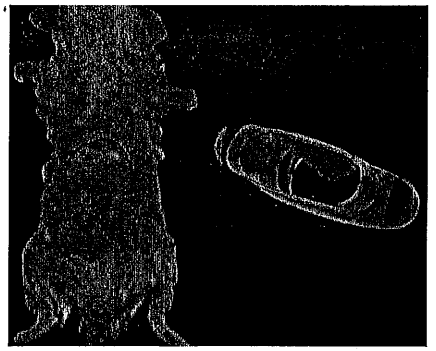
Fluorescent proteins: into the infrared

An engineered infrared fluorescent protein is the first member of a new class of genetically encodable probes, with special advantages over visible-wavelength fluorescent proteins for *in vivo* imaging.

Fluorescent proteins have come along way since the discovery of GFP from the jellyfish *Aequorea victoria* in the early 1960s. Today, both natural and engineered fluorescent proteins come in a veritable rainbow of colors, largely thanks to pioneering work done in Roger Tsien's laboratory at the University of California at San Diego, work that was recently recognized with part of the 2008 Nobel Prize in Chemistry.

In combination with fluorescence microscopy, genetically encodable fluorescent proteins are used to highlight various cell structures and monitor cellular processes in technicolored splendor. Fluorescent proteins are also useful for tracking cell populations, such as tumor cells, *in vivo*. Thus far, the best tools for imaging in live, whole animals have been far-red fluorescent proteins because both autofluorescence and light absorption by the animal's tissue are minimized at the far-red end of the visible spectrum. However, to produce far-red fluorescence, these proteins need to be excited by light at around or below 600 nanometers, which is substantially attenuated by hemoglobin before reaching deeply buried tissues. Despite continued improvement, so far there has been no success in developing a fluorescent protein with an excitation maximum around 700 nm, the absorption minimum of hemoglobin.

The Tsien lab recently took a new approach to develop a better fluorescent protein for *in vivo* imaging. Rather than begin with a red fluorescent protein scaffold and engineer it to make it brighter or shift the excitation further to the far-red, they decided to start fresh with a new scaffold to develop an infrared-emitting fluorescent protein. The researchers began with a bacterial phytochrome scaffold from *Deinococcus radiodurans*, which absorbs light at around 700 nm. This phytochrome binds a cofactor called biliverdin, which serves as a chromophore and just so happens to be found naturally in all animals. "The truncated phytochrome crystal structure was available, which facilitated protein engineering," explains Xiaokun Shu, a postdoc and the first author on this work. "Our goal with structure-based engineering was to rigidify



Mouse liver in a living mouse labeled with an infrared fluorescent protein (represented by red pseudocolor) delivered via an adenoviral vector. A tilted view of the liver is shown on the right. Figure courtesy of Xiaokun Shu and Roger Tsien.

BIOINFORMATICS

GOING WITH THE SKEWED FLOW

Computational and experimental biologists teamed up to develop a new software tool to analyze the rich data generated by new and powerful flow cytometers.

Imagine watching the flow of the East Australian Current; how would you describe all the fish swimming in front of you or find a rare species? Biologists face a similar situation when they look at blood flow. But unlike the tropical fish that come in various sizes, shapes and colors, most blood cells can only be distinguished by their surface protein markers.

Immunologists have used flow-cytometry technology for studying blood cells for decades. A flow cytometer analyzes the cells flowing through a fluorescence detector one by one and records the surface-marker profile of each cell. Recent advances in new fluorescent dyes and detection techniques have allowed simultaneous detection of close to 20 different markers. However, new computational methods are needed to properly analyze these rich data. At the Broad Institute of the Massachusetts Institute of Technology, Jill Mesirov, a computational biologist, recently collaborated with two experimental biology groups led by Philip De Jager and David Hafler to develop a software tool called flow analysis with automated multivariate estimation (FLAME), for such flow-cytometric data analysis.

"The traditional way is to look at the data two or three markers at a time and manually identify cell subpopulations, but when you have 12 different markers, you can't look at 12-dimensional space," says Mesirov. The current approach is inefficient and subjective. Worse, explains De Jager, "if you don't know what you are looking for, you get more and more limited in the successive two- or three-dimensional projections and fail to recognize the architecture of the whole cell population."

Two important characteristics of flow-cytometric data, asymmetric distribution and outliers, complicate the analysis. "The histograms that come out of a flow cytometer always have tails, which are not captured very well if one assumes symmetric distribution," says De Jager. "And the outlier can be also quite important," adds Mesirov. Others have used symmetric modeling or data transformation to analyze high-dimensional flow-cytometric data. However, symmetric modeling does not fit the flow-cytometric data, and "the problem of data transformation is that very different asymmetric distributions can, after transformation, yield the same Gaussian distribution," explains Mesirov.

To model robustly against such asymmetry and outliers for precise identification of subpopulations from the high-dimensional data, the authors used a non-Gaussian statistical model based on the multivariate skewed *t* distribution. "Mathematically it was obvious that a new modeling approach was needed" for flow-cytometric data, says Mesirov. A postdoc in her laboratory, Saumyadipta Pyne, designed the high-dimensional mixture model and then worked with another postdoc, Kui Wang, in Geoffrey McLachlan's lab at the University

'Injecting' yeast

Daniel Riveline[1,2] & Paul Nurse[1]

Yeast is a powerful genetic model system, but its rigid cell wall has prohibited microinjection. Using microfabricated channels to constrain the fission yeast *Schizosaccharomyces pombe*, we sheared local regions of individual cells with a piezoelectric unit. The cells remained viable, we detected actin patches in the cell after introduction of fluorescent phalloidin into the medium, and the cytokinetic ring was disrupted after injection of the myosin II inhibitor blebbistatin.

Both the budding and fission yeasts are powerful models for the genetic analysis of eukaryotic cell biology. However, their rigid cell wall has prevented the use of microinjection, a productive method in other systems. For example, the functions of small Rho GTPases had been discovered using injection of active GTPases into fibroblasts[1], and the phenomenon of RNA interference had been investigated using injection of RNA into *Caenorhabditis elegans*[2]. Here we report a procedure that allows introduction of molecules into the fission yeast *Schizosaccharomyces pombe* by an 'injection'-like process.

The fission yeast cell is a rod, 15 μm in length and 4 μm in diameter, with a rigid carbohydrate cell wall. Cells grow by elongation from the hemispheric ends and divide by medial fission. Initial experiments established that yeast cells must be immobilized for effective injection. Mammalian cells attached to a surface substratum remain fixed during injection, but *S. pombe* cells became detached even when fixed to adhesive lectin-coated surfaces.

To solve this problem, we prepared topographical patterned substrates to trap cells in channels (Online Methods). We prepared these chambers using standard soft lithography methods[3] with polydimethylsiloxane (PDMS) and a commercially available electron microscopy grid. A negative replica made of 5 μm wide bars allowed large numbers of channels to be prepared in one step (**Fig. 1a**). Division time in cells held within the PDMS substrate was unchanged.

However, even with yeast cells efficiently immobilized, it was difficult to penetrate the cell wall using conventional injection methods. To solve this problem, we used a piezo-impact micromanipulator, previously used to pierce biological barriers such as in intracytoplasmic sperm injection[4]. The goal was to create a hole in the rigid fission yeast cell by a local shear force. We pressed the flexible tip of a pipette against the grid pattern in the proximity of the cell and perpendicularly to its long axis and then pushed against the cell[5] (**Fig. 1b**). This procedure induced a mechanical stress, buckling the cell. Then we activated the piezo machine for ~10 s. Materials flowing out of the vibrating pipette could enter the cell locally (Online Methods). The surrounding medium contained 1.2 M sorbitol as an osmotic stabilizer to prevent cell lysis. After the manipulation, we removed the pipette. Alternatively, we placed the materials to be injected in the medium and repeated the same procedure, except we applied the shear by using the motorized manipulator. Note that our method is not a classical injection because the tip of the pipette is not introduced into the cell.

Injection of fluorescent phalloidin has been widely used for studying actin cytoskeletal structures[6–8]. We demonstrated the effectiveness of our approach by 'injecting' Alexa Fluor 488–labeled phalloidin, which binds actin irreversibly on the timescale of our experiments. We visualized actin patches, which indicated that the

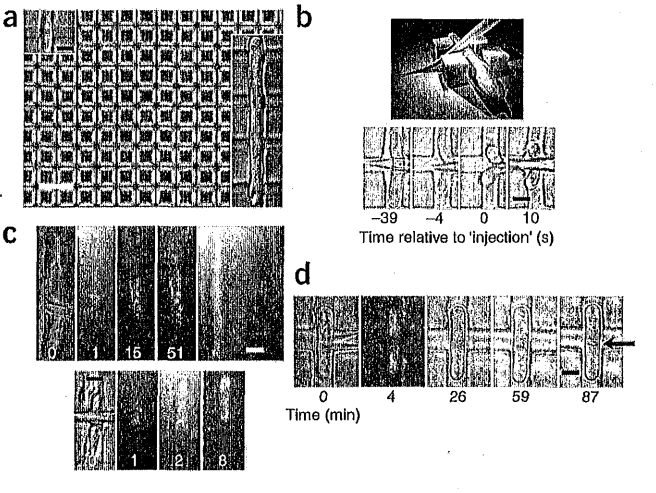


Figure 1 | Injecting a trapped yeast. (**a**) The topographical pattern. Insets show a haploid fission yeast cell (top left) and an elongated fission yeast cell (right, *cdc13*$^+$ switch-off). (**b**) Schematic (top) and phase-contrast images (bottom) of the 'injection' procedure. Pressure of the pipette bends the cell and the piezo-micromanipulator is then activated (at time 0 s) for about 10 s. (**c**) Merged phase-contrast and fluorescence images of *cdc13*$^+$ switch-off cells (top) and wild-type diploid cells (bottom) after 'injection' of Alexa Fluor 488–labeled phalloidin. Note that the procedure is cell-specific, because a neighboring cell (top right) and the sister cell (bottom right) are not labeled. Numbers indicate time relative to injection in minutes. (**d**) Phase contrast (0, 26, 59 and 87 min) and fluorescence (4 min) images of an injected diploid cell. A septum (arrow) forms after injection of a low dose of Alexa Fluor 488–labeled phalloidin (at time 0 min). Scale bars, 25 μm (5 μm in inset) (**a**) and 5 μm (**b–d**).

[1]Laboratory of Yeast Genetics and Cell Biology, The Rockefeller University, New York, New York, USA. [2]Laboratoire de Spectrométrie Physique, Centre National de la Recherche Scientifique (CNRS), Unité Mixte de Recherche 5588, Université Joseph Fourier, Saint-Martin-d'Hères, France. Correspondence should be addressed to D.R. (daniel.riveline@ujf-grenoble.fr).

RECEIVED 29 DECEMBER 2008; ACCEPTED 17 MARCH 2009; PUBLISHED ONLINE 7 JUNE 2009; DOI:10.1038/NMETH.1335

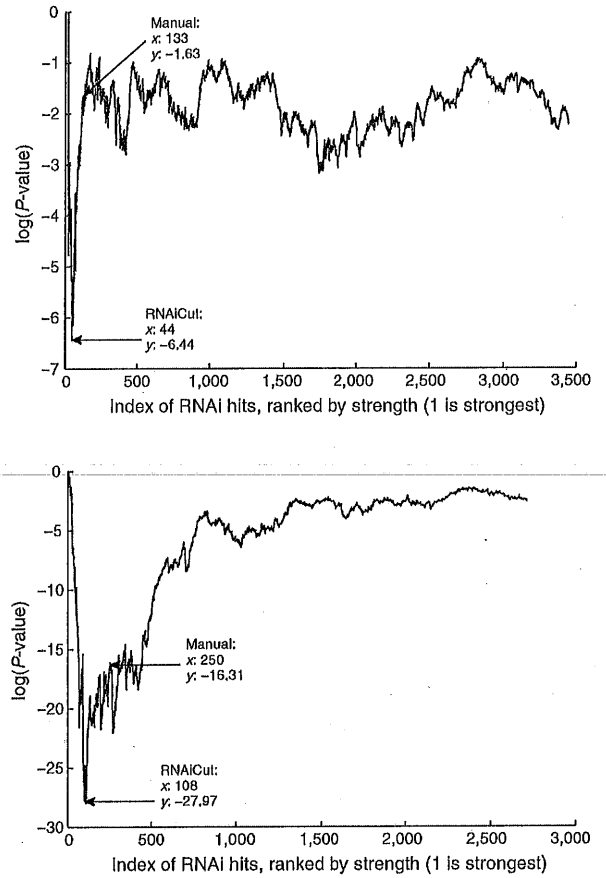


Figure 1 | RNAiCut results for insulin-triggered MAPK pathway screen in *D. melanogaster*[4]. Genes with positive (top) and negative (bottom) *Z* scores in the screen are ordered on the *x* axes from left to right based on the decreasing magnitude of *Z* scores. The *y* axis denotes the *P*-value, as a function of *k*, of finding a random PPI subnetwork as well connected as the one containing the *k* highest-scoring genes from the RNAi screen.

the core signaling pathway. RNAiCut was robust to *Z*-score noise generated by randomly scrambling close *Z* scores (Supplementary Fig. 11 and Supplementary Table 5).

We offer an online server (http://rnaicut.csail.mit.edu) for interpreting functional genomic experiments. Although we developed RNAiCut using a fly PPI network, RNAiCut can also be run on non-fly and non-PPI networks (Supplementary Fig. 12). This tool will help functional genomics research by enabling hit-list gene selection using orthogonal datasets.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank J. Bienkowska, M. Booker and K.G. Lu. I.M.K. was partially supported by the John Reed Fund.

Irene M Kaplow[1,6], Rohit Singh[2,6], Adam Friedman[3–5], Chris Bakal[2–4], Norbert Perrimon[3,4] & Bonnie Berger[1,2,5]

[1]Department of Mathematics and [2]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [3]Department of Genetics and [4]Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts, USA. [5]Harvard–Massachusetts Institute of Technology Division of Health Sciences and Technology, Boston, Massachusetts, USA. [6]These authors contributed equally to this work.
e-mail: perrimon@receptor.med.harvard.edu or bab@mit.edu

1. Salwinski, L. *et al. Nucleic Acids Res.* **32**, D449–D451 (2004).
2. Stark, C. *et al. Nucleic Acids Res.* **34**, D535–D539 (2006).
3. König, R. *et al. Cell* **135**, 49–60 (2008).
4. Friedman, A. & Perrimon, N. *Nature* **444**, 230–234 (2006).
5. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).

