Opinion

Fast, cheap and somewhat in control

Adam P Arkin*†‡, Daniel A Fletcher*†

Addresses: *Howard Hughes Medical Institute, Department of Bioengineering, University of California, Berkeley, CA 94720, USA. †Physical Biosciences Division, E.O. Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. ‡Virtual Institute of Microbial Stress and Survival, Berkeley, CA 94710, USA.

Correspondence: Adam P Arkin. E-mail: aparkin@lbl.gov

Published: 30 August 2006

Genome ***Biology*** 2006, **7:**114 (doi:10.1186/gb-2006-7-8-114)

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2006/7/8/114

Abstract

Efforts to manipulate living organisms have raised the question of whether engineering principles of hierarchy, abstraction and design can be applied to biological systems. Here, we consider the practical challenges to controlling living organisms that must be surmounted, or at least managed, if synthetic biology and cellular bioengineering are to be productive.

Manipulating biology through the ages

The natural world around us is not quite so natural. Over many generations, societies have engaged in a struggle to mold nature to serve the needs, real and perceived, of their members. From cultivating grains to mining coal, we have sought to address local and global demands for food, shelter, health, and convenience through technology guided by science. One of the earliest and most profound of human engineering inventions was biotechnology in the form of farming, starting about 12,000 years ago. This was later transformed into a true domestication of animal and plant species that might be defined as "genetic alteration through conscious or unconscious selection" [1]. It provided the key foundation for the spread and stability of human societies and has become one of the most central and longest-lived sciences there is. In more modern times, a scientific/rational basis for domestication and control of biological organisms has been sought both to make breeding organisms for human purposes quicker and more successful, and to limit the spread of infectious disease and other invasive species. The eradication of smallpox and the near-eradication of polio stand as reminders of how advanced medical, industrial and social engineering can change the health of an entire world. The precision of molecular biology has led to a whole new form of domestication and industry. Launched by the first mass production of a human protein (somatostatin) in bacteria, industrial genetic engineering became one of the most transforming industries of the 20th century [2]. After the introduction of genetically engineered herbicide- and insect-resistant crops in 1995, genetically engineered maize is now more than 20% of the US crop, and approximately 80% of the US soybean crop is now genetically engineered [3].

Engineered biological systems are being used to address a wide variety of society's needs. Examples include the production of insulin and more than 200 other biopharmaceuticals and countless natural products, industrial catalysts and bioenergy substrates such as sugars, ethanol, methane and hydrogen, in microbes, eukaryotic cells and higher organisms; engineered resistance, ripening, oil production and nutrient overexpression in plants [4]; and more exotic successes such as the use of cytokine-expressing *Mycobacterium bovis* BCG as an effective treatment for certain forms of bladder cancer [5]. These successes have largely been 'one offs', however; each one is a special case, and while lessons were learned, they do not provide a definitive roadmap for the next advance. Could it be any different? Could each small success make the solution of major problems easier? The increasing number of such special cases, as well as society's growing need for solutions to energy and environmental problems, presages the need for a more rational and integrated approach to engineering biology.

We can learn lessons from other engineering fields. Products such as the personal computer and cellular phone have at

their foundation deep fundamental theory and technology from solid-state physics, materials science, and computational and information theory. These foundations enable the predictable control of materials and processes through the application of physical laws to meet specific objectives. In addition, the industry surrounding these devices has defined a set of standards and protocols that make the parts and systems of these devices (often) interoperable, extensible and, most importantly, allow the efficient scale-up of manufacture and distribution of the technology. The practice of engineering has a broadly successful track record of addressing social needs (as well as creating additional problems) when applied to materials like silicon and steel, where the physical rules are known and the complexity is limited or very well controlled. Engineering as a practice has not, however, been successful in such endeavors as controlling the weather or avoiding natural disasters, due largely to the scale, complexity and uncertainty inherent to those problems. So we must ask: is the conventional paradigm of engineering appropriate for biology? Can we develop, or deal with, the lack of a coherent theoretical and physical foundation for living systems? Or is control of biology destined for the same fate as rainmaking?

Why should biology be engineerable?

