Programmed Development of Biological Organisms

The development of natural biological organisms provides a stunning example of how noisy, genetically identical cells or viruses can achieve complex, deterministic outcomes. I propose to study developmental processes using a combination of simulation and physical reconstruction of the governing biological networks. Construction of synthetic biological systems in support of understanding and analyzing natural biological systems can serve a number of purposes. *First*, recapitulation of observable functions provides a powerful test of our understanding of how system components interact to form a functioning whole (Sprinzak & Elowitz, 2005). Differences between models used to design synthetic systems and the physical instantiations of synthetic systems themselves point to gaps in our understanding of the governing components and interactions (Chan et al., 2005). Likewise, differences between synthetic and natural systems can lead to insight into what other functions the natural biological systems might be serving. *Second*, evolving surrogate synthetic systems can act as a tool to understand how the parts of the natural system contribute to system fitness, as well as provide a platform to understand how evolutionary processes can optimize system performance. *Third*, constructing more defined, measurable, and manipulable developmental systems provides a foundation for the rational engineering of novel self-replicating, self-assembling systems (Knight, 2005). Specifically, I will construct synthetic systems designed to control development to study the following natural developmental processes:

1. Bacteriophage development – Bacteriophage T7 is an obligate lytic phage that infects *E. coli* and is an excellent model organism for understanding how the genetic elements of a developmental program interact to form a functioning whole (Studier & Dunn, 1983). I propose to construct synthetic T7 genomes to better understand how the developmental program of T7 is organized and/or optimized to solve particular problems, such as host resource allocation across different developmental stages.

2. Bacterial development –*B. subtilis* processes and responds to one or more signals of nutrient deprivation and decides upon various courses of action such as degradive enzyme production, cannibalization, competence, motility, biofilm formation, and sporulation (Burkholder & Grossman, 2000; Phillips & Strauch, 2002; Gonzalez-Pastor et al., 2003). I will construct synthetic networks to control individual cell decisions as well as engineered signaling systems to coordinate these decisions across populations to understand how *B. subtilis* is able to apportion different responses to stress.

3. Metazoan development – Over the course of its development, *C. elegans* senses nutrient conditions, population levels, and temperature to decide whether to continue normal adult development or to 'wait it out' and develop into a more stress-resistant state called the dauer diapause (Golden & Riddle, 1984). I will use synthetic constructs to better understand how *C. elegans* senses and processes its internal and external state and then commits to a particular course of development.

Bacteriophage development

During development, bacteriophage T7 must direct host resources to different developmental stages such as the takeover of host gene expression machinery, replication of its genome, production and assembly of structural proteins, and finally lysis. I am particularly interested in how T7 development may be optimized to control the allocation of host cell resources across these different stages. To begin, we need models that can predict how changes to the existing genetic structure affect allocation of resources during development. T7 is particularly suited for building such models for at least three reasons. First, much of T7's development is decoupled from the host cell, and thus we can ignore much of the host's complexity when modeling development (Endy & Brent, 2001). Second, since genome entry is coupled to transcription, gene ordering determines the timing and level of gene expression (Garcia & Molineux, 1995). Finally, T7 genetic elements are sufficiently characterized to enable construction of empirical computational models for how the combination of these genetic elements is able to direct the processes of development (Endy et al., 1997). However, thus far these models are not able to predict the effects of some perturbations to the genome (Endy et al., 2000).

My work in Drew Endy's lab has focused on improving our understanding of T7 gene expression. We constructed a single-molecule base-pair resolved gene expression simulator called Tabasco that can explicitly model biophysical mechanisms thought to govern T7 development (Kosuri et al., in preparation). In addition, we made measurements of absolute copy numbers of various T7 mRNA during infection to better understand regulation and provide more stringent comparisons between models and measurements of gene expression (Keller et al., in preparation). However, discrepancies between these models and measurements still exist. We could continue to make our models ever more detailed in hope of better characterization of the natural system. Instead, we decided to design and construct a surrogate phage that is easier to understand and model. The new phage, T7.1, is a more direct encoding of our computational models of the T7 genome by virtue of the definition and physical separation of each known genetic element (Chan et al., 2005). Chimeras of two sections (~30%) of T7.1 with wild-type T7 were viable and provided us with significant encouragement to move forward.

I am currently designing a component library for a next generation phage, T7.2. At the CGR, I will construct a set of phages from this library that will extend the original goals of T7.1 by only encoding functions we believe to understand, while actively removing those that we do not. Specifically, iterations of T7.2 will remove genes with unknown or unimportant functions, replace natural transcriptional and translational elements with standard characterized ones, and encode reporters to make T7.2 more measurable. I expect that T7.2 will be easier to model because the system encoded by the redesigned genome will more likely include only those functions encoded within the model, and model testing and validation will be easier due to T7.2's increased manipulability and measurability. In addition, because we will have putative functions for most or all elements encoded on the T7.2 genome, experiments involving evolutionary optimization of T7.2 to particular environments will be more interpretable. In particular, evolution experiments in T7.2 could act as a debugging tool for understanding how altering gene expression patterns in T7 affects fitness in different environments (Dekel & Alon, 2005).

