IDENTIFICATION AND CONTROL OF GENE NETWORKS IN LIVING ORGANISMS VIA SUPERVISED AND UNSUPERVISED LEARNING

Michael E. Driscoll, Timothy S. Gardner

Center for BioDynamics and Department of Biomedical Engineering, Boston University
Boston, MA, USA

Abstract:
The stunning varieties of cellular behavior rival that of any engineered system. Cellular systems are capable of energy transduction, organic synthesis, signal processing, and environmental sensing – to name a few. They operate with energetic efficiency, at nanometer scales, and across a range of environmental conditions. Cellular functions such as cell growth, aerobic respiration, photosynthesis, and locomotion are governed by networks of interacting genes, proteins, and metabolites. A principal challenge in biology is to uncover the structure and dynamics of these networks, and connect these properties to biological functions. Such knowledge will enable advances in the development of medicines, microbes optimized for environmental remediation, and biologically derived energy sources. Here, we review recent approaches for the determination of the structure of genetic and biochemical networks regulating cellular function.

Keywords:, biotechnology, statistical inference, learning algorithms

SYSTEMS BIOLOGY AND GENE REGULATORY NETWORKS

The investigation of cells from a systems-level perspective is a relatively recent trend, enabled by parallel advances in computational and experimental technology, such as high-throughput gene sequencing. For much of the last century, biologists have had a gene-centric view of living systems. The "one gene, one enzyme" doctrine, first proposed by Beadle and Tatum [1941], was gradually replaced by a "one gene, one function" view, where every gene is matched with a biological function. However, the complete genomic sequences of model organisms, from microbes to mammals, have revealed that most organisms contain far fewer genes than had been expected. The implication is clear: the versatility of cellular systems does not arise from single genes per se, but from networks of interacting genes. The emerging field of "systems biology" seeks to address this reality, with the aim of characterizing ensembles of genes and proteins, rather than single genes in isolation.

Gene networks regulate much of a cell's internal activity, with many of the network's components having regulatory as opposed to structural (e.g. enzymatic or metabolic) roles. In the example of the DNA-damage response network in *E. coli*, tens of genes regulate the response of hundreds of genes involved in DNA repair and cell survival. The role of such networks is to integrate a variety of environmental stimuli and produce an appropriate response, in terms of the expression of structural genes required by the cell to survive.

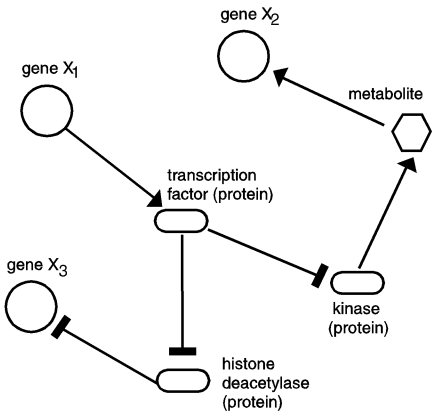


Figure 1. **A** A gene regulatory network consisting of three genes, three proteins, and a metabolite. Activation is represented with arrows, repression is represented with bars. **B** A reduced model of this network, consisting only of indirect gene interactions.

Gene networks consist not only of genes but of proteins and small molecules called metabolites. Genes are "expressed" upon the directed synthesis of a protein encoded by their DNA, by way of an RNA intermediate. Proteins in turn possess most of the enzymatic, mechanical, electrical and other properties needed to execute cell functions.

Gene, proteins, and metabolites regulate one another in myriad ways. Proteins bind to DNA at highly specific regions called cis-regulatory sites, and influence the expression of nearby genes. Proteins also act on one another directly, mediating the addition of catalytically important chemical groups, or combine to form multi-protein complexes, such as the those involved in gene expression. Metabolites can also bind to proteins and alter their activity. Experimentally, it is often difficult to determine the physical interactions underlying a regulatory relationship. In models of these networks, it is common to simplify and describe these interactions in terms of whether and how a gene inhibits or promotes the activity of another gene (Fig).

Identifying the structure of these regulatory networks could impact diverse areas of technology including medicine, public health, bioprocess engineering, bioremediation, and renewable energy. Below, we survey some existing experimental and analytical approaches for network identification, and a few promising applications of this technology.

MAPPING AND MODELING GENE REGULATORY NETWORKS

Uncovering the structure and dynamics of gene networks involves both experimental observation and computational analysis. Two approaches are commonly taken: one measures the physical interactions in the network, the other measures changes in the output of the network under different conditions, and infers the interactions.