Given the complexity of biology, an engineering approach based on design may seem an unlikely route to success. Living systems, unlike classical engineered systems, grow and evolve, and have material properties that are not easily controlled or predicted and that are often sensitive to their local environment. Indeed, traditionally directed evolution through selective breeding of, for example, pest-resistant or high-yield crops has been the main method for obtaining a desired outcome in biology. This is true at the biomolecular level as well, in which creation of new function in proteins and nucleic acids is often accomplished through directed evolution rather than *de novo* design [6,7]. The technology for direct, rational manipulation of an organism's DNA has improved in precision and efficiency, greatly increasing our ability to produce therapeutics, natural products, antibodies and enzymes in heterologous systems. Yet actually achieving, let alone optimizing, the production of a given target in a given system is still a time-consuming art driven by decades of empirical observation. Engineering of more complex behaviors will require a more principled understanding of biological system design.

Basic research over the past few decades has given us confidence that there is at least some organized structure to the workings of cells - a structure that may be altered or even rebuilt through an intelligent process of engineering. This view is emerging from multiple disciplines. Comparative genomics is helping to uncover the structure and evolution of the genome. Large-scale tracking of DNA expression, protein synthesis, molecular interactions and intermediates is revealing groups of molecules that work and play together. Fluorescent imaging of living cells is identifying time-dependent changes in protein activity and localization that correlate with behavior. Reconstitution of biochemical and biophysical processes from 'minimal systems' of proteins has built confidence that top-down and bottom-up approaches to biology meet somewhere in the middle. Systems biology has sought to integrate these results and data to reverse-engineer an understanding of biological network function and dynamics. Finally, the infrastructure for storing and disseminating information on biological systems, and for modeling them, has grown concurrently. In turn, this allows the rapid access and cross-comparison of information that is critical to establishing data quality and creating interoperability standards that will enable biologists to leverage their efforts and build scalable systems.

The key observation that biological systems exhibit some degree of modularity underlies the current belief that useful and 'engineerable' design principles exist [8]. Whether at the level of protein motifs with similar binding properties or groups of proteins that carry out specific functions in a variety of distinct settings, the modular parts of biological systems are used and reused to generate and control the apparently complex behavior of living organisms. The bold question that was asked at the dawn of recombinant DNA research, and continues to be asked today, is whether a growing understanding of this modularity and new tools to manipulate it can be used to engineer new and useful behavior. Attempts to directly answer this question - and to think about its consequences - have resulted in the formation of a loose assembly of scientists, engineers, ethicists and other thinkers engaged in what has become known as 'synthetic biology'.

What sets synthetic biology apart from molecular biology and its closely allied fields of genetic and metabolic engineering is the ambition to formalize the process of designing cellular systems, in the way that traditional engineering disciplines have formalized design and manufacture, so that complex behaviors can be achieved for practical ends. Such behaviors will require larger biochemical circuits, typically encoded in DNA, for control. To achieve this, synthetic biologists look to move beyond the qualitative and often *ad hoc* engineering pathways that have underlain the slow progress to this point. The goal, instead, is to create a systematic engineering science founded on the standardization of a cellular 'chassis' - the types of parts available, their manufacture, their characterization and protocols for their interconnection - analogous to those that underlie and enable the scalability of mechanical, electrical and civil engineering. But the analogy with traditional engineering should not be taken too far, as there are challenges to engineering biology that no internal combustion engine or microprocessor has faced.

What are the engineering challenges?

Despite much effort, the dream of engineering biology has not yet led to simple and rapid construction of biological organisms that address specific problems. The reasons for this are twofold. One is the lack of a technology infrastructure that enables production of biological parts and easy assembly of these parts into systems, a challenge that has been addressed in a recent review [9]. The second, which we focus on here, is the difficulty of predicting what biological components will do, even when the parts are readily obtainable and much is known about them individually [10]. On this issue, lessons learned from engineering bridges, boats and planes are of little help, because the operating conditions under which biological systems function are significantly different from those of familiar macroscopic systems. Thermal fluctuations that drive stochastic behavior can typically be ignored or managed in traditional engineering, but often not in cells. And *in situ* evolutionary change in parts and control systems are simply not problems for inanimate objects - not so for biology. In fact, biology's success - its ability to grow and evolve new solutions and test fitness through competition - has depended on just those behaviors that frustrate predictability. Any engineering of biology to serve our needs must recognize, understand and manage this drive towards variation and the evolutionary competition with other organisms. Some of these issues are already under practical consideration in relation to genetically modified organisms [11].