Enabling IMAC purification of low abundance recombinant proteins from *E. coli* lysates

To the Editor: Currently, the most widely used method for purifying recombinant proteins for biochemical and especially structural studies is immobilized metal affinity chromatography (IMAC), in which a metal-binding polyhistidine tag (His tag) serves as a small purification handle on the target protein. IMAC is a powerful and generic purification method, with high recovery yields and low costs. Additionally, the His tag is compatible with most downstream applications because it is small and relatively inert[1,2]. *Escherichia coli* is by far the most popular expression host owing to its supremacy regarding cost, biomass production and technical simplicity[3,4]. However, a serious drawback of IMAC is the often-experienced failure to purify low-abundance His-tagged proteins from *E. coli* lysates; increasing the culture size and thereby increasing the amount of available His-tagged protein does not result in increased yield. We examined this issue and propose that it is tightly linked to metal-ion leakage from the columns induced by the *E. coli* lysate.

We used His-tagged GFP (His_6-GFP) to examine the effect of *E. coli* lysate on the protein binding capacity of IMAC columns. Application of the soluble fraction of *E. coli* lysate lacking recombinant protein expression to a 1 ml HiTrap Chelating HP column (GE Healthcare) partly loaded with His_6-GFP, caused extensive migration of His_6-GFP whereas application of wash buffer did not (Supplementary Fig. 1a). We confirmed this using different column materials and concluded that *E. coli* lysate severely reduces the binding capacity of the column (data not shown). By separating a lysate into high- and low-molecular-weight components we found that the reduced binding capacity was brought about by low-molecular-weight components, and not high-molecular-weight components (Supplementary Fig. 1b), implying that the underlying cause for the reduced target protein binding is not the result of native *E. coli* proteins competing with the His-tagged protein for the immobilized nickel-ion binding sites. We determined the amount of nickel present on the different columns before and after sample load and found that the decrease in binding capacity correlated with loss of immobilized nickel ions from the column (Supplementary Fig. 1c).

IMAC is very sensitive to the presence of metal chelators[1], and the *E. coli* lysate contains many unspecific weak chelators such as dicarboxylic acids from the citric acid cycle. Under stress conditions, *E. coli* can also produce highly specific metal chelators, metallophores[5]. We speculated that such metallophores, if produced, would be mainly associated with the periplasmic space of *E. coli* but not with the cytosol. We therefore hypothesized that removing the periplasmic material before cell lysis could improve His-tagged recombinant protein purification yields. We subjected *E. coli* cells to osmotic shock to remove the periplasmic material before cell lysis (Supplementary Methods). His_6-GFP did not migrate substantially on IMAC columns treated with lysate devoid of periplasmic

Osmotically driven flows in microchannels separated by a semipermeable membrane

Kåre Hartvig Jensen,[a] Jinkee Lee,[b] Tomas Bohr[c] and Henrik Bruus†*[a]

Received 24th October 2008, Accepted 25th March 2009
First published as an Advance Article on the web 20th April 2009
DOI: 10.1039/b818937d

We have fabricated lab-on-a-chip systems with microchannels separated by integrated membranes allowing for osmotically driven microflows. We have investigated these flows experimentally by studying the dynamics and structure of the front of a sugar solution travelling in 200 μm wide and 50–200 μm deep microchannels. We find that the sugar front travels at a constant speed, and that this speed is proportional to the concentration of the sugar solution and inversely proportional to the depth of the channel. We propose a theoretical model, which, in the limit of low axial flow resistance, predicts that the sugar front should indeed travel with a constant velocity. The model also predicts an inverse relationship between the depth of the channel and the speed, and a linear relation between the sugar concentration and the speed. We thus find good qualitative agreement between the experimental results and the predictions of the model. Our motivation for studying osmotically driven microflows is that they are believed to be responsible for the translocation of sugar in plants through the phloem sieve element cells. Also, we suggest that osmotic elements can act as on-chip integrated pumps with no movable parts in lab-on-a-chip systems.

I. Introduction

Osmotically driven flows are believed to be responsible for the translocation of sugar in plants, a process that takes place in the phloem sieve element cells.[1] These cells form a micro-fluidic network which spans the entire length of the plant measuring from 10 μm in diameter in small plants to 100 μm in diameter in large trees.[1] The mechanism driving these flows is believed to be the osmotic pressures that build up relative to the neighboring water-filled tissue in response to loading and unloading of sugar into and out of the phloem cells in different parts of the plant.[1] This mechanism, collectively called the pressure-flow hypothesis, is much more efficient than diffusion, since the osmotic pressure difference caused by a difference in sugar concentration creates a bulk flow directed from large concentrations to small concentrations, in accordance with the basic needs of the plant.

Experimental verification of flow rates in living plants is difficult,[2] and the experimental evidence from artificial systems backing the pressure-flow hypothesis is scarce and consists solely of results obtained with centimetric sized setups.[3–5] However, many theoretical and numerical studies of the sugar translocation in plants have used the pressure-flow hypothesis[6–8] with good results. To verify that these results are indeed valid, we believe that it is of fundamental importance to conduct a systematic survey of osmotically driven flows at the micrometre scale. Finally, osmotic flows in microchannels can act as migration enhancers[9] or as microscale on-chip pumps with no movable parts. Examples of previous off-chip osmotic pumps are the device developed by Park *et al.*[10] and the osmotic pills developed by Shire Laboratories and pioneered by Theeuwes.[11] Also, there is a direct analogy between osmotically driven flows powered by concentration gradients, and electroosmotically driven flows in electrolytes[12,13] powered by electrical potential gradients.

II. Experimental setup

A. Chip design and fabrication

To study osmotically driven flows in microchannels, we have designed and fabricated a microfluidic system consisting of two layers of 1.5 mm thick polymethyl methacrylate (PMMA) separated by a semipermeable membrane (Spectra/Por Biotech Cellulose Ester dialysis membrane, MWCO 3.5 kDa, thickness ~40 μm), as sketched in Fig. 1(a)–(d). Channels of length 27 mm, width 200 μm and depth 50–200 μm were milled in the two PMMA layers by use of a MiniMill/Pro3 milling machine.[14,15] The top channel contains partly the sugar solution, and partly pure water, while the bottom channel always contains only pure water. To facilitate the production of a steep concentration gradient by cross-flows, a 200 μm wide cross-channel was milled in the upper PMMA layer perpendicular to and bi-secting the main channel. Inlets were produced by drilling 800 μm diameter holes through the wafer and inserting brass tubes into these. By removing the surrounding material, the channel walls in both the top and bottom layers acquired a height of 100 μm and a width of 150 μm. After assembly, the two PMMA layers were positioned such that the main channels in either layer were facing each other. Thus, when clamping the two layers together using two

[a]*Center for Fluid Dynamics, Department of Micro- and Nanotechnology, Technical University of Denmark, DTU Nanotech Building 345 East, DK-2800 Kongens Lyngby, Denmark*
[b]*School of Engineering and Applied Science, Harvard University, Cambridge, Massachusetts, 02138, USA*
[c]*Center for Fluid Dynamics, Department of Physics, Technical University of Denmark, DTU Physics Building 309, DK-2800 Kongens Lyngby, Denmark*
† *E-mail:* bruus@nanotech.dtu.dk

Protein Quantification in Complex Mixtures by Solid Phase Single-Molecule Counting

Lee A. Tessler,[†] Jeffrey G. Reifenberger,[‡] and Robi D. Mitra*,[†]

Center for Genome Sciences, Department of Genetics, Washington University in St. Louis School of Medicine, 4444 Forest Park Avenue, St. Louis, Missouri 63108, and Helicos Biosciences, One Kendall Square, Cambridge, Massachusetts 02139

Here we present a procedure for quantifying single protein molecules affixed to a surface by counting bound antibodies. We systematically investigate many of the parameters that have prevented the robust single-molecule detection of surface-immobilized proteins. We find that a chemically adsorbed bovine serum albumin surface facilitates the efficient detection of single target molecules with fluorescent antibodies, and we show that these antibodies bind for lengths of time sufficient for imaging billions of individual protein molecules. This surface displays a low level of nonspecific protein adsorption so that bound antibodies can be directly counted without employing two-color coincidence detection. We accurately quantify protein abundance by counting bound antibody molecules and perform this robustly in real-world serum samples. The number of antibody molecules we quantify relates linearly to the number of immobilized protein molecules (R^2 = 0.98), and our precision (1–5% CV) facilitates the reliable detection of small changes in abundance (7%). Thus, our procedure allows for single, surface-immobilized protein molecules to be detected with high sensitivity and accurately quantified by counting bound antibody molecules. Promisingly, we can probe flow cells multiple times with antibodies, suggesting that in the future it should be possible to perform multiplexed single-molecule immunoassays.

Our ability to detect and quantify proteins has lagged behind our ability to analyze nucleic acids. Closing this gap by developing more sensitive and quantitative protein analysis methods would greatly aid efforts to understand cellular processes[1,2] and to search for protein biomarkers that reveal disease state.[3,4] The application of single-molecule detection (SMD) methods to proteins holds great promise in this regard for five reasons: (1) Recent advances have made SMD methods inexpensive, robust, and reliable.[5,6] (2) SMD methods can enable the detection of low-abundance proteins,[7–9] which is especially important because the poor sensitivities of current proteomic methods are limiting progress in the area of biomarker discovery.[10,11] (3) SMD methods can enable protein quantification by employing single-molecule counting, which can be significantly more accurate than bulk methods.[12,13] (4) SMD methods can enable analysis of protein–protein interactions by detecting single-molecule colocalization.[14] (5) SMD methods for proteins affixed to a surface could enable highly multiplexed immunoassays. For example, by creating ~20 overlapping pools of labeled antibodies using a logarithmic pooling strategy like the one used to decode bead-based random microarrays,[15] a single assay could detect the protein targets of all 6 000 nonredundant human proteome antibodies[16] with only ~20 binding rounds.

There are several obstacles that have hampered the development of single-molecule immunoassays. One is the lack of a good surface for the SMD of surface-immobilized proteins. An ideal surface would be resistant to nonspecific antibody adsorption, while still allowing for the specific binding of antibodies to their target molecules. Efforts have been made to develop better surfaces,[17–28] however, the nonspecific adsorption on these surfaces has not been characterized with single-molecule resolution, with a few exceptions.[18,19] To work around high background

* Corresponding author. E-mail: rmitra@genetics.wustl.edu. Phone: (314)-362-2751. Fax: (314) 362-2156.

[†] Washington University in St. Louis School of Medicine.

[‡] Helicos Biosciences.

(1) Ghaemmaghami, S.; Huh, W.; Bower, K.; Howson, R. W.; Belle, A.; Dephoure, N.; O'Shea, E. K.; Weissman, J. S. *Nature* **2003**, *425*, 737–741.

(2) Cohen, A. A.; Geva-Zatorsky, N.; Eden, E.; Frenkel-Morgenstern, M.; Issaeva, I.; Sigal, A.; Milo, R.; Cohen-Saidon, C.; Liron, Y.; Kam, Z.; Cohen, L.; Danon, T.; Perzov, N.; Alon, U. *Science* **2008**, *322*, 1511–1516.

(3) Omenn, G. S.; States, D. J.; Adamski, M.; Blackwell, T. W.; Menon, R.; Hermjakob, H.; Apweiler, R.; Haab, B. B.; Simpson, R. J.; Eddes, J. S.; Kapp, E. A.; Moritz, R. L.; Chan, D. W.; Rai, A. J.; Admon, A.; Aebersold, R.; Eng, J.; Hancock, W. S.; Hefta, S. A.; Meyer, H.; Paik, Y. K.; Yoo, J. S.; Ping, P. P.; Pounds, J.; Adkins, J.; Qian, X. H.; Wang, R.; Wasinger, V.; Wu, C. Y.; Zhao, X. H.; Zeng, R.; Archakov, A.; Tsugita, A.; Beer, I.; Pandey, A.; Pisano, M.; Andrews, P.; Tammen, H.; Speicher, D. W.; Hanash, S. M. *Proteomics* **2005**, *5*, 3226–3245.

(4) Anderson, L. *J. Physiol. (London, U.K.)* **2005**, *563*, 23–60.

(5) Harris, T. D.; Buzby, P. R.; Babcock, H.; Beer, E.; Bowers, J.; Braslavsky, I.; Causey, M.; Colonell, J.; Dimeo, J.; Efcavitch, J. W.; Giladi, E.; Gill, J.; Healy, J.; Jarosz, M.; Lapen, D.; Moulton, K.; Quake, S. R.; Steinmann, K.; Thayer, E.; Tyurina, A.; Ward, R.; Weiss, H.; Xie, Z. *Science* **2008**, *320*, 106–109.

(6) Roy, R.; Hohng, S.; Ha, T. *Nat. Methods* **2008**, *5*, 507–516.

(7) Sauer, M.; Zander, C.; Muller, R.; Ullrich, B.; Drexhage, K. H.; Kaul, S.; Wolfrum, J. *Appl. Phys. B: Lasers Opt.* **1997**, *65*, 427–431.

(8) Li, L.; Tian, X.; Zou, G.; Shi, Z.; Zhang, X.; Jin, W. *Anal. Chem.* **2008**, *80*, 3999–4006.

(9) Loscher, F.; Bohme, S.; Martin, J.; Seeger, S. *Anal. Chem.* **1998**, *70*, 3202–3205.

(10) Moul, J. W. *Clin. Prostate Cancer* **2003**, *2*, 87–97.

(11) Munkarah, A.; Chatterjee, M.; Tainsky, M. A. *Curr. Opin. Obstet. Gynecol.* **2007**, *19*, 22–26.

(12) Wold, B.; Myers, R. M. *Nat. Methods* **2008**, *5*, 19–21.

(13) Marioni, J. C.; Mason, C. E.; Mane, S. M.; Stephens, M.; Gilad, Y. *Genome Res.* **2008**, *18*, 1509–1517.

(14) Wallrabe, H.; Periasamy, A. *Curr. Opin. Biotechnol.* **2005**, *16*, 19–27.

(15) Gunderson, K. L.; Kruglyak, S.; Graige, M. S.; Garcia, F.; Kermani, B. G.; Zhao, C. F.; Che, D. P.; Dickinson, T.; Wickham, E.; Bierle, J.; Doucet, D.; Milewski, M.; Yang, R.; Siegmund, C.; Haas, J.; Zhou, L. X.; Oliphant, A.; Fan, J. B.; Barnard, S.; Chee, M. S. *Genome Res.* **2004**, *14*, 870–877.

10.1021/ac901068x CCC: $40.75 © XXXX American Chemical Society

Anal. Chem. **2009,** *81,* 5111–5118

analytical chemistry
feature

Droplets for Ultrasmall-Volume Analysis

Daniel T. Chiu, Robert M. Lorenz, and Gavin D. M. Jeffries

University of Washington Seattle

By using methods that permit the generation and manipulation of ultrasmall-volume droplets, researchers are pushing the boundaries of ultrasensitive chemical analyses. (To listen to a podcast about this feature, please go to the *Analytical Chemistry* Web site at pubs.acs.org/ancham.)