The "physical" approach measures the protein-protein or protein-DNA interactions in a network, capturing cells at a given time and condition, and identifying the components of these interactions through chemical assays. Although these assays tend to be fraught with false-positives and systematic bias, they often represent the best form of direct evidence for a regulatory interaction. The downside of this approach is that it cannot tell us the functional nature of these interactions; we cannot know their strength or whether they are activating or inhibitory.

The "inference" approach measures how the concentration of gene products [1] change over a series of time-points or across conditions, and typically uses microarray technology [2] to measure gene expression. In this approach, the computational analysis plays a major role since interactions are inferred indirectly from experimental data. The recovered model is also indirect; it describes only the overall relationships between genes, but may not describe physical interactions(Fig B). Nonetheless, this approach has shown great potential of describing and predicting network functions. Moreover, due to the widespread availability of microarrays, this inference approach is growing in popularity and is the focus of our discussion below.

Network Inference via Supervised Learning

The strategies used for network inference in biology often involve some form of parameter estimation. Such methods are alternately called system identification, machine learning, or supervised learning. Here we will refer to them as supervised learning. In such an approach, a model structure is chosen to represent the genetic network based on multiple considerations, including the physical nature of interactions, prior information experimental constraints, and the type, quality and quantity of data available. The parameters of the model are then estimated such that the model is able to best reproduce the experimental data collected. If the approach is successful, the estimated model may then be used to make predictions about a network's function or behavior

[1] This term can refer to either RNA intermediates or proteins, as both are considered products of a gene's expression.

[2] A microarray consists of thousands of chemical probes, each specific to a gene product, arranged on a silicon or glass substrate about the size of a coin. They allow one to measure the expression (i.e., the concentrations of gene products) of thousands of genes simultaneously in a single experiment.

under untested conditions. These approaches are called "supervised" because the model is learned from a training data consisting of a set of system responses (e.g., RNA concentrations or protein activities) to a set of known inputs (e.g., perturbations to the expression of specific genes).

Unsupervised learning approaches also have been widely applied to the analysis and inference of gene networks. In unsupervised learning, a structured training data set is not available. Normally the data set consists of the network's expression responses to some inputs as measured by microarrays, with the inputs being unknown. A variety of approaches have been applied to identify common patterns in the expression of genes, including clustering [Pilpel et al., 2001] and principle component analysis. The clustering of expression data can be used to identify groups of genes that are coordinately regulated, as it assumed that the expression responses of these genes will be highly correlated. To be successful at identifying network features, unsupervised approaches must be supplied with more than just the expression response data. One such example is the incorporation of the DNA sequences near expressed genes, or cis-regulatory sites, and the proteins known to bind to these sequences[Bussemaker et al., 2001].

The application of supervised learning to gene networks, on the other hand, is more recent and the remainder of our review will focus on such methods. There are three principle challenges in applying supervised learning to gene network inference: (1) the selection of the model structure, (2) the fitness metric and computational search scheme used to estimate parameters, and (3) the design of experiments. The most important of these is the selection of the model structure, because it influences choices for the other two challenges and ultimately determines the utility of the approach in practical applications.

It is often presumed that in order to understand cell function at a "system level", it is necessary to build expansive computational models that integrate much of the nature of the physical details of gene, protein and metabolite interactions in a cellular network. But such a goal is probably unrealistic both computationally and experimentally – cells are too complex. Somehow, the physical interactions must be translated into a simplified model that still preserves system properties of the network. The model our group has chosen uses a linear approximation of network interactions which limits the number of required experiments, but retains valuable information on network structure and behavior. A variety of other models of gene expression, each with advantages and disadvantages, have been considered, but few models have been rigorously evaluated against experimental data. Thus, it remains an open question which models are most appropriate for which situations.

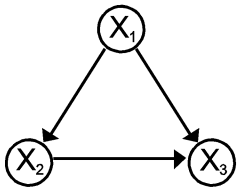


Figure 2. An idealized three-gene network, with regulatory interactions represented by edges.

Model Classes of Gene Networks

Here we discuss three types of model classes: Boolean functions, Bayesian networks, and systems of ordinary differential equations. Each model is illustrated using the simple gene network illustrated in Fig. 2. In each model, the states of the genes (e.g. the concentration or activity of gene products) are represented by the variables X_1, X_2, and X_3.