Finally, I will test how transcriptional feedback mechanisms in T7 may act to direct host resource allocation across the different stages of T7 development and buffering the phage to variations in infection conditions. T7 encodes three proteins that inhibit host and phage RNA polymerase. Simulations predict that the loss of feedback inhibition of the phage RNA polymerase leads to deficiencies in the phage's ability to direct host translational resources, especially towards gene expression late in development (Kosuri, unpublished). Since the timing and level of gene expression in T7 is dependent upon gene location, I will use T7.2's increased manipulability to alter the genomic positioning of feedback genes to explore if the natural organization is optimized to apportion the proper amount of resources over the course of gene expression.

Bacterial Development

Bacterial populations often exhibit multi-cellular behavior (Kaiser, 2001). *Bacillus subtilis* copes with nutrient deprivation by coordinating and distributing a suite of responses across the cell population such as degradive enzyme production, cannibalization, competence, motility, biofilm formation, and sporulation (Burkholder & Grossman, 2000; Phillips & Strauch, 2002; Gonzalez-Pastor et al., 2003, Kearns et al., 2005). Many of these processes are mutually exclusive. For example, a cell that has decided to sporulate cannot simultaneously activate processes such as competence or motility, and vice versa. While any individual cells will choose one particular state, populations of genetically identical cells will display a wide range of developmental outcomes. In nature, it can be advantageous for bacteria to be able to tune the apportioning of these developmental outcomes to their particular environment (Dekel & Alon, 2005). To better understand how natural pathways make, coordinate, and tune such decisions, I will use a synthetic approach to reconstruct the pathway from more defined individual components.

The decision of an individual bacterium in a population to enter into a developmental state is governed by a network of kinases, phosphates, transcriptional regulators, feed-back and feed-forward loops (Phillips & Strauch, 2002). This network modulates the activities of a few key regulators, among them Spo0A, AbrB, ComA, CodY, and SinR (Mole et al., 2003, Kunst et al, 1994; Sonenshein, 2005; Kearns et al., 2005). Cell-cell signaling is mediated by a set of 11 regulatory proteins (Rap) that are inhibited by 7 small pentapeptide signaling molecules (Phr) (Bongiorni et al., 2005; Perego, 1997). The signaling pathways are integrated into the decision network by modulating the activity of the phosphorelay that governs Spo0A activity (Phillips & Strauch, 2002). Others have made both computational models and experimental tests of those models for smaller subsets of the pathway (Voigt et al., 2005; Suel & Elowitz, unpublished).

I will begin studying this system by systematically removing as many of the major components involved in the developmental decision process as possible (almost all proteins are nonessential). Then I will characterize small subsets of the regulatory pathway. For example, I will individually reconstruct the phosphorelay controlling the activity of Spo0A, the Rap/Phr signaling systems, and the pathways of other regulators, such as CodY and AbrB. Then I can reconstruct these modules to recapitulate and

understand how *B. subtilis* is able to undergo and manage such a variety of responses to stress. Unlike rebuilding T7, this approach of removing and then reconstructing the pathways within the same organism may be confounded by unknown components that remain and have important functions. However, these unknown components are more likely to be elucidated by looking at simpler modules rather than at the entire system.

Metazoan development

If the reconstruction of decision pathways in *B. subtilis* is sufficiently encouraging, I will explore developmental decisions in higher organisms. Like bacteria, metazoans often have to decide between different developmental fates before becoming reproductive adults. One compelling example is the decision of an individual *C. elegans* to either continue normal development or proceed into a more stress-resistant, yet functionally crippled state called the dauer diapause (Riddle & Albert, 1997). *C. elegans*, like *B. subtilis,* processes cues from secreted pheromones from neighboring worms, low internal nutrient conditions, and increased temperature to direct development towards increasing percentages of dauer larvae (Golden & Riddle, 1984). The pheromone, daumone, is basally produced by all worms. Ciliary sensory neurons sense and process daumone through a set of signal transduction pathways that also sense energy levels such as cGMP and insulin pathways, resulting in final activation of the transcription factors DAF-12 and DAF-16 responsible for carrying out much of the dauer response (Antebi, 2005; Jeong et al., 2005). While many of the components and interactions of the signaling pathways are still unknown, the recent purification, identification, and synthetic synthesis of daumone by Jeong et al. should facilitate future genetic screens. Similar to the bacterial decision systems, I hope to better understand dauer development by making computational and/or phenomenological models of the system, and then testing the efficacy of those models explicitly by reprogramming the developmental processes to recapitulate the natural stress responses. Such an approach will benefit from the ease of manipulability, genetics, and measurability of *C. elegans*.