Engineering exogenous protein or gene circuits into a new host organism also faces problems of integration due to 'parasitic' effects and cross-talk with existing pathways. Parasitic effects that arise due to direct interaction among new components or through indirect interactions via their effects on the organism into which they are introduced - the chassis - such as sickening it or draining inputs to other pathways, often play a dominant role in preventing circuit function. To be broadly useful, the features of a biological component and the organism into which it is introduced must be characterized such that its function is predictable. Most circuit designs rely, at least in part, in transferring natural components from other organisms into the host chassis. There are basic problems of adapting the part for operation in the new host by, for example, adjusting codon usage, and as the part did not coevolve with the other parts of the chassis it might cross-react with other components in unforeseen ways. Zarrinpar *et al*. [12] elegantly demonstrated this in yeast by showing that a yeast Pbs2 protein binds specifically to SH3 (Src homology 3) domains from yeast but is promiscuous with SH3 domains from other organisms. These issues are compounded when design moves away from the single cell and towards multicellularity, as some researchers are now attempting [13,14].

To demonstrate the challenges of engineering biology to control behavior, we use two simple examples: one addresses how evolution could degrade biological circuit performance over time; and the other addresses how noise, possibly external to the system, can have dramatic effects on system behavior. In the first example, we borrow a model for competition between multiple strains of a microbe under nutrient-limiting conditions in continuous culture [15]. We assume that there are two quasispecies of bacterium, one bearing a functional version of our synthetic biological circuit and the other carrying a disabled version created by deleterious mutation from the first. The deleterious mutation rate is a function of circuit size in base pairs (bp), basal mutation rate in base pairs per generation, the generation time itself, and a factor related to the specific circuit design, which gives the fraction of mutations leading to loss of function of the circuit. The basal mutation rate, m, in *Escherichia coli* is approximately 5.4×10^{-10} base-pairs per cell per generation; the fraction of mutations that actually disable the circuit, f, is a free parameter, which for this simulation we set to 1/1,000. The circuit size for a small two-gene circuit is approximately 2,000 bp. We vary this parameter in the following calculations, holding f constant for convenience, although, in reality, for each circuit design f is a variable. We also assume that having a functioning circuit places a metabolic load on the cell that slows its growth rate. This load has been observed experimentally in a number of cases, but there are few hard and fast rules [16-18]. A disabling mutation can release this load, leading to faster growth of the mutant population. The deleterious mutation rate sets a threshold above which, over time, the nonfunctional mutant population can outcompete the wild-type population.

Figure 1 shows the results of a calculation of the time for the mutant population to become the majority under different circuit sizes and growth differences, starting from an initial condition of a pure wild-type population. Even with relatively modest growth differences and relatively small circuit sizes it is only a matter of days before the mutant takes over the population. You *et al*. [19] observed this in a synthetic population-control circuit in which mutants that escape the circuit control arise in 3-6 days (the You circuit is approximately 4,000 bp long). In many ways, the above example is a best-case calculation because of the restriction to only the simplest of mutation mechanisms and the resulting population structure, as well as the relative mildness of the growth differences explored. In practice, the use of selectable markers could, of course, aid in preventing such takeovers. However, these markers may be disabled when the circuit function is disabled, and it is not obvious how to design for this when complex behaviors are encoded. Error-tolerant (robust) designs can minimize f, but the principles for applying this to biological design are still in their early development [20-22].

A second example of how an engineered biological system can evade control arises when we consider the biophysics of reaction networks in cells. It has become clear that the