Boolean Models . In Boolean models, the state of each gene in the network as is represented as either "on" or "off", and is considered to be a (Boolean) function of the state of all other genes in the network. In our example network, might represent the behavior of gene 3 as a logical function of its two regulators, genes 1 and 2:

$$X_3 = X_1 \wedge X_2$$

If a Boolean model is assumed, inferring the network means identifying all of the logical functions which govern the behavior of each gene. Boolean models are valuable for applications where intermediate levels of gene expression are unimportant [Kauffman, 1989]. Inference methods based on Boolean models have also been developed and analyzed, [Ideker et al., 2000] but few have been experimentally validated.

Bayesian Network Models . Bayesian networks are used to model regulatory interactions between genes as probabilistic relationships. In such a network, the expression level each gene is represented as a continuous random variable. The probability density function (PDF) for that random variable is assumed to be jointly dependent upon the concentration of other genes in the network. In this framework, the task of reverse engineering the network is to identify two sets of parameters: the dependencies between variables (the edges in the network), and the nature of these dependencies. These parameters are typically learned from large

data sets, but occasionally some of these parameters may be supplied as prior information. In our example network of Fig.2, we would hope to discover a joint probability density function showing that X_2 is dependent on X_1, whereas X_3 depends on both X_1 and X_2. To aid the estimation of these relationships, the joint probability is broken into the product of conditional probabilities which are then estimated. In our three-gene network, this joint probability can be expressed as

$$P(X_1, X_2, X_3) = P(X_1)P(X_2|X_1)P(X_3|X_2, X_1)$$

Further simplifying assumptions, such as using discrete random variables with just two states (corresponding to a gene's being 'ON' or 'OFF' as in Boolean models), or limiting the number of edges in the network can simplify the inference task and reduce the requirements for experimental data.

Bayesian models are well-suited to dealing with incomplete data sets, and allow for the incorporation of prior data about a regulatory network's structure. The development of Bayesian inference methods for gene networks has received considerable attention in recent years, and has been applied successfully to experimental expression data to identify regulatory links in gene networks[Segal et al., 2001]. However, one shortcoming of Bayesian models it can be difficult to incorporate feedback, a property commonly found in gene networks.

Ordinary Differential Equation Models . Systems of ordinary differential equations present a natural and semi-physical model for regulatory gene networks. In this approach, the time-evolution of each gene's concentration is described as a function of all gene concentrations in the network. In this framework, inferring the network is a matter of determining these functions. In our simple network of Fig. 2, a recovered model might describe the behavior of X_2 and X_3 as:

$$\frac{dX_1}{dt} = K, \quad (K = constant) \tag{1}$$

$$\frac{dX_2}{dt} = f(X_1), \tag{2}$$

$$\frac{dX_3}{dt} = f(X_1, X_2). \tag{3}$$

These functions may be linear[van Someren et al., 2001, Liang et al., 1998] or non-linear[Weaver et al., 1999]. Gene networks have been observed to be sparse; that is, most genes are regulated by a small fraction of the total set of genes in the network. Thus, the functions for each gene have relatively few input variables. This property makes the challenge of identifying these functions, and recovering the network, more tractable.

A variety of approaches may be applied to infer these interaction functions. One common approach is to employ some form of multiple regression. In particular, we have developed a method based on a linear model of the network and have used it to correctly infer a model of regulation in *E. coli* controlling DNA damage response and repair[Gardner et al., 2003]. Moreover, we showed this model could correctly predict regulatory features and behaviors of the network. These results are presented briefly.