What are ultrasmall volumes, and why would one want to perform chemical analyses in them? The next question might be, how does one go about doing chemistry and chemical analysis in ultrasmall volumes? The goal of this article is to clarify the first two questions and to offer various approaches to address the third. Our intent is to introduce concepts and challenges of ultrasmall volume analysis and describe the advances being made in the field. Because of the limited amount of space, we had to confine our referencing to highlight only a few select areas and recent advances.

The motivations to conduct chemistry in small, confined volumes are diverse. The chemistry of life takes place in a small, enclosed volume that we call a cell; it's a volume of picoliters, and it is further subdivided into even smaller volumes, from femtoliters to attoliters, that are defined by subcellular organelles. To understand how chemistry takes place in these tight spaces, it is often beneficial to re-create these very small volumes, such as with the use of lipid vesicles,[1–3] so we can perform controlled experiments to understand the effects of molecular confinement. Likewise, to conduct chemistry on a single cell or subcellular structure and analyze its molecular contents requires the ability to create similarly small volumes to overcome diffusion and to prevent dilution.[4] Another motivation to downsize volumes is the need to create large, dense arrays of parallelized small volumes for high-throughput screening, combinatorial chemistry and biology, or chemical synthesis.[5]

SMALL VOLUMES AND ULTRASMALL VOLUMES

Because of the diverse set of motivations and applications that require small, confined volumes, the range of volumes involved varies just as widely, from microliters[6] and nanoliters[7] to femtoliters and attoliters.[1,4,8–13] At first glance, a nanoliter might seem rather small, but it is actually 6 orders of magnitude larger than a femtoliter. Correspondingly, the technologies used to control a femtoliter volume can be quite different than those used to manipulate a nanoliter volume; the resultant chemistry also can be quite different, because of the large differences in the surface-to-volume ratios between these two volume scales. For example, the rate of change in concentration caused by evaporation scales as the fifth power of the surface-to-volume ratio,[14] and thus can be extremely rapid for a femtoliter volume but relatively slow when nanoliters are involved. Similarly, detection technologies used to probe femtoliters need to be ~1 million× more sensitive than those used for nanoliters: a solution containing 1 μM of dissolved species corresponds to ~600 million molecules in a nanoliter but only ~600 molecules in a femtoliter. As a result, powerful measurement techniques that have been successfully adapted for the analysis of nanoliters, such as NMR[7,15] and MS,[16] currently lack the needed sensitivity for the analysis of femtoliter volumes. A clear distinction exists, therefore, in the challenges and approaches used to manipulate and probe nanoliters and femtoliters.

To distinguish these two regimes, we refer to ultrasmall volumes as volumes that are 10 pL or less; this is comparable to or smaller than a typical mammalian cell. Most ultrasmall volumes we describe here will be in the femtoliter range. In contrast, small volumes are in the range of nanoliters to hundreds of picoliters; these volumes are used in most traditional microfabricated devices. Small volumes are too small to be pipetted in air with micropipettes (which go down to 1–2 μL) but large enough to be stably manipulated in air using surface tension, such as by electrowetting,[17] in through-hole well plates,[18] or in enclosed microchannels.[19]

Although evaporation is a concern, it is still possible to manipulate small volumes for a short period of time in air. However, ultrasmall volumes must always be immersed in, but separated from, another fluid by a physical barrier. To illustrate, a 1-nL droplet takes about 20 seconds (s) to evaporate in air, whereas a 1-fL droplet would evaporate in ~5 ms. Similarly, it is possible to define a small volume or a nanoliter plug in a microfluidic device even when the volume of interest and its surrounding solutions are in direct contact without any physical

10.1021/ac900306q CCC: $40.75 © 2009 American Chemical Society
Published on Web 06/09/2009

analytical chemistry
feature

Fusion Protein Modeller

Fluorescent-Protein-Based Biosensors: Modulation of Energy Transfer as a Design Principle

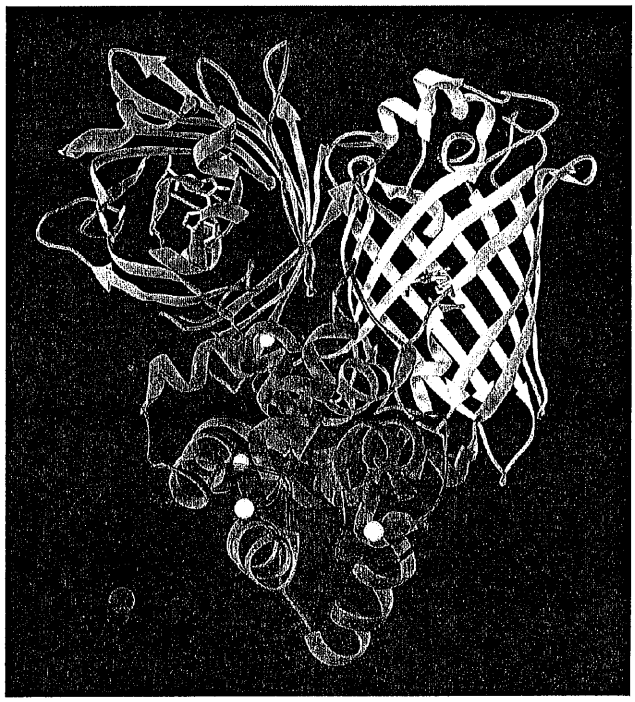
Robert E. Campbell

University of Alberta

Genetically-encoded biosensors based on FRET between fluorescent proteins of different hues enable quantitative measurement of intracellular enzyme activities and small molecule concentrations. (To listen to a podcast about this feature, please go to the *Analytical Chemistry* website at pubs.acs.org/journal/ancham.)

The definition of a "biosensor"—that is, a detection system that relies on a biomolecule for molecular recognition and a transducer to produce an observable output[1]—is firmly established in the lexicon of analytical chemistry. However, in recent years, biochemists appropriated the term biosensor to refer to genetically-encoded, designed proteins that are self-sufficient detection systems for a variety of targets, including small molecules and enzymes. The molecular recognition component of a biosensor is a protein, and this is also true for genetically-encoded biosensors. The primary difference between conventional biosensors and genetically-encoded biosensors is the nature of the transducer. Traditionally, the transducer is any one of a wide assortment of synthetic and modified surfaces that are electrochemically or optically sensitive to the action of the biomolecule. In contrast, the choice of transducer for a soluble and discrete genetically-encoded biosensor is constrained to being genetically encoded. Fortunately, one particular transducer format has proven to be exceptionally versatile for the construction of these biosensors: the modulation of Förster (or fluorescence) resonance energy transfer (FRET) between genetically fused fluorescent proteins (FP) of differing hues.[2]

The use of recombinant FPs for live-cell fluorescence imaging made its dramatic debut in 1994, when an image of a nematode worm with a green fluorescent neuron appeared on the cover of *Science*.[3] Ever since, the popularity and number of applications of FPs have increased at a dramatic pace. The single overwhelming advantage of FPs as fluorophores for live-cell applications is that, unlike synthetic dyes or quantum dots, they do not need to be manually introduced into the cell. A cell or organism containing an appropriate FP transgene can synthesize the FP polypeptide



TONY B. GINES/MICHAEL W. DAVIDSON

using intrinsic cellular transcriptional and translational machinery.[3] The nascent FP polypeptide then folds and autonomously undergoes posttranslational modifications to form a visible wavelength fluorophore buried deep within its β-barrel structure.[4] Through the efforts of protein engineers, a complete spectrum of FP variants—ranging from blue to far red[5]—are now available, all of which are close structural homologues of the archetypical *Aequorea* green FP (GFP).

The most common applications of FPs do not involve genetically-encoded biosensors but rather use FPs as reporters of gene expression or as markers for the localization of specific organelles or recombinant fusion proteins in live cells.[6,4] Design and construction of gene chimeras for these applications are relatively simple, and considerations are generally limited to the relative merits of N- versus C-terminal fusions and the length of flexible

10.1021/ac802613w CCC: $0.00

Single-Molecule Approaches to Stochastic Gene Expression

Arjun Raj and Alexander van Oudenaarden

Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; email: avano@mit.edu

Annu. Rev. Biophys. 2009. 38:255–70

First published online as a Review in Advance on February 5, 2009

The *Annual Review of Biophysics* is online at biophys.annualreviews.org

This article's doi: 10.1146/annurev.biophys.37.032807.125928

1936-122X/09/0609-0255$20.00

Key Words

single-molecule mRNA protein, gene expression noise

Abstract

Both the transcription of mRNAs from genes and their subsequent translation into proteins are inherently stochastic biochemical events, and this randomness can lead to substantial cell-to-cell variability in mRNA and protein numbers in otherwise identical cells. Recently, a number of studies have greatly enhanced our understanding of stochastic processes in gene expression by utilizing new methods capable of counting individual mRNAs and proteins in cells. In this review, we examine the insights that these studies have yielded in the field of stochastic gene expression. In particular, we discuss how these studies have played in understanding the properties of bursts in gene expression. We also compare the array of different methods that have arisen for single mRNA and protein detection, highlighting their relative strengths and weaknesses. In conclusion, we point out further areas where single-molecule techniques applied to gene expression may lead to new discoveries.

Quantitative analysis of gene expression in a single cell by qPCR

Kiyomi Taniguchi, Tomoharu Kajiyama & Hideki Kambara

We developed a quantitative PCR method featuring a reusable single-cell cDNA library immobilized on beads for measuring the expression of multiple genes in a single cell. We used this method to analyze multiple cDNA targets (from several copies to several hundred thousand copies) with an experimental error of 15.9% or less. This method is sufficiently accurate to investigate the heterogeneity of single cells.

Molecular biology in general and the human genome project in particular have provided a massive amount of information about genomes and proteins, and there is now much interest in understanding living systems at the molecular level. Because the smallest metabolically functional unit of a living organism is a single cell, many people are trying to analyze the molecular components in single cells. But information from typical ensemble measurements of mRNA and proteins is averaged from millions of cells. The individual cells from the same tissue may actually differ from each other and have different roles[1]. There have recently been several attempts to investigate the heterogeneity of gene expression in individual embryo cells, neurons, immunocytes and cancer cells[2–9]. These single-cell approaches have attracted much attention[10–12], and efforts are underway to develop tools for analyzing the components in single cells[13–15].

Our ultimate goal is to establish a method for quantifying all the mRNA in a single cell. This requires developing technologies to efficiently and reproducibly extract the mRNA needed to produce cDNA from a single cell and quantify the cDNA for each target gene. Previous work in this area has primarily used whole RNA or cDNA amplification coupled with DNA chips or quantitative PCR (qPCR). More recently, a single-cell mRNA analysis method using whole cDNA amplification coupled with digital counting of amplified cDNA with an ultra-high-throughput DNA sequencer was reported[16]. It was effective at finding many splicing variants, but in terms of quantitative analysis, it seems to have ambiguity owing to the amplification processes. As qPCR is considered to be the most accurate quantitative method at the present time, the best way to achieve accurate quantitative analysis is to use direct qPCR from a cDNA pool without pre-amplification. Although there have been several reports of qPCR being used to quantify expression of a few genes in a single cell, the analysis of multiple genes in a single cell is rather difficult because it requires dividing the sample, which reduces the sample size for measurement and therefore the detection sensitivity.

Although the use of oligo(dT)-immobilized beads for creating a cDNA library from multiple cells has been reported[17–19], in most cases the library has been used for reverse transcriptase PCR followed by agarose-gel electrophoresis analysis. The combination of qPCR and a bead-supported single-cell cDNA library is attractive for quantifying the expression of multiple genes by permitting repeated use of the library. However, it is difficult to reuse a cDNA library because of adsorption of PCR products on the beads and tube surfaces and desorption of immobilized cDNA from the beads during thermal cycling.

We overcame these difficulties by adding surfactant to the reaction mixture and by lowering the temperature during qPCR to reduce the thermal damage to the bead surfaces, which causes cDNA desorption. This method allowed accurate qPCR-mediated quantification of as few as several copies of mRNA from multiple genes in a single cell (**Supplementary Protocol**).

The important factors for producing single-cell cDNA libraries and using them for qPCR are the number of oligo$(dT)_{30}$-immobilized beads (capture beads), the 3′ bias in the cDNA production and the selection of transcriptase.

The number of oligo$(dT)_{30}$ probes on a capture bead was about 1.5×10^5. We estimated the number of mRNA molecules in a single cell to be 10^5–10^6. To optimize the number of capture beads, we estimated the mRNA capture rate together with the reverse transcription rate. As the number of mRNA molecules varied between cells, we used a model RNA (10^1–10^9 molecules, *EEF1G*) for the estimation. We created model cDNA libraries with various numbers of capture beads. For all of them, the number of cDNA molecules increased with the number of capture beads up to 10^7 capture beads (cDNA production efficiency was almost 100% for model RNA with less than 10^6 copies) (**Fig. 1a**). The reverse transcription efficiency with 10^8 capture beads was much lower for model RNA with over 10^6 copies. Therefore, the number of capture beads should be kept below 10^8. We used 10^7 capture beads. The capture beads included about 1.5×10^{12} oligo$(dT)_{30}$ probes.

The cDNA production efficiencies with oligo$(dT)_{30}$ probes for four model RNA (*TBP*, *SDHA*, *B2M* and *EEF1G*; 10^3 molecules each) were almost the same (**Fig. 1b**). Moreover, the production efficiencies with four types of probes (oligo$(dT)_{30}$, gene-specific, oligo$(dT)_{25}$VN (where V is an A, G and C mixture, and N is an A, G, C and T mixture) and locked nucleic acid (LNA)) for a model RNA (*SDHA*; 10^3 molecules) were the same (**Supplementary Fig. 1a** and **Supplementary Methods**).

It has been reported that a cDNA library produced with oligo(dT) probes has a 3′ bias, so cDNA lengths are not uniform

Hitachi Central Research Laboratory, Tokyo, Japan. Correspondence should be addressed to H.K. (hideki.kambara.se@hitachi.com).