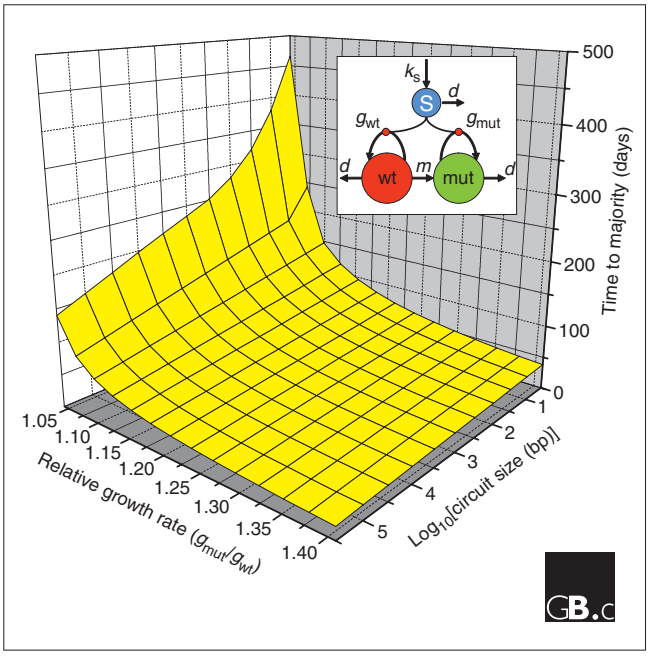

Figure 1
The takeover of a nonfunctional mutant with a higher growth rate in a population. The predicted time (in days) to a nonfunctional mutant strain of a synthetic microbe becoming the majority of the population as a function of the log of the circuit size and the ratio of growth rate of the mutant to that of wild type. The circuit size is a proxy for the cross-section of the circuit for deleterious mutation, which is also a function of growth rate, basal mutation rate *m* and circuit architecture. The larger the size and larger the growth advantage of the mutant strain, the faster the population loses function. The inset shows a schematic of the underlying model which tracks competitive growth (*g*) of a wild-type (wt) population and mutant (mut) population on a common resource (S) in continuous culture. *ks* is the influx rate of resource into the bioreactor and *d* is the dilution rate of cell and substrate out of the reactor. The parameter *m* is proportional to circuit size and is the rate of production of non-functional (and growth competitive) mutants from the wild-type population.

discrete and stochastic nature of chemical reactions can play an important role in cellular behavior, in part because many cellular processes are governed by small numbers of molecules. A number of recent papers describe synthetic biological constructs for exploring the effect of noise on these low-molecular number processes [23-30]. Even in cases where the numbers of molecules are not too small, stochastic effects can have surprising consequences. Theoretically, the addition of a small amounts of external noise to a ubiquitous biological network motif, the enzymatic futile cycle (in which a protein undergoes continuous cycles of phosphorylation and dephosphorylation under the control of kinases and phosphatases), can lead to different qualitative behavior than that predicted by the deterministic equations [31]. In fact, different types of noise can lead to dramatically different behaviors of the futile cycle, including different signal amplification, switching and oscillation properties. Figure 2 shows a simple analysis (from [31]) that shows the bifurcation from monostable to bistable behavior that occurs in a futile cycle as the distribution of the small noise term added to the forward enzyme is changed. There might be different exogenous noise sources under different environmental conditions in which the engineered organism finds itself, and thus this subcircuit could behave in 'unexpected' ways. In turn, these might impact greatly on the fitness of the organism [32-34].

The first of these two simple examples demonstrates that two of the key properties of a cellular chassis and its environment that need to be engineered are the basal mutation rate and the robustness to circuit load, which should, generally, both be minimized. Furthermore, the particular design choices made in part choice and in the mechanisms by which these parts are hooked together to make a system will affect the value of *f*, the deleterious mutation rate. The second of these examples shows how even a simple biochemical system can exhibit complex unintended behaviors if the environment in which it operates changes only its noise properties (even if the mean values stay the same). Thus a designed cell that is passing through uncertain or multiple environments will have to be designed to minimize or even to exploit these effects. In fact, outside bioreactors, engineered organisms - other than a few agricultural examples - survive poorly in real-world environments where conditions and competition with other organisms are less controlled [35]. Both the cases described here demonstrate special considerations that must be applied to the engineering of biological systems in order to meet the challenges to the scalability of engineered organisms.

Immediate goals and future prospects

As we have for millennia, we are shaping the biological world to meet our needs. There are major problems that cry out for biologically engineered solutions, such as those in cell and tissue engineering, gene therapy, biologically derived materials, biocatalysis and natural product synthesis, optimization of agricultural yield and nutrition, pest and disease control and much more. Synthetic biology, with its focus on elucidating and harnessing design principles of living systems, aims to tackle these problems. But unlike other engineering disciplines, synthetic biology has not developed to the point where there are scalable and reliable approaches to finding solutions. Instead, the emerging applications are most often kludges that work, but only as individual special cases. They are solutions selected for being fast and cheap and, as a result, they are only somewhat in control (with apologies to Errol Morris).