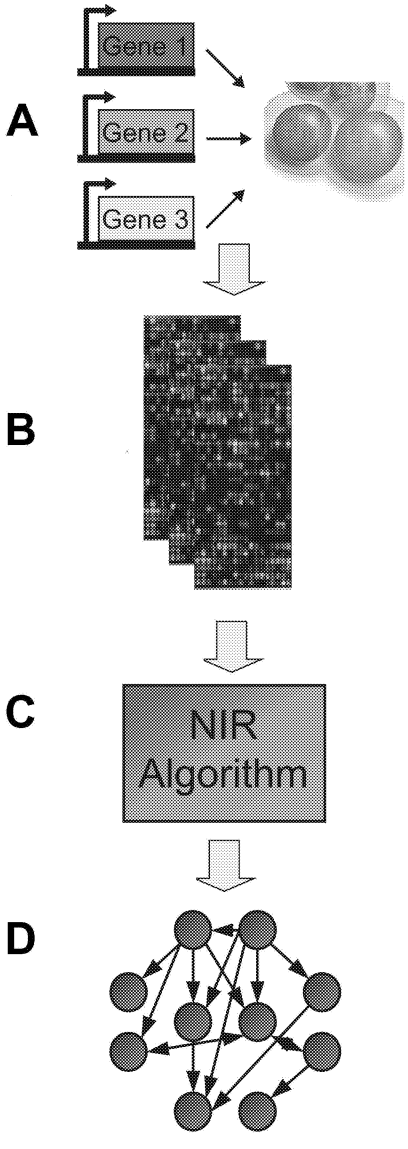


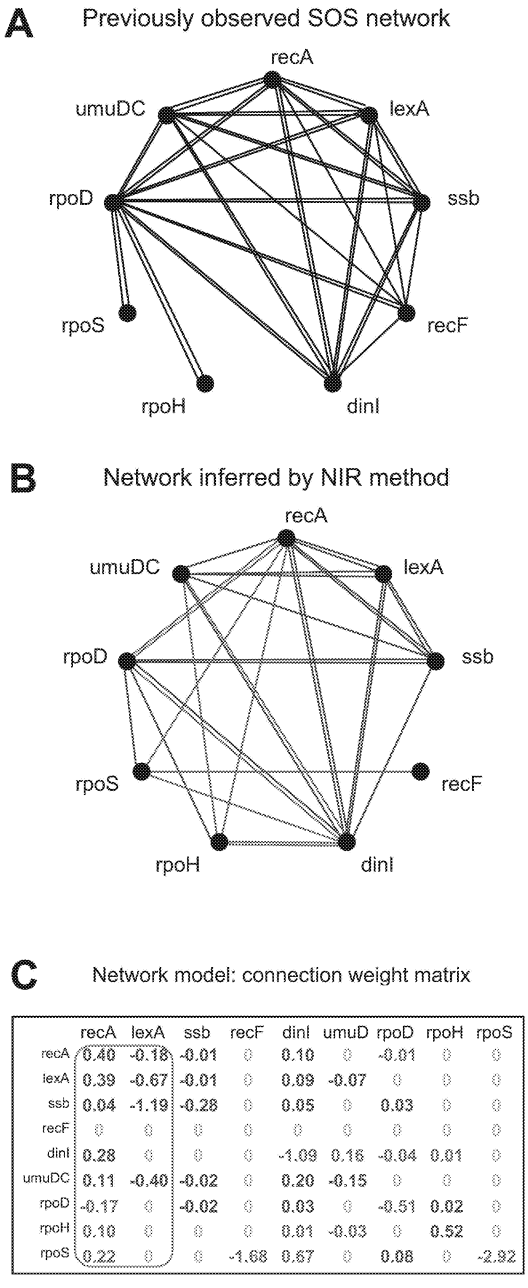
Figure 3. Overview of the NIR method. **A**: A structured set of perturbations is delivered to cells, such as the overexpression of one or more genes in each experiment. **B**: Gene expression (or, if possible, protein and metabolite activity) is measured for all genes in the network. **C**: This data set is analyzed by the NIR algorithm to infer a model of the perturbed network. **D**: The resulting model may then be used for analysis and prediction of network function.

NETWORK INFERENCE VIA MULTIPLE REGRESSION (NIR)

In our inference method, called Network Identification via multiple Regression (NIR) gene interactions are represented by a linear model. The rate of synthesis of RNA from each gene is represented as a weighted sum of the RNA concentrations of the other genes in the network. Training data are collected by overexpressing individual genes and then measuring the steady-state RNA levels of all genes in the network. The weights of the model are then learned using multiple regression analysis. Thus, the influence of each gene on each other gene (if any) is determined by the calculated weights.

We tested the NIR method on the SOS network in *E. coli*. This network regulates the cell's response to DNA damage and involves more than 100 genes. As the SOS network is well described in the literature, it serves as a good network for validating the NIR method. As a starting point, we applied the NIR method to a nine-gene subset at the core of the network. We used an extra copy of each gene to individually alter each gene's expression in nine separate experiments, and we measured the resulting changes in RNA concentrations. The NIR method was able to correctly identify 25 of the previously identified regulatory relationships between the nine genes, as well as 14 relationships that may be novel regulation pathways or possibly false positives (Fig.4). These results were obtained with a noise-to-signal ratio of approximately 68%. Moreover, the network model obtained by the NIR algorithm correctly identified the *recA* and the *lexA* genes, the known principal regulators of the SOS response, as having the strongest influence (largest regulatory weights) on the other genes in the network (Fig. 4C). Thus, the model can be used to suggest which genes should be perturbed to elicit a maximal response from the network – a capability of great value in optimizing bacteria for environmental remediation or bio-production of compounds.