RECEIVED 19 FEBRUARY; ACCEPTED 27 APRIL; PUBLISHED ONLINE 14 JUNE 2009; DOI:10.1038/NMETH.1338

Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell sorting based on enzymatic activity†

Jean-Christophe Baret,‡[a] Oliver J. Miller,‡[a] Valerie Taly,[a] Michaël Ryckelynck,[a] Abdeslam El-Harrak,[a] Lucas Frenz,[a] Christian Rick,[a] Michael L. Samuels,[b] J. Brian Hutchison,[b] Jeremy J. Agresti,[c] Darren R. Link,[b] David A. Weitz[c] and Andrew D. Griffiths*[a]

Received 5th February 2009, Accepted 14th April 2009
First published as an Advance Article on the web 23rd April 2009
DOI: 10.1039/b902504a

We describe a highly efficient microfluidic fluorescence-activated droplet sorter (FADS) combining many of the advantages of microtitre-plate screening and traditional fluorescence-activated cell sorting (FACS). Single cells are compartmentalized in emulsion droplets, which can be sorted using dielectrophoresis in a fluorescence-activated manner (as in FACS) at rates up to 2000 droplets s^{-1}. To validate the system, mixtures of *E. coli* cells, expressing either the reporter enzyme β-galactosidase or an inactive variant, were compartmentalized with a fluorogenic substrate and sorted at rates of ~300 droplets s^{-1}. The false positive error rate of the sorter at this throughput was <1 in 10^4 droplets. Analysis of the sorted cells revealed that the primary limit to enrichment was the co-encapsulation of *E. coli* cells, not sorting errors: a theoretical model based on the Poisson distribution accurately predicted the observed enrichment values using the starting cell density (cells per droplet) and the ratio of active to inactive cells. When the cells were encapsulated at low density (~1 cell for every 50 droplets), sorting was very efficient and all of the recovered cells were the active strain. In addition, single active droplets were sorted and cells were successfully recovered.

Introduction

The compartmentalization of assays in wells makes microtitre-plates the most flexible and most widely used screening platform in use today. However, reducing assay volumes to below 1–2 μl is problematic[1] and the maximum throughput, even when using sophisticated (and expensive) robotic handling, is little more than 1 s^{-1}. In contrast, fluorescence-activated cell sorting (FACS) is capable of analyzing and sorting cells at a rate of up to 7×10^4 cells s^{-1}.[2] However, during FACS, cell fluorescence is detected in a continuous aqueous stream[3] and the absence of compartmentalization limits the range of activities that can be screened: the fluorescent marker(s) must remain either inside or on the surface of the cells to be sorted. This makes detection of secreted enzymes using fluorogenic substrates impossible. Additionally, if the enzyme is intracellular, then the cell may be impermeable to the substrate or the product may freely diffuse out of the cell. Conventional FACS machines also require typically more than 10^5 cells in the starting population,[3] are very expensive and generate aerosols with serious biosafety ramifications.[4]

Microfluidic flow sorting systems have the potential to offer solutions to these problems, enabling the handling of small numbers of cells in inexpensive, sterile, aerosol-free, disposable devices.[2,5] Several approaches have already been demonstrated, including devices that sort cells by dielectrophoretic actuation, electrokinetic actuation, hydrodynamic flow-switching and optical forces (listed by Perroud *et al.*).[6] However, as with conventional FACS, the absence of assay compartmentalization limits their flexibility.

The versatility of conventional FACS can be increased by *in vitro* compartmentalization (IVC)[7] of assays in emulsion droplets, allowing selection for enzymatic activity.[8,9] However, the technique has three main limitations: complex double emulsions structures must be generated; the emulsions are highly polydisperse, limiting quantitative analysis; and the capacity to modify the contents of droplets after encapsulation is restricted.[10] These limitations can, however, be overcome by using droplet-based microfluidic systems, which allow the generation of highly monodisperse emulsions[11] and the fusion[12–14] and splitting[12,15,16] of droplets. It has even been possible to separate or sort droplets by charging them and steering them with an electric field[17] or by exploiting dielectrophoresis,[18] electrocoalescence,[19] localized heating,[20] or Rayleigh–Plateau instabilities.[21] It has not, however, been possible to selectively enrich specific subpopulations of droplets according to their fluorescence until now.

This manuscript describes a highly efficient droplet-based microfluidic FACS, optimized to sort picolitre-range droplets by dielectrophoresis.[17,18] This fluorescence-activated droplet sorting (FADS) system combines many of the advantages of microtitre-plate screening and fluorescence-activated cell sorting (FACS): assays are compartmentalized in emulsion droplets, which are the functional equivalent of microtitre-plate wells, but can be analyzed and sorted at high speed (as in FACS). Although several techniques for monitoring fluorescent reactions[22,23] and

[a]*Institut de Science et d'Ingénierie Supramoléculaires (ISIS), Université de Strasbourg, CNRS UMR 7006, 8 allée Gaspard Monge, BP 70028, F-67083 Strasbourg Cedex, France. E-mail: griffiths@isis.u-strasbg.fr*
[b]*RainDance Technologies, Inc., 44 Hartwell Avenue, Lexington, MA, 02421, USA*
[c]*Department of Physics and School of Engineering and Applied Sciences, Harvard University, Cambridge, USA*

† Electronic supplementary information (ESI) available: Model for cellular enrichment by FADS, supplementary Fig. S1–8, Table 1 and Movies S1–6. See DOI: 10.1039/b902504a

‡ These two authors contributed to the work equally.

Determination of Bacterial Antibiotic Resistance Based on Osmotic Shock Response

Scott M. Knudsen,† Marcio G. von Muhlen,† David B. Schauer,† and Scott R. Manalis*,†,‡

Departments of Biological Engineering and Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

We investigate the buoyant mass of bacterial cells in real time with the suspended microchannel resonator (SMR) as the population recovers from an osmotic shock. The density of the culture medium is chosen such that the bacteria initially have a positive buoyant mass which becomes negative as they recover from the hyperosmotic stress. This behavior can be used to differentiate between an antibiotic-resistant and an antibiotic-susceptible strain of the pathogenic bacteria *Citrobacter rodentium*, and we propose a general approach for exploiting the high precision of the SMR for rapid detection of antibiotic resistance.

The buoyant densities of bacterial cells are typically investigated using equilibrium sedimentation centrifugation experiments. Differences in cellular buoyant density have been used to separate different species of cells,[1,2] as well as to distinguish or separate different cell forms such as spores or minicells from normal cells.[3,4] It has also been shown that there are differences in buoyant density between the exponential and stationary culture phases.[5] However, centrifugation experiments can only look at populations of cells and cannot observe changes in density in real time. Microcantilever resonators[6,7] have been used to monitor growth of bacteria cells[8] and, in air, have detected bacteria binding events with single cell precision.[9] Although such resonators are highly sensitive in air or vacuum,[10,11] the presence of viscous damping in fluid has hindered the investigation of living bacteria at the single cell level. In this paper we observe changes in buoyant density of a population of *Escherichia coli* in real time using single-cell measurements of buoyant mass acquired on a suspended microchannel resonator (SMR), a device in which a microchannel is embedded within a resonating cantilever to reduce viscous damping.[12]

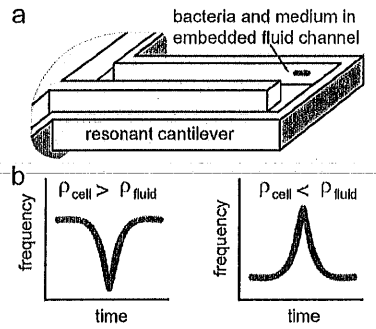


Figure 1. Principle of bacterial buoyant mass measurements on the SMR. (a) Bacteria suspended in a carrier medium pass through an embedded channel within a resonating cantilever, depicted here without an upper surface. (b) The position of a bacterium as it flows through the cantilever defines a transient frequency change, the height of which is proportional to the buoyant mass of the cell.

The resonant frequency of the SMR is determined by the physical properties of the cantilever itself and the mass of the solution within the channel. When a suspended particle of different density than the carrier fluid is passed through the channel, a transient frequency change occurs which is dependent upon the position of the particle within the cantilever (see Figure 1). The height of the resulting peak is proportional to the buoyant mass of the particle. In addition, the baseline frequency of the SMR provides a direct measure of the density of the carrier fluid throughout the experiment. We have previously reported the weighing of individual bacterial cells as they flow through the SMR.[13] Further, we have demonstrated that the density of a uniform population of particles can be investigated by measuring the buoyant mass of the particles in carrier fluids of different densities.[14] In this work, we investigate changes in the densities of single cells within a population by weighing each cell in a medium that is nearly isopycnic. Measured in this way, our density

* To whom correspondence should be addressed. E-mail: scottm@media.mit.edu.
† Department of Biological Engineering.
‡ Department of Mechanical Engineering.

(1) Carrera, M.; Zandomeni, R. O.; Sagripanti, J. L. *J. Appl. Microbiol.* **2008**, *105*, 68–77.
(2) Inoue, K.; Nishimura, M.; Nayak, B. B.; Kogure, K. *Appl. Environ. Microbiol.* **2007**, *73*, 1049–1053.
(3) Adler, H. I.; Fisher, W. D.; Cohen, A.; Hardigree, A. A. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *57*, 321–326.
(4) Jinks, D. C.; Guthrie, R.; Naylor, E. W. *J. Clin. Microbiol.* **1985**, *21*, 826–829.
(5) Makinoshima, H.; Nishimura, A.; Ishihama, A. *Mol. Microbiol.* **2002**, *43*, 269–279.
(6) Lang, H. P.; Berger, R.; Battiston, F.; Ramseyer, J. P.; Meyer, E.; Andreoli, C.; Brugger, J.; Vettiger, P.; Despont, M.; Mezzacasa, T.; Scandella, L.; Guntherodt, H. J.; Gerber, C.; Gimzewski, J. K. *Appl. Phys., A: Mater. Sci. Process.* **1998**, *66*, S61–S64.
(7) Thundat, T.; Chen, G. Y.; Warmack, R. J.; Allison, D. P.; Wachter, E. A. *Anal. Chem.* **1995**, *67*, 519–521.
(8) Gfeller, K. Y.; Nugaeva, N.; Hegner, M. *Biosens. Bioelectron.* **2005**, *21*, 528–533.
(9) Ilic, B.; Czaplewski, D.; Zalalutdinov, M.; Craighead, H. G.; Neuzil, P.; Campagnolo, C.; Batt, C. *J. Vac. Sci. Technol., B* **2001**, *19*, 2825–2828.
(10) Gupta, A.; Akin, D.; Bashir, R. *Appl. Phys. Lett.* **2004**, *84*, 1976–1978.
(11) Li, M.; Tang, H. X.; Roukes, M. L. *Nat. Nanotechnol.* **2007**, *2*, 114–120.
(12) Burg, T. P.; Sader, J. E.; Manalis, S. R. *Phys. Rev. Lett.* **2009**, *102*, 228103.
(13) Burg, T. P.; Godin, M.; Knudsen, S. M.; Shen, W.; Carlson, G.; Foster, J. S.; Babcock, K.; Manalis, S. R. *Nature* **2007**, *446*, 1066–1069.
(14) Godin,M.; Bryan, A. K.; Burg, T. P.; Babcock, K.; Manalis, S. R. *Appl. Phys. Lett.* **2007**, *91*, 123121.

10.1021/ac900968r CCC: $40.75

Vol 460|23 July 2009|doi:10.1038/nature08150

LETTERS

Partial penetrance facilitates developmental evolution in bacteria

Avigdor Eldar[1]*, Vasant K. Chary[2]*, Panagiotis Xenopoulos[2], Michelle E. Fontes[1], Oliver C. Losón[1], Jonathan Dworkin[3], Patrick J. Piggot[2] & Michael B. Elowitz[1]

Development normally occurs similarly in all individuals within an isogenic population, but mutations often affect the fates of individual organisms differently[1–4]. This phenomenon, known as partial penetrance, has been observed in diverse developmental systems. However, it remains unclear how the underlying genetic network specifies the set of possible alternative fates and how the relative frequencies of these fates evolve[5–8]. Here we identify a stochastic cell fate determination process that operates in *Bacillus subtilis* sporulation mutants and show how it allows genetic control of the penetrance of multiple fates. Mutations in an intercompartmental signalling process generate a set of discrete alternative fates not observed in wild-type cells, including rare formation of two viable 'twin' spores, rather than one within a single cell. By genetically modulating chromosome replication and septation, we can systematically tune the penetrance of each mutant fate. Furthermore, signalling and replication perturbations synergize to significantly increase the penetrance of twin sporulation. These results suggest a potential pathway for developmental evolution between monosporulation and twin sporulation through states of intermediate twin penetrance. Furthermore, time-lapse microscopy of twin sporulation in wild-type *Clostridium oceanicum* shows a strong resemblance to twin sporulation in these *B. subtilis* mutants[9,10]. Together the results suggest that noise can facilitate developmental evolution by enabling the initial expression of discrete morphological traits at low penetrance, and allowing their stabilization by gradual adjustment of genetic parameters.

Under nutrient-limited conditions, an individual *B. subtilis* cell can develop into a resilient dormant spore[11]. Many sporulation mutations reduce the fraction of cells that sporulate successfully (Fig. 1a, b)[4,12–14]. This makes sporulation an ideal model system in which to study the origins and impact of partial penetrance.

At the onset of sporulation, *B. subtilis* cells divide asymmetrically into smaller (forespore) and larger (mother-cell) compartments. Septation leads to the forespore-specific activation of the transcriptional regulator σ^F (Fig. 1c, d)[11]. This in turn activates expression of *spoIIR*, which initiates an intercompartmental signalling cascade that activates the mother-cell-specific regulator σ^E, causing mother-cell differentiation[11,15]. Deleting σ^E allows a second asymmetric septum to form, resulting in 'abortively disporic' cells with two DNA-containing immature forespores and a mother cell devoid of DNA[16,17]. Attenuation of *spoIIR* expression results in a partially penetrant mixture of successfully sporulating and abortively disporic cells[12,13].