Yet there is optimism in the field. Engineering biology is indeed a great challenge, but its potential benefits are even greater. Through the creative efforts of many investigators, solutions to robustness and noise suppression may be found - or we will at least understand why no solutions can be

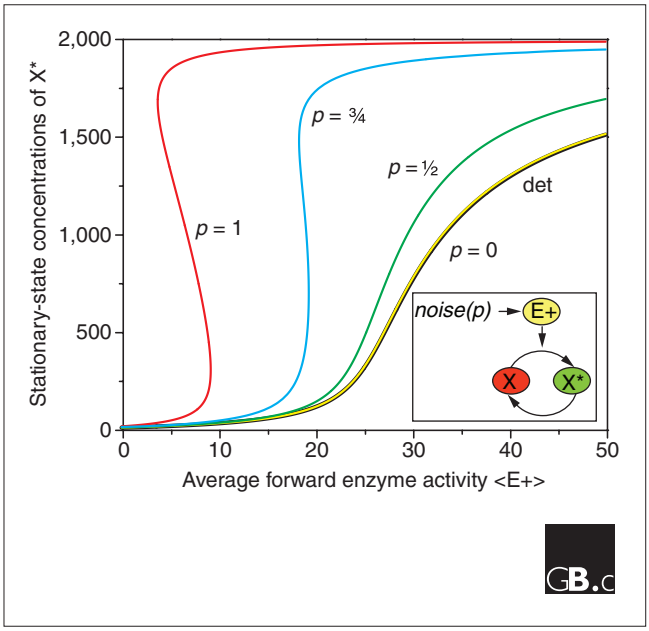

Figure 2
The effects of noise on an enzymatic futile cycle. The cycle is formed by the phosphorylation of a protein X to form X* through the action of a kinase, E+, which may or may not be subject to noise in its activity. Each curve is a plot of the stationary-state concentration of X*, from the system shown schematically in the inset, as a function of the average forward enzyme activity <E+>. The variable p is related to the noise power and determines the effective noise distribution around <E+>; $p = 0$ corresponds, for example, to approximately normally distributed noise whereas the other values correspond to different distribution shapes. The (black) curve labeled 'det' is the deterministic solution when E+ is not subject to noise. Whereas the deterministic system defined by $p = 0$ is monostable, the system with noise can be bistable and oscillate stochastically. From an analysis in [31].

found. Further effort and investment are required to develop robust theories and computational infrastructure for biological circuit design and synthesis, to establish standards in measurement and information about circuits and their interoperability, and to create new manufacturing technologies that allow production of large circuits, creation of novel chasses (with, for example, new genetic codes [36]), and environments for the development of artificial tissues. Excitement among those engaged in engineering biology stems from the fact that there are clear routes to progress on all of these fronts and from the incredible pull of the applications that are possible if these problems are solved.

Acknowledgements

We thank David Schaffer for a critical reading of the manuscript. In this work A.P.A. is supported by the Howard Hughes Medical Institute and by the Virtual Institute for Microbial Stress and Survival [37] supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics Program:GTL through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the US Department of Energy. D.A.F. acknowledges NSF and NIH funding during the course of this work.