The network model obtained by the NIR algorithm was also used to identify the genes that mediate the network response to a particular stimulus. The model was applied to public data obtained using microarrays to assay the response of *E. coli* to various stimuli. As illustrated in Fig. 5, the network model correctly identifies the *recA* gene as the key mediator of the SOS network response to UV irradiation and treatment with the quinolone antibiotic pefloxacin (both cause DNA damage), but not to novobiocin treatment (a quinolone that does not cause DNA damage).



	recA	lexA	ssb	recF	dinI	umuD	rpoD	rpoH	rpoS
recA	0.40	-0.18	-0.01	0	0.10	0	-0.01	0	0
lexA	0.39	-0.67	-0.01	0	0.09	-0.07	0	0	0
ssb	0.04	-1.19	-0.28	0	0.05	0	0.03	0	0
recF	0	0	0	0	0	0	0	0	0
dinI	0.28	0	0	0	-1.09	0.16	-0.04	0.01	0
umuDC	0.11	-0.40	-0.02	0	0.20	-0.15	0	0	0
rpoD	-0.17	0	-0.02	0	0.03	0	-0.51	0.02	0
rpoH	0.10	0	0	0	0.01	-0.03	0	0.52	0
rpoS	0.22	0	0	-1.68	0.67	0	0.08	0	-2.92

Figure 4. Inference of *E. coli* subnetwork using the NIR method. **A**: Previously known connections of the nine-gene subnetwork of the *E. coli* DNA damage response pathway. **B**: The connections identified by the NIR method. For visual clarity, strengths and directions of the identified connections are not labeled. **C**: The model is used to calculate the mean influence of each gene on expression changes in the other genes. The model identifies *recA* and *lexA* as the principal regulatory nodes in the network, which is consistent with existing knowledge.

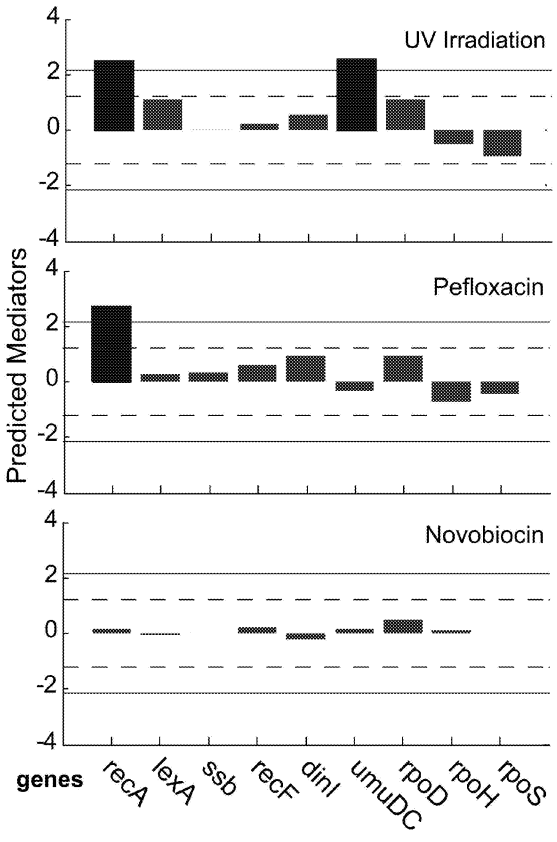


Figure 5. Prediction of genes mediating response to three different stimuli. The network model identified by NIR was used to predict the mediators of expression responses following UV irradiation and treatment with two antibiotics. The expression data were obtained from public microarray data sets. In the case of UV irradiation and pefloxacin treatment, both DNA-damaging, the *recA* gene is correctly predicted as the mediator of the expression response. For novobiocin, which does not damage DNA, *recA* is not predicted as the mediator of the expression response. Lines denote significance levels: $P = 0.3$ (dashed), $P = 0.1$ (solid).