To explore the effects of *spoIIR* mutations on sporulation penetrance, we constructed a set of strains, collectively denoted as *spoIIR*^{PP} mutants, in which the rate and/or the time of onset of *spoIIR* expression is specifically perturbed (Supplementary Methods

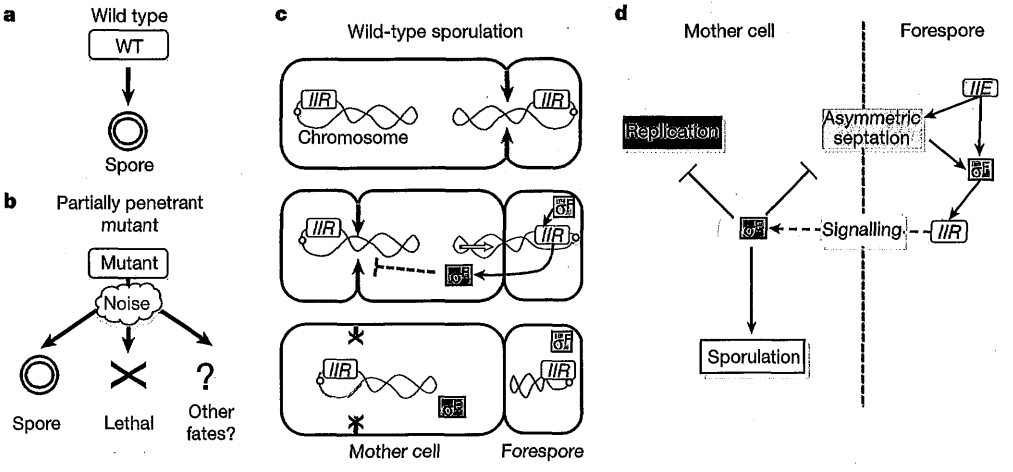


Figure 1 | Partial penetrance in the developmental process of sporulation. **a**, In wild-type sporulation, each sporulating cell produces a single spore. **b**, Partly penetrant mutants exhibit a mixture of normal sporulation, lethal failures (cross) and alternative viable fates ('?') due to cellular fluctuations (cloud). **c, d**, Events (**c**) and genetic interactions (**d**) leading to differentiation of the mother-cell and forespore compartments (see text).

[1]Howard Hughes Medical Institute and Division of Biology and Department of Applied Physics, California Institute of Technology, Pasadena, California 91125, USA. [2]Department of Microbiology and Immunology, Temple University School of Medicine, 3400 North Broad Street, Philadelphia, Pennsylvania 19140, USA. [3]Department of Microbiology, College of Physicians and Surgeons, Columbia University, New York, New York 10032, USA.
*These authors contributed equally to this work.

Vol 460|9 July 2009|doi:10.1038/nature08112

ARTICLES

Adaptive prediction of environmental changes by microorganisms

Amir Mitchell[1], Gal H. Romano[2], Bella Groisman[1], Avihu Yona[1], Erez Dekel[3], Martin Kupiec[2], Orna Dahan[1]* & Yitzhak Pilpel[1,4]*

Natural habitats of some microorganisms may fluctuate erratically, whereas others, which are more predictable, offer the opportunity to prepare in advance for the next environmental change. In analogy to classical Pavlovian conditioning, microorganisms may have evolved to anticipate environmental stimuli by adapting to their temporal order of appearance. Here we present evidence for environmental change anticipation in two model microorganisms, *Escherichia coli* and *Saccharomyces cerevisiae*. We show that anticipation is an adaptive trait, because pre-exposure to the stimulus that typically appears early in the ecology improves the organism's fitness when encountered with a second stimulus. Additionally, we observe loss of the conditioned response in *E. coli* strains that were repeatedly exposed in a laboratory evolution experiment only to the first stimulus. Focusing on the molecular level reveals that the natural temporal order of stimuli is embedded in the wiring of the regulatory network—early stimuli pre-induce genes that would be needed for later ones, yet later stimuli only induce genes needed to cope with them. Our work indicates that environmental anticipation is an adaptive trait that was repeatedly selected for during evolution and thus may be ubiquitous in biology.

Microorganisms are constantly faced with environmental stimuli and stresses. The cellular response to such challenges has been intensively studied in several model organisms[1–4]. The simplest response strategy to a stimulus is to monitor the environment and to respond directly to it using designated mechanisms (Fig. 1). The environmental stress response in yeast represents a more complicated strategy in which the responses to many stresses are partially overlapping[1,2]. Theoretical work has shown that when a population of microorganisms evolves under erratic environmental fluctuations, cells may not effectively monitor the environment, but rather use stochasticity to randomly alternate between potential states[5] (Fig. 1). Stochastic switching might thus ensure that a portion of the population is prepared in advance for the unpredicted challenge[6,7]. However, other, more predictable, environments offer organisms the opportunity to adopt an alternative regulation strategy of anticipating an environmental change based on a preceding signal. The capacity of some complex multicellular eukaryotes to capture the statistics that govern the temporal connection between events in their environment, known as classical Pavlovian conditioning, serves as a central paradigm in the study of learning[8]. Here we ask whether genetic regulatory networks of microorganisms adaptively evolved to capture the temporal connections between subsequent stimuli in their environment. Most recently, 'anticipatory regulation' was discovered[9] (Fig. 1)—an association between environmental changes in bacteria. Specifically, this study investigated the response of *E. coli* to temperature increase that is followed by a drop in oxygen availability upon its entry to the digestive tract. Interestingly, these two signals show a symmetrical associative regulation pattern—each signal affects the expression of genes needed to cope with both (Fig. 1). Remarkably, the authors successfully decoupled the two responses during a laboratory evolution experiment in which the two signals were presented out-of-phase from one another. The ability to decouple the two responses is an indication that the coupling seen in the wild type is not a trivial combined response to the two stresses.

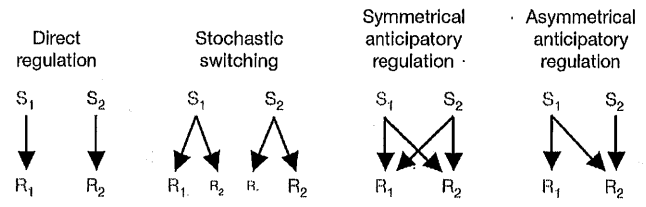


Figure 1 | Four possible regulation strategies in response to environmental stimuli. Under direct regulation, each of the stimuli, S_1 and S_2, activates exclusively the responses R_1 and R_2, respectively. Under stochastic switching, cells randomly sample either R_1 or R_2 in response to either S_1 or S_2. Under symmetrical anticipatory regulation, each of the stimuli activates both responses. Under asymmetrical anticipatory regulation, the stimulus that usually appears first in the ecology activates both responses, whereas the stimulus that appears later induces only the second response.

Here we show that biological systems that react to a unidirectional temporal order of environmental changes may manifest a more elaborate predictive capacity. This capacity is reflected in a corresponding asymmetric response strategy between subsequent stimuli, denoted S_1 and S_2 hereafter, and their designated responses R_1 and R_2, respectively. The first stimulus, S_1, activates both responses, R_1 and R_2, yet because the second stimulus, S_2, does not predict the appearance of S_1, it only activates its own response (Fig. 1). We propose three criteria to determine whether the observed cross-regulation pattern forms an adaptive anticipatory response strategy that could be selected for by evolution. First, asymmetric fitness advantage: pre-exposure to S_1 increases the fitness under S_2, yet pre-exposure to S_2 should not enhance fitness upon subsequent growth on S_1. This ensures that the natural order of stimuli was captured during evolution. Second,

[1]Department of Molecular Genetics, Weizmann Institute of Science Rehovot 76100, Israel. [2]Department of Molecular Microbiology and Biotechnology, Tel Aviv University, Tel Aviv 69978, Israel. [3]Department of Molecular Cell Biology, Weizmann Institute of Science Rehovot 76100, Israel. [4]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA.
*These authors contributed equally to this work.

LETTERS

A spatial gradient coordinates cell size and mitotic entry in fission yeast

James B. Moseley[1], Adeline Mayeux[2], Anne Paoletti[2] & Paul Nurse[1]

Many eukaryotic cell types undergo size-dependent cell cycle transitions controlled by the ubiquitous cyclin-dependent kinase Cdk1 (refs 1–4). The proteins that control Cdk1 activity are well described but their links with mechanisms monitoring cell size remain elusive. In the fission yeast *Schizosaccharomyces pombe*, cells enter mitosis and divide at a defined and reproducible size owing to the regulated activity of Cdk1 (refs 2, 3). Here we show that the cell polarity protein kinase Pom1, which localizes to cell ends[5], regulates a signalling network that contributes to the control of mitotic entry. This network is located at cortical nodes in the middle of interphase cells, and these nodes contain the Cdk1 inhibitor Wee1, the Wee1-inhibitory kinases Cdr1 (also known as Nim1) and Cdr2, and the anillin-like protein Mid1. Cdr2 establishes the hierarchical localization of other proteins in the nodes, and receives negative regulatory signals from Pom1. Pom1 forms a polar gradient extending from the cell ends towards the cell middle and acts as a dose-dependent inhibitor of mitotic entry, working through the Cdr2 pathway. As cells elongate, Pom1 levels decrease at the cell middle, leading to mitotic entry. We propose that the Pom1 polar gradient and the medial cortical nodes generate information about cell size and coordinate this with mitotic entry by regulating Cdk1 through Pom1, Cdr2, Cdr1 and Wee1.

Recurring cycles of cellular growth and division present a simple question: how do cells link these processes and divide only upon reaching an appropriate size? Genetic screens have identified components in the fission yeast controlling cell size at division[6,7]. The central factor Cdk1 drives entry into mitosis by phosphorylating a wide range of targets[8]. In small cells that are in early G2, Cdk1 is kept inactive by the kinase Wee1, which directly phosphorylates and inhibits Cdk1. As cell size increases during G2, the phosphatase Cdc25 removes this inhibitory phosphorylation to activate Cdk1 and allow mitotic entry. Thus, the mitotic entry control system coordinates a balance of Wee1 and Cdc25 activity with changes in cell size to generate a reproducible cell size at division. Although several factors regulating Wee1 and Cdc25 have been identified[3], the links between cell size and the Wee1-Cdc25-Cdk1 module have remained elusive.

We noted that the protein kinase Cdr2, which negatively regulates Wee1, is localized to a band of cortical nodes in the middle of interphase cells[9]. After mitotic entry, cytokinesis factors such as myosin II are recruited to similar nodes, which subsequently condense to form the cytokinetic ring[10]. The presence of Cdr2 in these medial nodes during G2, when the inhibition of Wee1 contributes to the G2–M transition, suggested that the nodes might act as potential sites of cell cycle regulation. To investigate this possibility further, we first analysed the composition of medial nodes. We found that the previously uncharacterized protein Blt1 (also known as SPBC1A4.05), identified using a proteomic approach (Supplementary Fig. 1), co-localized with Cdr2 in the medial interphase nodes, together with Mid1 (Fig. 1a, b), which was previously shown to localize to similar interphase structures[11]. Furthermore, physical interactions between Blt1–Mid1, Blt1–Cdr2 and Cdr2–Mid1 were detected by co-immunoprecipitation (Fig. 1d). The Cdr2-related kinase Cdr1, which directly phosphorylates and inhibits Wee1 *in vitro*[12–14], also co-localized with Blt1 in the nodes (Fig. 1c and Supplementary Fig. 2a). The presence of two components of the mitotic control system, Cdr1 and Cdr2, in nodes during G2 supports the possibility that these structures might function to regulate mitotic entry. To investigate the recruitment of these proteins to the interphase nodes, we monitored green fluorescent protein (GFP)-tagged node proteins in cells deleted for the other components, and found that Cdr2 localized to interphase nodes in *mid1*Δ, *blt1*Δ and *cdr1*Δ cells, but was required for all other proteins to reach the interphase cortical nodes (Fig. 2a and Supplementary Fig. 3a). We determined a hierarchy for recruitment of these proteins to interphase nodes by examining all combinations of GFP-tagged proteins and mutants, including the additional interphase node proteins Klp8 and Gef2 (Supplementary Figs 2, 3 and data not shown). These results indicate that medial cortical nodes are formed by the ordered, Cdr2-dependent assembly of multiple interacting proteins during interphase.

The presence of the mitotic regulators Cdr1 and Cdr2 in the interphase nodes prompted us to examine whether Blt1 also regulates mitotic entry. *blt1*Δ cells had an increased length at division, consistent with a delay in mitotic entry. This phenotype was synthetic with a temperature-sensitive *cdc25^{ts}* mutation (Supplementary Table 1), but not with *cdr1*Δ and *cdr2*Δ mutants, which act through Wee1 (refs 15–17). These genetic data are consistent both with the localization hierarchy described earlier and with a pathway in which Cdr2 functions upstream of Cdr1 and Blt1 to promote mitosis through the negative regulation of Wee1 (Fig. 2b).

To determine whether these nodes might act as sites of Wee1 regulation, we next examined the cellular distribution of Wee1. GFP–Wee1 under the control of the low-level P81nmt1 promoter was found in three cellular structures: the medial cortical nodes, the nucleus and the spindle pole body (Fig. 2c). GFP–Wee1 co-localized with Blt1–mCherry at the medial cortical nodes (Fig. 2d), and with the spindle pole body marker Sid4–mCherry (Supplementary Fig. 4). To confirm this localization pattern for endogenously expressed Wee1, we integrated a triple-GFP tag at the genomic *wee1$^+$* locus. Although present at low levels, Wee1–3×GFP was found in the medial cortical nodes and the nucleus (Fig. 2e). Wee1 localization to the medial cortical nodes depended on Cdr2 but not on Cdr1 or Blt1 (Supplementary Fig. 5), reinforcing the importance of Cdr2 in generating these structures. Moreover, a Cdr2(E177A) mutant, which localizes properly but is predicted to abolish Cdr2 kinase activity[9], disrupted the recruitment of Wee1 to the medial cortex and delayed entry into mitosis (Supplementary Fig. 6), indicating a role for Cdr2 kinase activity in signalling to Wee1. Because Wee1 is present in the same medial cortical nodes as its inhibitory network, we propose that

[1]The Rockefeller University, New York, New York 10065, USA. [2]Institut Curie, Centre de Recherche and CNRS UMR144, Paris 75248 cedex 05, France.