References

1. Janick J: *Perspectives on New Crops and New Uses. Volume 4.* Edited by Janick J. Phoenix, Arizona: ASHS Press; 1999:104-110.
2. Itakura K, Hirose T, Crea R, Riggs AD Heyneker HL, Bolivar F, Boyer HW: **Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin.** *Science* 1977, **198:**1056-1063.
3. **Adoption of Genetically Engineered Crops in the US** [http://www.ers.usda.gov/Data/BiotechCrops]
4. Capell T, Christou, P: **Progress in plant metabolic engineering.** *Curr Opin Biotechnol* 2004, **15:**148-154.
5. Kassouf W, Kamat AM: **Current state of immunotherapy for bladder cancer.** *Expert Rev Anticancer Ther* 2004, **4:**1037-1046.
6. Bloom JD, Meyer MM, Meinhold P, Otey CR, MacMillan D, Arnold FH: **Evolving strategies for enzyme engineering.** *Curr Opin Struct Biol* 2005, **15:**447-452.
7. Bornscheuer UT: **Trends and challenges in enzyme technology.** *Adv Biochem Eng Biotechnol* 2005, **100:**181-203.
8. Wolf DM, Arkin AP: **Motifs, modules and games in bacteria.** *Curr Opin Microbiol* 2003, **6:**125-134.
9. Endy D: **Foundations for engineering biology.** *Nature* 2005, **438:**449-453.
10. Sprinzak D, Elowitz MB: **Reconstruction of genetic circuits.** *Nature* 2005, **438:**443-448.
11. Kirk TK, Carlson JE, Ellstrand N, Kapuscinski AR, Lumpkin TA, Magnus DC, Magraw DB, Nester EW, Peloquin JJ, Snow AA, *et al.*: *Biological Confinement of Genetically Engineered Organisms.* Washington DC: National Academies Press; 2004.
12. Zarrinpar A, Park SH, Lim WA: **Optimization of specificity in a cellular protein interaction network by negative selection.** *Nature* 2003, **426:**676-680.
13. Basu S, Mehreja R, Thiberge S, Chen MT, Weiss R: **Spatiotemporal control of gene expression with pulse-generating networks.** *Proc Natl Acad Sci USA* 2004, **101:**6355-6360.
14. Basu S, Gerchman Y, Collins CH, Arnold FH, Weiss R: **A synthetic multicellular system for programmed pattern formation.** *Nature* 2005, **434:**1130-1134.
15. Hansen SR, Hubbell SJ: **Single-nutrient microbial competition: quantitative agreement between experimental and theoretically forecast outcomes.** *Science* 1980, **207:**1491-1493.
16. Flores S, de Anda-Herrera R, Gosset G, Bolivar FG: **Growth-rate recovery of *Escherichia coli* cultures carrying a multicopy plasmid, by engineering of the pentose-phosphate pathway.** *Biotechnol Bioeng* 2004, **87:**485-494.
17. Neubauer P, Lin HY, Mathiszik B: **Metabolic load of recombinant protein production: inhibition of cellular capacities for glucose uptake and respiration after induction of a heterologous gene in *Escherichia coli*.** *Biotechnol Bioeng* 2003, **83:**53-64.
18. Haddadin FT, Harcum SW: **Transcriptome profiles for high-cell-density recombinant and wild-type *Escherichia coli*.** *Biotechnol Bioeng* 2005, **90:**127-153.
19. You L, Cox RS, Weiss R, Arnold FH: **Programmed population control by cell-cell communication and regulated killing.** *Nature* 2004, **428:**868-871.
20. McAdams HH, Arkin A: **Towards a circuit engineering discipline.** *Curr Biol* 2000, **10:**R318-R320.
21. Gerdes SY, Scholle MD, Campbell JW, Balazsi G, Ravasz E, Daugherty MD, Somera AL, Kyrpides NC, Anderson I, Gelfand MS, *et al.*: **Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655.** *J Bacteriol* 2003, **185:**5673-5684.
22. Albert R, Jeong H, Barabasi AL: **Error and attack tolerance of complex networks.** *Nature* 2000, **406:**378-382.
23. Swain PS, Elowitz MB, Siggia ED: **Intrinsic and extrinsic contributions to stochasticity in gene expression.** *Proc Natl Acad Sci USA* 2002, **99:**12795-12800.
24. Cai L, Friedman N, Xie XS: **Stochastic protein expression in individual cells at the single molecule level.** *Nature* 2006, **440:**358-362.
25. Pedraza JM, van Oudenaarden A: **Noise propagation in gene networks.** *Science* 2005, **307:**1965-1969.

26. McAdams HH, Arkin A: **Stochastic mechanisms in gene expression.** *Proc Natl Acad Sci USA* 1997, **94:**814-819.
27. Weinberger LS, Burnett JC, Toettcher JE, Arkin AP, Schaffer DV: **Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity.** *Cell* 2005, **122:**169-182.
28. Golding I, Paulsson J, Zawilski SM, Cox EC: **Real-time kinetics of gene activity in individual bacteria.** *Cell* 2005, **123:**1025-1036.
29. Newman JR, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, Weissman JS: **Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise.** *Nature* 2006, **441:**840-846.
30. Raser JM, O'Shea EK: **Control of stochasticity in eukaryotic gene expression.** *Science* 2004, **304:**1811-1814.
31. Samoilov M, Plyasunov S, Arkin AP: **Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations.** *Proc Natl Acad Sci USA* 2005, **102:**2310-2315.
32. Wolf DM, Vazirani VV, Arkin AP: **Diversity in times of adversity: probabilistic strategies in microbial survival games.** *J Theor Biol* 2005, **234:**227-253.
33. Kussell E, Leibler S: **Phenotypic diversity, population growth, and information in fluctuating environments.** *Science* 2005, **309:**2075-2078.
34. Thattai M, van Oudenaarden A: **Stochastic gene expression in fluctuating environments.** *Genetics* 2004, **167:**523-530.
35. Cases I, de Lorenzo V: **Genetically modified organisms for the environment: stories of success and failure and what we have learned from them.** *Int Microbiol* 2005, **8:**213-222.
36. Wang L, Xie J, Schultz PG: **Expanding the genetic code.** *Annu Rev Biophys Biomol Struct* 2006, **35:**225-249.
37. **VIMSS: Virtual Institute for Microbial Stress and Survival** [http://VIMSS.lbl.gov]