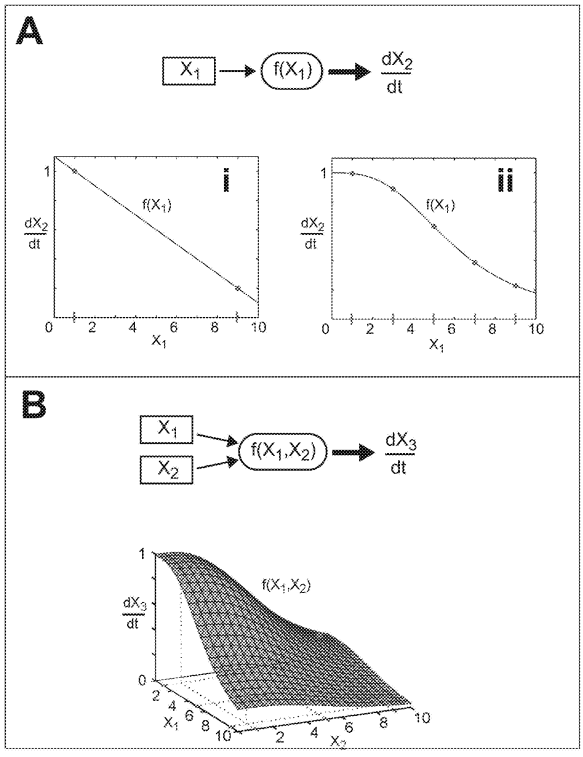


Figure 6. The quantitative relationship between the concentration of a regulator gene and the expression of a regulated gene can be represented as a function surface f, e.g., the line, curve, and surface in panels **A(i)**, **A(ii)** and **B**. A small number of experimental points can fully define simple surfaces such as the line in panel **A(i)**. (Data points in the figures represent experimental measurements of the input/output relationship, which are noise-free for illustration purposes.) But a larger number of experimental points are needed to fully define more complex input/output relationships such as the curve in panel **A(ii)**. In panel **A(ii)**, 5 points are adequate to define relationship for a single-gene input, but for two inputs (**B**), $5^2 = 25$ points are needed to sample the two-input surface as densely as the one-input surface.

SIMPLIFYING COMPLEXITY

The representation of a network or other system by a simplified model is sometimes called *smoothing*. In a sense, each interaction functions in the network can be represented as a surface (Fig. 2), and the details of the biochemistry are like bumps or wrinkles on the surface. Model simplifications ignore these bumps, but still capture the general shape and curvature of the surface. As the complexity (roughness) of a model representation increases (and hence its ability to describe the details), so does the amount of data needed to describe it. Thus, there is a trade-off between model scope and realism (complexity) and experimental/computational tractability. This trade-off is exaggerated in multivariate systems (which are the norm in biology). The amount of data needed increases exponentially with the number of dimensions in the system. This problem is sometimes referred to as the *curse of dimensionality* and is illustrated in Fig. 2. To make the inference of high-dimensional systems tractable, we must use simplified models of input/output relationships, such as a hyperplane or boolean function. However, such approximations may limit the range of questions addressable by the model.

The choice of model type and simplicity depends on several factors including the nature of the system being studied, the properties that are desired to be studied, and the type and amount of data that can be collected. This choice is the major challenge in applying statistical learning to gene networks.

In developing our NIR method, for example, we selected a model able to capture network properties of value in biomedical applications, namely the identification of the major network regulators and mediators of chemical and environmental stresses. On the other hand, this model may

not be appropriate for all behaviors observed in microorganisms such as the genetic interactions that orchestrate bacteriophage infection. Other models types of varying complexity may be better suited to such systems, including the Boolean and Bayesian models discussed here. We and other groups are currently exploring the feasibility and utility of such approaches.

DRUGS, SLUDGE, AND BIO-BATTERIES: APPLICATIONS OF GENE NETWORK INFERENCE

The elucidation of regulatory pathways in microbes has potential applications for bioremediation, the development and improvement of antibiotics, and renewable energy resources.

Since 1972, when a strain of *B.cepacia* capable of degrading petroleum in oil spills was engineered by Chakrabarty [1972], the environmental uses of microbes have been explored. Bacteria have extraordinarily flexible respiratory mechanisms, possessing multiple and overlapping pathways for obtaining energy from compounds in their environment. These pathways govern the movement of electrons from high redox-potential molecules, such as sugars, to lower-potential molecules such as fumarate, oxygen, or metallic ions. In particular, *S. oneidensis*'s ability to reduce soluble heavy metals, such as Uranium(VI), to an insoluble form has made it a leading candidate for a role in the bioremediation of contaminated waste sitesLovley et al. [1991].