Evolutionary selection between alternative modes of gene regulation

Ulrich Gerland[a,1] and Terence Hwa[b,1]

[a]Institute for Theoretical Physics, Arnold Sommerfeld Center for Theoretical Physics, Theresienstrasse 37, 80333 Munich, Germany; and [b]Center for Theoretical Biological Physics and Department of Physics, University of California at San Diego, La Jolla, CA 92093-0374

Edited by Curtis G. Callan, Jr., Princeton University, Princeton, NJ, and approved April 2, 2009 (received for review August 27, 2008)

Microorganisms employ a wealth of gene regulatory mechanisms to adjust their growth programs to variations in the environment. It was pointed out long ago [Savageau M (1977) *Proc Natl Acad Sci USA* 74: 5647–5651] that the particular mode of gene regulation employed may be correlated with the "demand" on the regulated gene, i.e., how frequently the gene product is needed in its natural habitat. An evolutionary "use-it-or-lose-it" principle was proposed to govern the choice of gene regulatory strategies. Here, we examine quantitatively the forces selecting for and against two opposing modes of gene regulation, in the context of an evolutionary model that takes genetic drift, mutation, and time-dependent selection into account. We consider the effect of time-dependent selection, with periods of strong selection alternating with periods of neutral evolution. Using a variety of analytical methods, we find the effective population size and the typical time scale of environmental variations to be key parameters determining the fitness advantage of the different modes of regulation. Our results support Savageau's use-it-or-lose-it principle for small populations with long time scales of environmental variations and support a complementary "wear-and-tear" principle for the opposite situation.

transcription control | design principles | molecular evolution | time-dependent selection

Much effort is currently devoted to the study of design principles for functional modules in molecular and cell biology (1). Typically, design principles are based on optimizing the functional performance of a module. However, different designs can be functionally equivalent. Already 30 years ago, Savageau raised the fundamental question of whether nature's choice between functionally equivalent module designs is random or whether there is an evolutionary selection criterion (2, 3). This question was posed in the context of a simple genetic switch, the elementary unit of gene regulatory modules. A basic function of a genetic switch is to assure that the expression of a specific gene is turned on when a signal, e.g., a nutrient, is present and turned off when the signal is absent. Importantly, the presence of the signal may vary on long time scales, e.g., a population of *Escherichia coli* cells can use lactose as the carbon source in a mammalian infant, but lactose may then become unavailable for a long time in the same host (2). The desired regulatory function can be implemented by a double-positive (++) mode of control, e.g., the signal activates a transcription factor that then activates the gene (Fig. 1*B*). The same function can also be obtained with a double-negative (−−) mode of control, e.g., the signal disables specific binding of a transcriptional repressor to its operator site, thereby relieving repression (Fig. 1*C*). Indeed, both of these control modes are ubiquitously used in bacterial gene regulation (4).

Savageau empirically examined many bacterial genetic switches for correlations between the mode of gene regulation and temporal patterns in the input signal (2, 3). The latter were estimated, for instance, from physiological measurements of nutrient absorption rates in the mammalian intestine, which are an indication for the likelihood that a specific nutrient reaches the colon where *E. coli* colonizes. The study suggested a strong correlation between the "demand" for the product of the regulated gene and the mode of control: Genes whose protein products were needed most of the time ("high demand") were found to be under (++) control, whereas genes whose products were rarely needed ("low demand") were under (−−) control. To rationalize this correlation, Savageau proposed an intriguing "use-it-or-lose-it" principle, wherein the mode of gene regulation should be chosen to maximize the usage of the regulator, so as to avoid the loss of functionality during the periods when they are not used. Indeed, an activating transcription factor is only needed to be functional (e.g., bind to its functional DNA-binding site) when the target gene needs to be expressed, whereas a repressor is only needed to be functional when default expression of the target gene needs to be turned OFF. Hence, the use-it-or-lose-it principle is consistent with regulation by an activator for genes under high demand and regulation by a repressor for genes under low demand.

The proposed qualitative principle calls for a quantitative theoretical formulation and analysis, as recognized already in the original work of Savageau (2). Indeed, a more recent theoretical study by Savageau (5) yielded some support for an evolutionary choice of repressors at low demand and activators at high demand. However, that study did not explicitly consider stochastic fluctuations in the form of genetic drift, which had been suggested to play an important role for the use-it-or-lose-it principle (2). Moreover, a recent article (6) challenges the evolutionary basis of the empirical correlations and discusses some ideas for alternative, functional explanations. Thus, an explicit theoretical formulation of the use-it-or-lose-it principle is clearly needed, together with an assessment of the conditions under which the principle may be applicable.

Here, we use the framework of theoretical population genetics to provide a quantitative formulation of the problem. On the one hand, this framework allows us to assess the conditions under which the use-it-or-lose-it principle is borne out and show that significant genetic drift is indeed an essential requirement (with a detailed discussion of our findings in comparison with those of the previous theoretical study (5) below). On the other hand, our framework reveals another, more general aspect to the problem: The use-it-or-lose-it principle is contrary to the well-established population genetics concept of genetic robustness (7), which focuses on the "mutational load," i.e., the average fitness reduction of individuals in a population incurred by mutations. One expects this load to be minimal when a transcriptional regulator is rarely used, because the fitness of a strain with a dysfunctional regulator is reduced only during the periods when the regulator is needed. We will loosely refer to the evolutionary design principle based on this argument as the "wear-and-tear" principle. We will show that, somewhat surprisingly, our quantitative formulation supports either of the two opposing principles, depending on

Author contributions: U.G. and T.H. designed research, performed research, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. E-mail: gerland@lmu.de or hwa@ucsd.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0808500106/DCSupplemental.

Adiabatic coarse-graining and simulations of stochastic biochemical networks

N. A. Sinitsyn[a,b], Nicolas Hengartner[b], and Ilya Nemenman[a,b,1]

[a]Center for Nonlinear Studies, and [b]Computer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545

Edited by William H. Press, University of Texas, Austin, TX, and approved April 10, 2009 (received for review September 18, 2008)

We propose a universal approach for analysis and fast simulations of stiff stochastic biochemical networks, which rests on elimination of fast chemical species without a loss of information about mesoscopic, non-Poissonian fluctuations of the slow ones. Our approach is similar to the Born–Oppenheimer approximation in quantum mechanics and follows from the stochastic path integral representation of the cumulant generating function of reaction events. In applications with a small number of chemical reactions, it produces analytical expressions for cumulants of chemical fluxes between the slow variables. This allows for a low-dimensional, interpretable representation and can be used for high-accuracy, low-complexity coarse-grained numerical simulations. As an example, we derive the coarse-grained description for a chain of biochemical reactions and show that the coarse-grained and the microscopic simulations agree, but the former is 3 orders of magnitude faster.

Computer simulations are often the method of choice to explore an agreement between a model and experimental data in systems biology. Unfortunately, even the simplest biochemical simulations often face serious conceptual and practical problems. First, they usually involve combinatorially many chemical species and reaction processes: for example, a single molecule with n modification sites can exist in 2^n microscopic states (1). Second, although it is widely known that some molecules occur in cells at very low copy numbers (e.g., the DNA), which give rise to important stochastic effects, it is less appreciated that the combinatorial complexity makes this true for many molecular species. Indeed, even for a large total number of molecules, typical abundances of a species may be small if the number of the species is combinatorial. Third, and perhaps the most profound difficulty, is that only very few of the kinetic parameters underlying the networks are experimentally observable.

Although some day, computers may be able to tackle the formidable problem of modeling combinatorially complex biochemical processes and then performing sweeps through parameter spaces in search of an agreement with experiments, this day is far away. More importantly, even if the computing power were available, it would not help in building a comprehensible interpretation of the modeled system and in identifying connections between its microscopic and macroscopic features.

Clearly, such an interpretation can be aided by coarse-graining, that is, by merging or eliminating certain nodes and/or reaction processes (this would be called blocking or decimation in statistical physics). Ideally, one wants to substitute multiple elementary (that is, single-step, Poisson-distributed) biochemical reactions with a few complex processes in a way that retains predictability of the system. Not incidentally, this would help with each of the 3 roadblocks mentioned above by reducing the number of interacting elements, increasing the copy numbers of agglomerated hyperspecies, and combining multiple microscopic rates into a smaller number of effective parameters.

Coarse-graining in biochemistry is well established, and the prime example is the Michaelis–Menten (MM) kinetics (2)

$$E + S \underset{k_{-1}[C]}{\overset{k_1[E][S]}{\rightleftharpoons}} C \xrightarrow{k_2[C]} E + P. \qquad [1]$$

Here k_1, k_2, and k_{-1} are kinetic rates, S, P, E, and C denote the substrate, the product, the enzyme, and the enzyme–substrate complex molecules, respectively, and $[\ldots]$ represent the abundances. The enzyme catalyzes the $S \to P$ transformation by merging with S to create an unstable complex C, which then dissociates either back into $E+S$ or forward into $E+P$, leaving E unmodified. If $[S] \gg [E]$, then the enzyme cycles many times before $[S]$ changes appreciably. Thus the enzyme equilibrates resulting in a coarse-grained reaction with the decimated enzyme species:

$$S \xrightarrow{\nu} P, \quad \nu = \frac{k_2[S][E]}{[S] + (k_2 + k_{-1})/k_1}. \qquad [2]$$

However, this simple reduction is insufficient when stochasticity is important: Each MM reaction consists of multiple elementary steps, thus approximating the number of the reactions as a Poisson variable (3) is not always valid. While some attempts have been made to extend deterministic coarse-graining to the stochastic domain (4–7), such systematic tools have not been found yet. In this article, we make a step towards the goal.

We start by noting that, in addition to the 3 conceptual problems, a technical one stands in the way of stochastic simulations in systems biology: Molecular species have diverse dynamical time scales, making the systems stiff and difficult to simulate. We propose to use this property to our advantage, finding a coarse-graining procedure exhibiting the following 4 features.

First, like in the deterministic MM case, fast variables must not only be treated differently from the slow ones, but they must be eliminated altogether. Otherwise, the coarse-graining would not decrease the complexity of the interpretation and of numerical simulations, which scale at least linearly with the number of the involved variables.

Second, the distinction between the fast and the slow variables must not be based on reaction rates. For example, for the MM scheme, all 3 reaction rates may be comparable, and coarse-graining is still possible due to the difference in the enzyme and the substrate abundances. Overall, if 2 species of different abundances are coupled by a reaction, then a relatively small change in the high-abundance one can have a dramatic effect on the low-abundance one, leading to the different dynamical time scales. We seek the notion of species-rather than reaction-based adiabaticity as a basis for the coarse-graining. This has an additional advantage: Having higher abundances, coarse-grained variables will be amenable to fast mesoscopic, Langevin-like (8) methods, instead of event-by-event simulations (9).

Third, real biological systems have more than just fast and slow variables; instead a whole spectrum of time scales is usually

Author contributions: N.A.S. and I.N. designed research; N.A.S., N.H., and I.N. performed research; and N.A.S., N.H., and I.N. 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: ilya@menem.com.

This article contains supporting information online at www.pnas.org/cgi/content/full/0809340106/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.0809340106

Efficient computation of optimal actions

Emanuel Todorov[1]

Departments of Applied Mathematics and Computer Science & Engineering, University of Washington, Box 352420, Seattle, WA 98195

Edited by James L. McClelland, Stanford University, Stanford, CA, and approved April 28, 2009 (received for review November 16, 2007)

Optimal choice of actions is a fundamental problem relevant to fields as diverse as neuroscience, psychology, economics, computer science, and control engineering. Despite this broad relevance the abstract setting is similar: we have an agent choosing actions over time, an uncertain dynamical system whose state is affected by those actions, and a performance criterion that the agent seeks to optimize. Solving problems of this kind remains hard, in part, because of overly generic formulations. Here, we propose a more structured formulation that greatly simplifies the construction of optimal control laws in both discrete and continuous domains. An exhaustive search over actions is avoided and the problem becomes linear. This yields algorithms that outperform Dynamic Programming and Reinforcement Learning, and thereby solve traditional problems more efficiently. Our framework also enables computations that were not possible before: composing optimal control laws by mixing primitives, applying deterministic methods to stochastic systems, quantifying the benefits of error tolerance, and inferring goals from behavioral data via convex optimization. Development of a general class of easily solvable problems tends to accelerate progress—as linear systems theory has done, for example. Our framework may have similar impact in fields where optimal choice of actions is relevant.

action selection | cost function | linear Bellman equation | stochastic optimal control

If you are going to act, you might as well act in the best way possible. But which way is best? This is the general problem we consider here. Examples include a nervous system generating muscle activations to maximize movement performance (1), a foraging animal deciding which way to turn to maximize food (2), an internet router directing packets to minimize delays (3), an onboard computer controlling a jet engine to minimize fuel consumption (4), and an investor choosing transactions to maximize wealth (5). Such problems are often formalized as Markov decision processes (MDPs), with stochastic dynamics $p(x'|x,u)$ specifying the transition probability from state x to state x' under action u, and immediate cost $\ell(x,u)$ for being in state x and choosing action u. The performance criterion that the agent seeks to optimize is some cumulative cost that can be formulated in multiple ways. Throughout the article we focus on one formulation (total cost with terminal/goal states) and summarize results for other formulations.

Optimal actions cannot be found by greedy optimization of the immediate cost, but instead must take into account all future costs. This is a daunting task because the number of possible futures grows exponentially with time. What makes the task doable is the optimal cost-to-go function $v(x)$ defined as the expected cumulative cost for starting at state x and acting optimally thereafter. It compresses all relevant information about the future and thus enables greedy computation of optimal actions. $v(x)$ equals the minimum (over actions u) of the immediate cost $\ell(x,u)$ plus the expected cost-to-go $\mathrm{E}[v(x')]$ at the next state x':

$$v(x) = \min_u \{\ell(x,u) + \mathrm{E}_{x'\sim p(\cdot|x,u)}[v(x')]\}. \qquad [1]$$

The subscript indicates that the expectation is taken with respect to the transition probability distribution $p(\cdot|x,u)$ induced by action u. Eq. **1** is fundamental to optimal control theory and is called the Bellman equation. It gives rise to Dynamic Programming (3) and Reinforcement Learning (2) methods that are very general but can be inefficient. Indeed, Eq. **1** characterizes $v(x)$ only implicitly, as the solution to an unsolved optimization problem, impeding both analytical and numerical approaches.

Here, we show how the Bellman equation can be greatly simplified. We find an analytical solution for the optimal u given v, and then transform Eq. **1** into a linear equation. Short of solving the entire problem analytically, reducing optimal control to a linear equation is the best one can hope for. This simplification comes at a modest price: although we impose certain structure on the problem formulation, most control problems of practical interest can still be handled. In discrete domains our work has no precursors. In continuous domains there exists related prior work (6–8) that we build on here. Additional results can be found in our recent conference articles (9–11), online preprints (12–14), and supplementary notes [supporting information (SI) *Appendix*].

Results

Reducing Optimal Control to a Linear Problem. We aim to construct a general class of MDPs where the exhaustive search over actions is replaced with an analytical solution. Discrete optimization problems rarely have analytical solutions, thus our agenda calls for continuous actions. This may seem counterintuitive if one thinks of actions as symbols ("go left," "go right"). However, what gives meaning to such symbols are the underlying transition probabilities—which are continuous. The latter observation is key to the framework developed here. Instead of asking the agent to specify symbolic actions, which are then replaced with transition probabilities, we allow the agent to specify transition probabilities $u(x'|x)$ directly. Formally, we have $p(x'|x,u) = u(x'|x)$.