Conclusions

This work would be particularly well suited for study at the CGR because of the Center's expertise in developing and optimizing high-throughput measurement technologies. For example, the proposed science will require high-throughput methods for detecting RNA and/or protein expression levels in order to compare performance characteristics of genetic components of one system to another, and to systematically compare natural systems to their synthetic surrogates. Much of the work will depend upon quickly constructing large sections of DNA. I hope to leverage new companies that have been developing the ability to quickly and cheaply synthesize long pieces of DNA *in vitro*. If however, these companies are not reliable and/or I do not have the resources to afford synthesis, I expect to scale-up existing techniques (Smith et al., 2003; Kodumal et al., 2003; Tian et al., 2004) towards quickly and accurately construct DNA molecules the size of T7 sections (~10,000 bp). This work would undoubtedly benefit from the robotics instruments CGR has available.

For bacteria and worms, I am currently focusing on decision making processes in development. In the future we could reconstruct entire developmental pathways from

initial decision making, to commitment, to executing the subsequent developmental program. Such work would depend upon increases in synthesis technologies as well as a deeper understanding of the components of these large developmental systems.

Finally to reiterate, I feel using a synthetic approach to understand how to construct biological systems to make and then implement developmental decisions across individuals and populations will be useful for at least three reasons. First, recapitulation of observable phenotypes gives us a working endpoint for trying to understanding natural system function (Chan et al., 2005). Second, separation of developmental systems' performance from evolutionary relics and/or optimizations will give us a better platform to study how longer time scale attributes such as a network's evolvability impact the design of natural biological systems (Kirschner & Gerhart, 1998). Third, understanding how to build such developmental decisions into cells will provide a starting point for programming novel and useful developmental functions in the future.

Works Cited

Antebi A. 2005. *Cell Metab*. **1**:157-8.
Bongiorni C et al. 2005. *J Bacteriol*. **187**:4353-61.
Burkholder WF, Grossman AD. 2000. In: *Prokaryotic Development*, pp. 151-156.
Chan LY, Kosuri S, Endy D. 2005. *Mol Sys Biol*. doi:10.1038/msb4100025.
Dekel E, Alon U. 2005. *Nature*. **436**:588-92.
Endy D, Brent R. 2001. *Nature*. **409**:391-5.
Endy D, Kong D, Yin J. 1997. *Biotech Bioeng*. **55**:375-89.
Endy D, You L, Yin J, Molineux IJ. 2000. *Proc Natl Acad Sci USA*. **97**:5375-80.
Garcia LR, Molineux IJ. 1995. *J Bacteriol*. **177**:4066-76.
Golden JW, Riddle DL. 1984. *Dev Biol*. **102**:368-78.
Gonzalez-Pastor JE, Hobbs EC, Losick R. 2004. *Science*. **301**:510-3.
Jeong PY et al. 2005. *Nature*. **433**:541-5.
Kaiser D. 2001. *Annu Rev Genet*. **35**:103-23.
Keller H, Kosuri S, Endy D. *in preparation*.
Kearns DB, et al. 2005. *Mol Microbiol*. **55**:739-49.
Kirschner M, Gerhart J. 1998. *Proc Natl Acad Sci USA*. **95**:8420-7.
Kodumal SJ, et al. 2003. *Proc Natl Acad Sci USA* **101**:15573-8.
Kosuri S, Kelly JR, Endy D. *in preparation*.
Kunst F et al. 1994. *Res Microbiol*. **145**:393-402
Knight TF. 2005. *Mol Sys Biol*. doi:10.1038/msb4100033.
Molle V et al. 2003. *Mol Microbiol*. **50**:1683-701
Perego M. 1997. *Proc Natl Acad Sci USA*. **94**:8612-7.
Phillips ZEV, Strauch MA. 2002 *Cell Mol Life Sci*. **29**:392-402.
Riddle DL, Albert PS. 1997. In: *C. elegans II*, pp. 739-68.
Smith HO et al. 2003. *Proc Natl Acad Sci USA*. **100**:15440-5
Sonenshein AL et al. 2005. *Curr Opin Microbiol*. **8**:203-7
Sprinzak D, Elowitz MB. 2005. *Nature*. **438**:443-8.
Studier FW, Dunn JJ. 1983. *Cold Spring Harb Symp Quant Biol*. **47**:999-1007.
Tian J et al., 2004. *Nature*. **432**:1050.
Voigt CA, Wolf DM, Arkin AP. 2005. *Genetics*. **169**:1187-202.