The versatility of *S. oneidensis* respiratory pathways is conferred, in part, by extensive gene networks regulating the concentrations of enzymes and other metabolic proteins. Examination of anaerobic growth conditions has shown manyfold increases in the rates of metal reduction and heightened expression of a number of electron transport genes, compared with aerobic conditions [Beliaev et al., 2002]. The pathways which *S. oneidensis* activates vary depending on the conditions under which it is grown; in the presence of oxygen, for example, *S. oneidensis* will not reduce Uranium(VI). The NIR method can be used to develop a model of the gene networks regulating these electron transport pathways, in particular those underlying (i) the transition from aerobic to anaerobic growth, and (ii) growth on various substrates, including environmental toxins. In combination with models of bacterial metabolism, such a model could help to manipulate and optimize *S. oneidensis* for remediation use in the field.

Network identification also has immediate utility in the development of improved antibiotics. Bacterial communities can exist on surfaces in the form of a biofilm, where cells are embedded in a matrix of dense, extracellular polysaccharides. Biofilms, such as those involved in gingivitis, cystic fibrosis, and surgical implants, represent as many as 60% of human infections and are notoriously difficult to eradicate with antibiotic therapy. Cells in biofilms survive antibiotic doses several orders of magnitude higher than those sufficient to kill free-floating bacteria. This increased antibiotic resistance is not necessarily the result of genetic mutations. Cells removed from a biofilm are as susceptible to killing by antibiotics as their free-floating counterparts [Spoering and Lewis, 2001]. In some cases, this antibiotic tolerance is due to reduced diffusion of drugs into biofilms, but this is not true in general. For example, fluoroquinolone antibiotics, such as ofloxacin and ciprofloxacin, readily penetrate biofilms and kill most of the cells. Yet a small number of cells survive regardless of the concentration of antibiotic applied[Keren et al., 2004]. These cells are called persisters. Persisters are believed to repopulate a biofilm after discontinuation of antibiotic treatment, and thus cause recurrent infections. In addition, persistence in microorganisms may amplify the problem of genetically mediated resistance by giving cells the opportunity to develop advantageous mutations.

The mechanisms of persistence are not well understood. But it is likely a dynamic response of the cell, orchestrated by multiple stress response pathways. A better understanding of these stress response networks will be of great value in identifying productive targets for novel anti-biofilm compounds with greater lethality and lower rate of resistance. As described above, the network model obtained with the NIR method can also be used to identify the genes that mediate the effects of a particular compound. Thus the network model will also be of great value in the optimization of candidate antibiotic compounds, and will enable the development of novel classes of drugs that account for and utilize the complex regulatory properties of gene networks.

The rich array of metabolic processes which operate within microbes have several applications in renewable energy. Two strategies being pursued at present involve using microbes to generate an electric current directly and using microbes to produce pure hydrogen gas (which is then harvested as an energy source). In both cases, the initial energy sources are readily-available carbohydrates, such as glucose or cellulose.

The generation of a direct electric current with the recently isolated microbe *R. ferrireducens* involves some of the very same respiratory pathways used in bioremediation. As its name implies, this bacterium has the capacity to transfer electrons onto Fe(III), reducing it to Fe(II) as part of its

breakdown of glucose. Remarkably, it has been discovered to be able to transfer electrons directly onto an electrode. In a simple fuel-cell, comprised of anaerobic two-chambered vessel with the microbe, glucose medium, and anode in one chamber separated by a cation-selective membrane from the cathode chamber, over 80% of the theoretic yield of electrons were transferred into current [Chaudhuri and Lovley, 2003]. A better understanding of microbial respiratory networks may allow for this same ability of *R. ferrireducens* to be engineered into other, more well-studied microbes.

The bacterial production of hydrogen gas is catalyzed by enzymes known as hydrogenases, which couple the production of hydrogen gas to the reduction of metabolites (usually $NADH$) generated by the catabolic pathways of sugars. The theoretical stochiometric yield for these reactions would be 12 mols of H_2 for every mol of glucose, and in yields approaching this have been achieved experimentally [Woodward et al., 2000]. However, a number of practical obstacles remain before this technology could be commercially applied, such as finding enzymes which are stable at higher temperatures (thereby improving the kinetics of the reactions).

It is worth noting that CO_2 is a byproduct of these respiratory pathways, and hence might be thought to contribute to global warming as existing energy technologies do. However, microbes can reduce all of this CO_2 into sugars via photosynthetic pathways. Using microbes as vessels for carbon sequestration on a large-scale, to reset the balance of atmospheric gases, is itself a bioremediation application under investigation.