Thus, our agent has the power to reshape the dynamics in any way it wishes. However, it pays a price for too much reshaping, as follows. Let $p(x'|x)$ denote the passive dynamics characterizing the behavior of the system in the absence of controls. The latter will usually be defined as a random walk in discrete domains and as a diffusion process in continuous domains. Note that the notion of passive dynamics is common in continuous domains but is rarely used in discrete domains. We can now quantify how "large" an action is by measuring the difference between $u(\cdot|x)$ and $p(\cdot|x)$. Differences between probability distributions are usually measured via Kullback–Leibler (KL) divergence, suggesting an immediate cost of the form

$$\ell(x,u) = q(x) + \mathrm{KL}(u(\cdot|x)||p(\cdot|x)) = q(x) + \mathrm{E}_{x'\sim u(\cdot|x)}\left[\log \frac{u(x'|x)}{p(x'|x)}\right]. \qquad [2]$$

The state cost $q(x)$ can be an arbitrary function encoding how (un)desirable different states are. The passive dynamics $p(x'|x)$ and controlled dynamics $u(x'|x)$ can also be arbitrary, except that

Author contributions: E.T. designed research, performed research, analyzed data, and wrote the paper.

The author declares no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

See Commentary on Page 11429.

[1]E-mail: todorov@cs.washington.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0710743106/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.0710743106

Development of GFP-based biosensors possessing the binding properties of antibodies

Tej V. Pavoor, Yong Ku Cho, and Eric V. Shusta[1]

Department of Chemical and Biological Engineering, University of Wisconsin, Madison, WI 53706

Edited by Frances H. Arnold, California Institute of Technology, Pasadena, CA, and approved May 29, 2009 (received for review March 16, 2009)

Proteins that can bind specifically to targets that also have an intrinsic property allowing for easy detection could facilitate a multitude of applications. While the widely used green fluorescent protein (GFP) allows for easy detection, attempts to insert multiple binding loops into GFP to impart affinity for a specific target have been met with limited success because of the structural sensitivity of the GFP chromophore. In this study, directed evolution using a surrogate loop approach and yeast surface display yielded a family of GFP scaffolds capable of accommodating 2 proximal, randomized binding loops. The library of potential GFP-based binders or "GFAbs" was subsequently mined for GFAbs capable of binding to protein targets. Identified GFAbs bound with nanomolar affinity and required binding contributions from both loops indicating the advantage of a dual loop GFAb platform. Finally, GFAbs were solubly produced and used as fluorescence detection reagents to demonstrate their utility.

alternative scaffold | directed evolution | yeast surface display | thermal stability | loop randomization

Antibodies have long been a mainstay of biological and medical research, and the current use of antibodies as therapeutics has further expanded their portfolio of applications. More recently, to address various challenges such as the reduced stability and production yields of the antibody fragments that are frequently used in in vitro evolution platforms, and in large part as a result of intellectual property concerns, the field of alternative binding scaffolds has emerged (1). By mutagenizing solvent-exposed loop regions or inserting diverse loop repertoires into nonantibody protein scaffolds, specific binding attributes can be conferred to proteins that naturally have desirable properties, such as high stability and production titers. In this way, alternative scaffolds such as the 10th human fibronectin type III domain (2), anticalins (3–5), designed ankyrin repeat proteins (6), and Affibodies (7, 8), among others, have been developed to bind to targets with antibody-like affinity.

Green fluorescent protein (GFP) has also been explored as a potential alternative scaffold. To date, GFP has been used for a wide variety of different applications (9) including Ca^{2+} detection (10), visualization of protein–protein interactions (11), and as a reporter for protein folding (12). Considerable effort has also been expended in attempts to develop GFP as a binding scaffold that would have 2 potential advantages over the aforementioned alternative scaffolds. First, by combining binding attributes with the intrinsic fluorescence of the GFP protein, the proteins could act as single step detection reagents in applications such as fluorescence-based ELISAs, flow cytometry, and intracellular targeting/trafficking in live cells. Second, GFP fluorescence requires that the protein is properly folded (13) offering an in situ metric for folding fidelity, absent from other alternative scaffolds. Such a folding probe could assist both assessment of library fitness upon binding loop introduction, and subsequent selection of properly folded, soluble clones.

Several attempts have been made to confer binding capability to GFP by inserting binding loops into various solvent-exposed turns that connect the β-strands of the GFP β-barrel structure. The regions of GFP that are most amenable to insertion of amino acids have been determined (turns Gln-157-Lys-158 and Glu-172-Asp-173) (14, 15), although fluorescence is diminished substantially, and when random loops were inserted, the resultant library fluorescence decreased to 2.5% of wild type (14). Selection of GFP-inserted peptide libraries for targeting various intracellular compartments has also been performed (16). In addition, antibody heavy chain CDR3 sequences have been inserted into several loop regions of superfolder GFP, a GFP variant evolved for high stability and improved folding kinetics (17), to create libraries of single CDR3-inserted GFP. Results from this study indicated that insertion at many sites substantially reduces GFP fluorescence as seen previously with standard GFP variants (18). Three loop regions of the superfolder GFP, however, tolerated single loop CDR insertions (including Asp-173-Gly-174) such that it was possible to isolate fluorescent binders against protein targets using T7 phage display, with the best being a 470 nM lysozyme binder (19). This level of affinity is in the realm of that found for peptide binders (20), likely as a consequence of its single-binding loop design. Affinity of GFP-based binding proteins could therefore in principle benefit from display of multiple binding loops that could act together to form a cooperative binding interface. However, the lone examples of multiple loop insertion into GFP include insertion of hemagglutinin peptide (21) or random loops (22) into 2 loops on opposite faces of GFP. While suitable for the authors' goals, these insertion locations would not be ideal for forming a cooperative binding interface. Moreover, GFP fluorescence of the resulting clones in the case of the random loop libraries was not demonstrated (22). Thus, to date, robust fluorescent multiple loop-inserted GFP repertoires have not been described, even using the superfolder GFP as a template, likely because studies have used preexisting GFP variants that while bright and stable have not been optimized for binding loop insertion. Thus, in this study, the GFP scaffold itself was evolved to maintain its fluorescence properties in the presence of 2 inserted binding loops, and we demonstrated that scaffolds designed in this way were capable of accepting a diverse loop repertoire from which fluorescent binding proteins could be isolated.

Results

Effects of Single and Dual Loop Insertions on GFP Expression and Fluorescence. The initial goal of this study was to evaluate the capability of monomeric yeast enhanced green fluorescent protein (GFPM) (23, 24) (see *Materials and Methods* for details) to accommodate dual loop insertions. The Glu-172-Asp-173 turn region was chosen as 1 insertion site since earlier studies have shown that GFP can retain its fluorescence upon insertions of various lengths at this location (14, 15) (Fig. 1*A*). The second location selected was turn Asp-102-Asp-103 because of its proximity (≈1.6 nm) to Glu-172-Asp-173 on the same face of the β-barrel allowing the eventual possibility of improved affinity for targets through cooperative

Author contributions: T.V.P., Y.K.C., and E.V.S. designed research; T.V.P. and Y.K.C. performed research; T.V.P., Y.K.C., and E.V.S. analyzed data; and T.V.P., Y.K.C., and E.V.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed at: Department of Chemical and Biological Engineering, University of Wisconsin, 1415 Engineering Drive, Madison, WI 53706. E-mail: shusta@engr.wisc.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0902828106/DCSupplemental.

Impossibility of successful classification when useful features are rare and weak

Jiashun Jin[1]

Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213

Communicated by David L. Donoho, Stanford University, Stanford, CA, April 9, 2009 (received for review February 21, 2009)

We study a two-class classification problem with a large number of features, out of which many are useless and only a few are useful, but we do not know which ones they are. The number of features is large compared with the number of training observations. Calibrating the model with 4 key parameters—the number of features, the size of the training sample, the fraction, and strength of useful features—we identify a region in parameter space where no trained classifier can reliably separate the two classes on fresh data. The complement of this region—where successful classification is possible—is also briefly discussed.

higher criticism | phase diagram | region of impossibility | region of possibility | threshold feature selection

An overwhelming trend in modern research activity is the tendency to gather very large databases and use them to search for good data-based classifier rules. For example, currently, a very large number of research teams in the medical sciences seek to gather and study gene expression microarray data in hopes of obtaining empirical rules that separate healthy patients from those affected by a disease—thus allowing for automatic diagnosis.

Much of the current surge of enthusiasm for such studies stems from the advent of high-thoughput methods that automatically, on each subject, make measurements of a very large numbers of features. In genomics, proteomics, and metabolomics it is now common to take several thousand automatic measurements per study subject. The opportunity to survey so many features at once is thought to be valuable: optimists will say that "surely somewhere among these many features will be a few useful ones allowing for successful classification!"

Advocates of the optimistic viewpoint must contend with the growing awareness in at least some fields that many published associations fail to replicate—i.e., the published classification rules simply do not work when applied to fresh data. Such failure has been the focus of meetings and special publications (1). Although there may be many reasons for failure to replicate (2, 3), we focus here on one specific cause: there may simply be too many useless features being produced by high-throughput devices, so that, even where there really are decisive features to be found in the high-throughput measurements, they simply cannot be reliably identified.

In fact, we establish in this article a specific "region of impossibility" for feature selection in classifier design. We identify settings with large numbers of measurements, some useful, some useless, where the subset of useful measurements, *if only it were known a priori*, would allow for training of a successful classifier; however, when the subset of useful features is not known, we show that no classifier-training procedure can be effective.

Specifically, we study a model problem introduced in refs. 4 and 5 where there are a large number of features, many of which are useless and a few of which are useful. In this model we consider a two-class classification problem where there are parameters controlling the fraction of useful features, the strength of the useful features, and the ratio between the number of observational units (e.g., patients) and the number of measured features (e.g., gene expression measurements). We identify a region in parameter space where, *with* prior knowledge of at least some useful features, success is possible, but *absent* such prior knowledge about the subset of useful features, no classifier built from the dataset is likely to separate the two classes on fresh data.

In companion work (5), we show that in the complement of this region, a specific method for classifier training—Higher Criticism Threshold feature selection (4)—does work, and so the results here are definitive.

Classification When Features Are Rare and Weak

Consider a two-class classification setting where we have a set of labeled training samples (Y_i, X_i), $i = 1, 2, \ldots, n$. Each label $Y_i = 1$ if X_i comes from class 1 and $Y_i = -1$ if X_i comes from class 2, and each feature vector $X_i \in R^p$. For simplicity, we suppose that the training set contains equal numbers of samples from each of the two classes, and that the feature vector obeys $X_i \sim N(Y_i\mu, I_p)$, $i = 1, 2, \ldots, n$, for an unknown mean contrast vector $\mu \in R^p$. Also, we suppose that the feature covariance matrix is the identity matrix. Extension to correlated cases is possible if side information about the feature covariance is available (see ref. 6, for example).

Following the two companion papers (4, 5), we consider the following *rare/weak feature model* (RW model), where the vector μ is nonzero in only an ϵ faction of coordinates, and the nonzero coordinates of μ share a common amplitude μ_0. Formally speaking, let $I_1, I_2, \ldots, I_p$ be samples from Bernoulli(ϵ), and let

$$\mu(j) = \mu_0 \cdot I_j, \quad 1 \le j \le p.$$

Let Z denote the vector of z scores corresponding to the training set: $Z(j) = (1/\sqrt{n}) \sum_{i=1}^{n} Y_i \cdot X_i(j)$. The j th z score arises in a formal normal-theory test of whether the j th feature is useless or useful. Under our assumptions, $Z \sim N(\sqrt{n}\mu, I_p)$; thus each coordinate of Z has expectation either 0 or $\tau = \sqrt{n}\mu_0$.

We assume $p \gg n$, ϵ is small, and τ is either small or moderately large (e.g., $p = 10,000$, $n = 100$, $\epsilon = 0.01$, $\tau = 2$). Because zero coordinates of μ are entirely noninformative for classification, the useful features are those with nonzero coordinates in μ. The parameters ϵ and τ can be set to make such useful features arbitrarily rare (by setting ϵ close to 0) and weak (setting τ small); we denote an instance of the rare/weak model by $RW(\epsilon, \tau; n, p)$.

Formally, our goal is to use the training data to design a classifier for use on fresh data. If we are given a new unlabeled feature vector X, we must then label it with a class prediction, i.e., attach a label $\hat{Y} = 1$ or $\hat{Y} = -1$. We hope that our predicted label $\hat{Y}$ is typically correct. The central problem is for which combinations (ϵ, τ, n, p) it is possible to train a classifier that can label Y correctly, and for which combinations is it not possible to do so?

Linking Rarity and Weakness to Number of Features. We now adopt an *asymptotic* viewpoint. We let the number of features p be the driving problem size descriptor, and for the purposes of calculation, we let p tend to infinity, and other quantities vary with p. We have checked that our asymptotic calculations are descriptive of actual classifier performance in realistic finite-sized problems, say

Author contributions: J.J. designed research, performed research, and wrote the paper.

The author declares no conflict of interest.

Freely available online through the PNAS open access option.

[1]E-mail: jiashun@stat.cmu.edu.

Stochastic Simulation of Signal Transduction: Impact of the Cellular Architecture on Diffusion

Michael T. Klann,* Alexei Lapin, and Matthias Reuss
Institute of Biochemical Engineering and Center Systems Biology, Universität Stuttgart, Stuttgart, Germany

ABSTRACT The transduction of signals depends on the translocation of signaling molecules to specific targets. Undirected diffusion processes play a key role in the bridging of spaces between different cellular compartments. The diffusion of the molecules is, in turn, governed by the intracellular architecture. Molecular crowding and the cytoskeleton decrease macroscopic diffusion. This article shows the use of a stochastic simulation method to study the effects of the cytoskeleton structure on the mobility of macromolecules. Brownian dynamics and single particle tracking were used to simulate the diffusion process of individual molecules through a model cytoskeleton. The resulting average effective diffusion is in line with data obtained in the in vitro and in vivo experiments. It shows that the cytoskeleton structure strongly influences the diffusion of macromolecules. The simulation method used also allows the inclusion of reactions in order to model complete signaling pathways in their spatio-temporal dynamics, taking into account the effects of the cellular architecture.

INTRODUCTION

The cellular response to external signals depends on the signal transduction from the plasma membrane to the respective targets of the signal. The biochemical pathway for the signal transduction process is known for many signals, e.g., epidermal growth factor (1). While biochemical reactions affect the number of the molecules carrying the signal, transport processes are needed to deliver these molecules to their targets on various locations in the cell, mostly the nucleus where they trigger the expression of certain genes. Some signaling cascades involve the active transportation of signaling molecules along the cytoskeleton by motor proteins (2). However, in many cases the translocation of the signaling molecules depends on undirected diffusion in the cell (3).