FUTURE WORK

Although a great deal of effort has been focused on the development of various modeling and inference approaches, experimental evaluation of such schemes has been relatively limited to date. Rigorous testing and refinement of these approaches is needed to better determine where, when and how to apply them.

When inference methods have been tested, they have generally been applied only to RNA concentration data, but such methods could just as easily be extended to proteins and metabolites. However, large-scale measurements of protein concentrations, protein activity states, and metabolite concentrations are still difficult to obtain. Though still young, a variety of technologies including mass-spectrometry, high-resolution electrophoresis, and protein arrays are showing increasing promise. As these technologies become better developed, it will be possible to use inference algorithms to explore the dynamic and quantitative properties of protein signaling cascades and metabolic networks. This capability will be of tremendous value in understanding the mechanisms by which such networks mediate, distinguish and integrate environmental signals in microbes and higher organisms. For the present time, we expect that inference methods will continue to prove valuable in analyzing and predicting the behavior of gene regulatory networks.

REFERENCES

G. W. Beadle and E.L. Tatum. Genetic Control of Biochemical Reactions in Neurospora. *Proc Natl Acad Sci U S A*, 27:499–506, 1941.

A. S. Beliaev, D. K. Thompson, T. Khare, H. Lim, C. C. Brandt, Guangshan. Li, Alison. E. Murray, John. F. Heidelberg, Carol. S. Giometti, John. 3rd. Yates, K. H. Nealson, J. M. Tiedje, and J. Zhoui. Gene and protein expression profiles of Shewanella oneidensis during anaerobic growth with different electron acceptors. *OMICS*, 6(1):39–60, 2002.

H. J. Bussemaker, H. Li, and E. D. Siggia. Regulatory element detection using correlation with expression. *Nat Genet*, 27(2):167–171, Feb 2001.

A.M. Chakrabarty. Genetic Basis of Biodegradation of Salicylate in Pseudomonas. *J Bacteriol*, 112:815–823, 1972.

Swades. K. Chaudhuri and Derek. R. Lovley. Electricity generation by direct oxidation of glucose in mediatorless microbial fuel cells. *Nat Biotechnol*, 21(10):1229–1232, Oct 2003.

T. S. Gardner, D. di Bernardo., D. Lorenz, and J. J. Collins. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science*, 301(5629):102–105, Jul 2003.

T. E. Ideker, V. Thorsson, and R. M. Karp. Discovery of regulatory interactions through perturbation: inference and experimental design. *Pac Symp Biocomput*, pages 305–316, 2000.

S. A. Kauffman. *Origins of Order: Self-Organization and Selection in Evolution.* Oxford University Press, Oxford, 1989.

I. Keren, N. Kaldalu, A. Spoering, Y. Wang, and K. Lewis. Persister cells and tolerance to antimicrobials. *FEMS Microbiol Lett*, 230(1): 13–18, Jan 2004.

S. Liang, S. Fuhrman, and R. Somogyi. REVEAL, a general reverse engienering algorithm for inference of genetic network architectures. *Proc Pacific Symp Biocomp*, 3:18–29, 1998.

D. R. Lovley, E. J. P. Phillips, Y. A. Gorby, and E. R. Landa. Microbial reduction of uranium. *Nature*, 350:413–416, 1991.

Y. Pilpel, P. Sudarsanam, and G. M. Church. Identifying regulatory networks by combinatorial analysis of promoter elements. *Nat Genet*, 29(2):153–159, Oct 2001.

E. Segal, B. Taskar, A. Gasch, N. Friedman, and D. Koller. Rich probabilistic models for gene expression. *Bioinformatics*, 17 Suppl 1:243–252, 2001.

A. L. Spoering and K. Lewis. Biofilms and planktonic cells of Pseudomonas aeruginosa have similar resistance to killing by antimicrobials. *J Bacteriol*, 183(23):6746–6751, Dec 2001.

E. P. van Someren, L. F. A. Wessels, and E. Reinders, M. J. T. andBacker. Searching for limited connectivity in genetic network models. *Proc 2nd Int Conf on Systems Biol,*, pages 222–230, Nov 2001.

D. C. Weaver, C. T. Workman, and G. D. Stormo. Modeling regulatory networks with weight matrices. *Proc Pacific Symp Biocomp*, 4:112–123, 1999.

J. Woodward, M. Orr, K. Cordray, and E. Greenbaum. Enzymatic production of biohydrogen. *Nature*, 405(6790):1014–1015, Jun 2000.