To model signal transduction realistically, the stochasticity caused by low particle numbers should be incorporated (4–6). In addition, the spatial aspects must be taken into account (7–11). These include the microscopic and heterogeneous cellular architecture as well as molecular crowding (12,13). Besides the existence of spatial restrictions, the mobility of signaling molecules can further decrease as a result of unspecific and transient binding to the cytoskeleton (14). Computer simulations enable a separate and combined analysis of the different effects in silico. This article focuses solely on the spatial aspects and investigates the influence of molecule size as well as differences in the architecture of the cytoskeleton network on the diffusion of inert tracer molecules. By adjusting cytoskeleton parameters so that the diffusion results fit to measured data, one may also derive additional information about the (cytoskeleton) structures in cells.

Blum et al. (15) were the first to calculate the effect of the cytoskeleton on diffusion. However, an analytical solution for the diffusion in a cytoskeleton structure was only possible in a regular lattice structure. Monte Carlo simulations on the particle level facilitate geometries that are more realistic. A particle-based model to analyze diffusion due to forbidden spaces in the cell was first used by Ölveczky and Verkman (16) and later in the software environment of Smoldyn, but only using plates or cubes as fixed obstacles (17). MCell, another simulation environment, allows complex compartment geometries but does not include a cytoskeleton (18). Pogson et al. (19) included binding to actin structures but did not explore the effect on diffusion. The stochastic simulation framework presented here tackles the diffusion problem based on a realistic model for the cytoskeleton (see Fig. 1). In this framework, binding to the cytoskeleton and reactions can be included as well to build a realistic signal transduction simulation (20). In addition, all structures and molecules of the simulation can be visualized to provide a three-dimensional impression of the intracellular processes (21).

The diffusion of molecules in the cell is experimentally determined using fluorescence recovery after photobleaching or fluorescence correlation spectroscopy (FCS) methods (22–25). Overall, a size-dependent hindrance of diffusion was observed. In addition, the measurements not only revealed the reduced diffusion but also subdiffusion. In this case, the diffusion coefficient changes over time due to spatial heterogeneity. Subdiffusion depends on the level of crowding (26,27). Diffusion, and to a greater extent, subdiffusion, can limit biochemical reaction rates (28) and thus, affect the dynamics of signal transduction (29).

The hindrance in diffusion depends on various parameters such as cytoskeleton volume fraction, filament alignment, or particle size. Empirical formulas can describe some

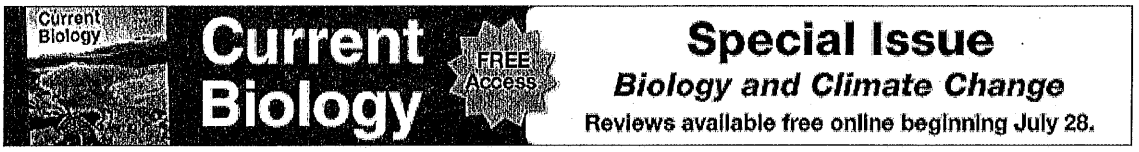
Submitted November 27, 2008, and accepted for publication March 27, 2009.

*Correspondence: klann@ibvt.uni-stuttgart.de

Editor: Jason M. Haugh.

0006-3495/09/06/5122/8 $2.00

doi: 10.1016/j.bpj.2009.03.049



Biophysical Journal, Volume 96, Issue 7, 2935-2948, 8 April 2009
doi:10.1016/j.bpj.2008.12.3943

Previous Article | Table of Contents | Next Article
Add Bookmark

ARTICLE

Absolute Quantitation of Bacterial Biofilm Adhesion and Viscoelasticity by Microbead Force Spectroscopy

Peter C.Y. Lau †, ‡, **John R. Dutcher** †, §, **Terry J. Beveridge** †, ‡ **and Joseph S. Lam** †, ‡, ✉, ✉

† Biophysics Interdepartmental Group, University of Guelph, Guelph, ON N1G 2W1, Canada

‡ Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON N1G 2W1, Canada

§ Department of Physics, University of Guelph, Guelph, ON N1G 2W1, Canada

✉ Corresponding author

Abstract

Bacterial biofilms are the most prevalent mode of bacterial growth in nature. Adhesive and viscoelastic properties of bacteria play important roles at different stages of biofilm development. Following irreversible attachment of bacterial cells onto a surface, a biofilm can grow in which its matrix viscoelasticity helps to maintain structural integrity, determine stress resistance, and control ease of dispersion. In this study, a novel application of force spectroscopy was developed to characterize the surface adhesion and viscoelasticity of bacterial cells in biofilms. By performing microbead force spectroscopy with a closed-loop atomic force microscope, we accurately quantified these properties over a defined contact area. Using the model gram-negative bacterium *Pseudomonas aeruginosa*, we observed that the adhesive and viscoelastic properties of an isogenic lipopolysaccharide mutant wapR biofilm were significantly different from those measured for the wild-type strain PAO1 biofilm. Moreover, biofilm maturation in either strain also led to prominent changes in adhesion and viscoelasticity. To minimize variability in force measurements resulting from experimental parameter changes, we developed standardized conditions for microbead force spectroscopy to enable meaningful comparison of data obtained in different experiments. Force plots measured under standard conditions showed that the adhesive pressures of PAO1 and wapR early biofilms were 34 ± 15 Pa and 332 ± 47 Pa, respectively, whereas those of PAO1 and wapR mature biofilms were 19 ± 7 Pa and 80 ± 22 Pa, respectively. Fitting of creep data to a Voigt Standard Linear Solid viscoelasticity model revealed that the instantaneous and delayed elastic moduli in *P. aeruginosa* were drastically reduced by lipopolysaccharide deficiency and biofilm maturation, whereas viscosity was decreased only for biofilm maturation. In conclusion, we have introduced a direct biophysical method for simultaneously quantifying adhesion and viscoelasticity in bacterial biofilms under native conditions. This method could prove valuable for elucidating the contribution of genetic backgrounds, growth conditions, and environmental stresses to microbial community physiology.

Article Information

PDF (1138 kb)
Full Text with Thumbnail Figures
Full Text with Large Figures
Supporting Material
Export Citation

PubMed

Articles by Peter C.Y. Lau
Articles by Joseph S. Lam

Related Articles

Viscoelastic Indentation of Extremely Soft Biological Sample...
Power-Law Rheology of Isolated Nuclei with Deformation Mappi...
Rheology of Passive and Adhesion-Activated Neutrophils Probe...

...more

Copyright © 2009 Biophysical Society. All rights reserved.
Biophysical Journal, Volume 96, Issue 8, L50-L52, 22 April 2009
doi:10.1016/j.bpj.2009.01.035

Previous Article | Table of Contents | Next Article
Add Bookmark

LETTER

Tracking of Single Quantum Dot Labeled EcoRV Sliding along DNA Manipulated by Double Optical Tweezers

Andreas Biebricher †, ✉, **Wolfgang Wende** ‡, **Christophe Escudé** §, **Alfred Pingoud** ‡ and **Pierre Desbiolles** †

† Laboratoire Kastler Brossel, ENS, Université Pierre et Marie Curie Paris 6, CNRS UMR 8552, Paris, France
‡ Justus-Liebig-Universität Gießen, Institut für Biochemie, FB 8, Gießen, Germany
§ Muséum National d'Histoire Naturelle, INSERM U565, CNRS UMR 8646, Paris, France

✉ Corresponding author

Abstract

Fluorescence microscopy provides a powerful method to directly observe single enzymes moving along a DNA held in an extended conformation. In this work, we present results from single EcoRV enzymes labeled with quantum dots which interact with DNA manipulated by double optical tweezers. The application of quantum dots facilitated accurate enzyme tracking without photobleaching whereas the tweezers allowed us to precisely control the DNA extension. The labeling did not affect the biochemical activity of EcoRV checked by directly observing DNA digestion on the single molecule level. We used this system to demonstrate that during sliding, the enzyme stays in close contact with the DNA. Additionally, slight overstretching of the DNA resulted in a significant decrease of the 1D diffusion constant, which suggests that the deformation changes the energy landscape of the sliding interaction. Together with the simplicity of the setup, these results demonstrate that the combination of optical tweezers with fluorescence tracking is a powerful tool for the study of enzyme translocation along DNA.

Article Information

PDF (236 kb)
Full Text with Thumbnail Figures
Full Text with Large Figures
Supporting Material
Export Citation

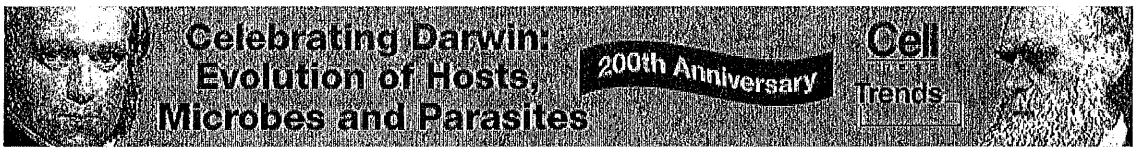
PubMed

Articles by Andreas Biebricher
Articles by Pierre Desbiolles

Related Articles

Neurexin/Neuroligin Interaction Kinetics Characterized by Co...
Single-Particle Tracking of Membrane Protein Diffusion in a ...
New Light on Quantum Dot Cytotoxicity

...more



Biophysical Journal, Volume 96, Issue 10, 4013-4023, 20 May 2009
doi:10.1016/j.bpj.2009.02.064

Previous Article | Table of Contents | Next Article
Add Bookmark

ARTICLE

Optimal Feedback Strength for Noise Suppression in Autoregulatory Gene Networks

Abhyudai Singh and Joao P. Hespanha

Department of Electrical and Computer Engineering, University of California, Santa Barbara, California

Corresponding author

Abstract

Autoregulatory feedback loops, where the protein expressed from a gene inhibits or activates its own expression are common gene network motifs within cells. In these networks, stochastic fluctuations in protein levels are attributed to two factors: intrinsic noise (i.e., the randomness associated with mRNA/protein expression and degradation) and extrinsic noise (i.e., the noise caused by fluctuations in cellular components such as enzyme levels and gene-copy numbers). We present results that predict the level of both intrinsic and extrinsic noise in protein numbers as a function of quantities that can be experimentally determined and/or manipulated, such as the response time of the protein and the level of feedback strength. In particular, we show that for a fixed average number of protein molecules, decreasing response times leads to attenuation of both protein intrinsic and extrinsic noise, with the extrinsic noise being more sensitive to changes in the response time. We further show that for autoregulatory networks with negative feedback, the protein noise levels can be minimal at an optimal level of feedback strength. For such cases, we provide an analytical expression for the highest level of noise suppression and the amount of feedback that achieves this minimal noise. These theoretical results are shown to be consistent and explain recent experimental observations. Finally, we illustrate how measuring changes in the protein noise levels as the feedback strength is manipulated can be used to determine the level of extrinsic noise in these gene networks.

Article Information

PDF (370 kb)
Full Text with Thumbnail Figures
Full Text with Large Figures
Supporting Material
Export Citation

PubMed

Articles by Abhyudai Singh
Articles by Joao P. Hespanha

Related Articles

Postsynaptic Mad Signaling at the Drosophila Neuromuscular J...
Erratum
DNA as Membrane-Bound Ligand-Receptor Pairs: Duplex Stabilit...

...more

Published June 29, 2009

Accuracy and precision in quantitative fluorescence microscopy

Jennifer C. Waters

Harvard Medical School, Department of Cell Biology, Boston, MA 02115

The light microscope has long been used to document the localization of fluorescent molecules in cell biology research. With advances in digital cameras and the discovery and development of genetically encoded fluorophores, there has been a huge increase in the use of fluorescence microscopy to quantify spatial and temporal measurements of fluorescent molecules in biological specimens. Whether simply comparing the relative intensities of two fluorescent specimens, or using advanced techniques like Förster resonance energy transfer (FRET) or fluorescence recovery after photobleaching (FRAP), quantitation of fluorescence requires a thorough understanding of the limitations of and proper use of the different components of the imaging system. Here, I focus on the parameters of digital image acquisition that affect the accuracy and precision of quantitative fluorescence microscopy measurements.

What information is present in a fluorescence microscopy digital image?

Quantitative microscopy measurements are most often made on digital images. A digital image is created when the optical image of the specimen formed by the microscope is recorded by a detector (usually a charge-coupled device [CCD] camera [Moomaw, 2007; Spring, 2007] or photomultiplier tube [PMT; Art, 2006]) using a two-dimensional grid of equally sized pixels. The pixels spatially sample the optical image, such that each pixel represents a defined finite sized area in a specific location in the specimen.

Correspondence to jennifer_waters@hms.harvard.edu

During acquisition of the digital image, the photons that are detected at each pixel are converted to an intensity value that is correlated to, but not equal to, the number of detected photons (Pawley, 2006c). In fluorescence microscopy, the intensity value of a pixel is related to the number of fluorophores present at the corresponding area in the specimen. We can therefore use digital images to extract two types of information from fluorescence microscopy images: (1) spatial, which can be used to calculate such properties as distances, areas, and velocities; and (2) intensity, which can be used to determine the local concentration of fluorophores in a specimen.

Accuracy and precision

Every quantitative measurement contains some amount of error. Error in quantitative fluorescence microscopy measurements may be introduced by the specimen, the microscope, or the detector (Wolf et al., 2007; Joglekar et al., 2008). Error shows itself as inaccuracy and/or imprecision in measurements. Inaccuracy results in the wrong answer. For example, with an inaccurate pH meter one might carefully measure the pH of a basic solution many times, each time finding the pH to be 2.0. Imprecision, on the other hand, results in variance in repeated measurements and therefore uncertainty in individual measurements. With an imprecise pH meter, repeated measurements of a solution with pH 7.0 might have a distribution ranging from 5.0–9.0, with an average value of 7.0. Although the average value of these repeated measurements is accurate, any individual measurement may be inaccurate. The importance of accuracy is obvious. Precision is equally important in quantitative fluorescence microscopy because we are often forced to make only one measurement (for example, one time-point in a live-cell time-lapse experiment). In addition, we are usually measuring biological specimens that have some level of natural variability, so variance seen in measurements made on different cells will be caused by both biological variability and that which is introduced when making the measurement. To use a fluorescence microscope and digital detector to quantitate spatial and intensity information from biological specimens, we must understand and reduce the sources of inaccuracy and imprecision in these types of measurements.

Signal, background, and noise

In quantitative fluorescence microscopy, we want to measure the signal coming from the fluorophores used to label the object of interest in our specimen. For example, consider live cells expressing GFP-tubulin in which we wish to measure the amount of tubulin polymer. The signal we are interested in is the photons emitted from GFP bound to tubulin incorporated into microtubules. We use the pixel intensity values in the digital image to localize the tubulin polymers and make conclusions about the quantity of microtubules. However, the intensity values in the digital images of

The Rockefeller University Press $30.00
J. Cell Biol. Vol. 185 No. 7 1135–1148
www.jcb.org/cgi/doi/10.1083/jcb.200903097