BIOGRAPHICAL SKETCH

Provide the following information for the key personnel in the order listed on Form Page 2.
Photocopy this page or follow this format for each person.

NAME	POSITION TITLE
Adam P. Arkin, Ph.D.	Assistant Professor

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, and include postdoctoral training.)*

INSTITUTION AND LOCATION	DEGREE	YEAR(s)	FIELD OF STUDY
Carleton College, MN	B.A.	1988	Chemistry
Massachusetts Institute of Technology, MA	Ph.D.	1992	Physical Chemistry
Stanford University (Chemistry), CA	Postdoc	1992-95	Nonlinear Chem. Systems
Stanford University (Developmental Biology), CA	Postdoc	1995-1997	Modeling Development

Professional Experience

July 1999- Present Assistant Professor, Departments of Bioengineering and Chemistry, University of California, Berkeley

Faculty Scientist, Computational and Theoretical Biology Department, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA

January 1998-July 1999 Staff Scientist, Computational and Theoretical Biology Department, Physical Biosciences Division, E.O. Lawrence Berkeley National Laboratory, University of California, Berkeley, CA

Address

1 Cyclotron Road, MS 3-144, Berkeley, California, 94720

Teaching

Chemistry 130A: Biophysical Chemistry

Selected Publications

Refereed Journals

1. **Arkin, A.P.**, Youvan D.C. (1992) An Algorithm for Protein Engineering: Simulations of Recursive Ensemble Mutagenesis. *Proc. Natl. Acad. Sci. USA* **89**(16):7811-7815.
2. **Arkin, A.P.**, Ross, J. (1994) Computational Functions in Biochemical Reaction Networks. *Biophysical Journal*. **67**:560-578.
3. **Arkin, A.P.**, Ross, J. (1995) Statistical Construction of Chemical Reaction Mechanisms from Measured Time-Series. *J. Phys. Chem.* **99**: 970-979.
4. McAdams, H., **Arkin. A.P.** (1997) Stochastic Mechanisms in Gene Expression. *Proc. Natl. Acad. Sci., USA* . **94**(3):814.
5. Swanson, C., **Arkin, A.P.**, Ross, J. (1997) An Endogenous Calcium Oscillator May Control Early Embryonic Division. *Proc. Natl. Acad. Sci., USA* . **94**(4):1194.

6. **Arkin, A.P** ,.Shen, P.-D., Ross, J. (1997) A Test Case of Correlation Metric Construction of a Reaction Pathways from Measurements. *Science*. **277**(5330): 1275.
7. McAdams, H. H., **Arkin, A.P.** (1998) Simulation of Prokaryotic Genetic Networks. *Annu. Rev. Biophys. Biomol. Struct.* **27**: 199-244
8. **Arkin, A.P.**, Ross, J., McAdams, H.H. (1998) Stochastic Kinetic Analysis of a Developmental Pathway Bifurcation in Phage-λ *Escherichia coli.Genetics.* **149**(4):1633-1648.
9. McAdams, H.H., **Arkin, A.P** (1999) Genetic Regulation at the Nanomolar Scale: It's a Noisy Business! *TIGS.* **15**(2): 65-69.
10. **Arkin, A.P.** (1999) Signal Processing by Biochemical Reaction Networks. In: Biodynamics. J. Walleczek, ed. Cambridge University Press, Cambridge.In Press.

Other Significant Publications

1. with Gary Stix. (1991) Protein Probe: Remote Sensing Technique Screens Bacterial Colonies. *Scientific American*. May issue. p. 123.
2. **Arkin, A.P.**, Youvan, D.C. (1992) Digital Imaging Spectroscopy. In: The Photosynthetic Reaction Center J. Deisenhofer & J.R. Norris eds. 133-154.

Current and Pending Grants

N00014-99-F-0458 (A.Arkin) 9/1/99-12/31/01 10%
ONR/DARPA $65,055,$251,069,~$150,000
Instant Cell Analysis, BioSpice, Cellular Device and Exquisite Detection: Towards and Interactive Biology

A multisite program is proposed to bring biological research to a level of real-time interactivity similar to computers. The set of subtasks to be covered include: (1) non-enzymatic and rapid methods of DNA sequencing and protein analysis; (2) Modeling genetic regulatory networks (BioSpice) and demonstration of models through cell engineering; (3) Creating cellular devices from genomic data and (4) exquisite detection (detection of 1 molecule in 10^20).

LDRD (Adam Arkin, PI) 10/1/99-9/30/01 10%
DOE $150,000
Integrated Computational Biology of Stress Response in *D. radiodurans* and *B. subtilis*.
We propose to build a coherent computational biology program at LBNL by linking research in DNA modeling; protein fold recognition, comparative modeling, and *ab initio* prediction of individual gene products; and molecular recognition of protein-protein and protein-nucleic acid complexes; and modeling biochemical and regulatory pathways, using *Deinococcus radiodurans* and *Bacillus subtilis* as our test beds.

LDRD (Adam Arkin, PI) 10/1/99-9/30/01 10%
DOE $200,000
Integration of multiscale bioinformatic tools from sequence and structure to networks.
All three grants show some overlap with the current project. See section 4.E for a description of the overlap and the synergy amongst these projects.

4. Research Plan

Abstract

This proposal outlines an integrated computational and experimental program to elucidate the molecular basis for type-1 pili phase variation in uropathic *E. coli* and the sporulation initiation program of *Bacillus subtilis*. Both these pathways represent genetic switches that respond to environmental signals in order to change the behavior of a population of bacteria. In addition, both these switches are "imperfect" in that the population splinters into heterogeneous subpopulations that have different behaviors. Both pathways are also involved in the pathogenic potential of the organism or a closely related microbe. Type-1 pili mediate adherence of uropathic *E. coli* to the urinary track epithelium, and the sporulation pathway in *Bacillus subtilis* is nearly identical to that in *B. cereus B. anthracis* both of which are human and animal pathogens. The biochemical and genetic networks underlying these switches are of sufficient complexity that it is nearly impossible to reason about their function qualitatively. Thus, we propose to further development an integrated database and data-mining tool called Bio/Spice specialized for the kinetic simulation and analysis of biochemical and genetic reaction networks. We have found such a tool necessary for systematizing, consistency checking and hypothesis testing experimentally derived information about any biological process of even moderate complexity (without the tool an experimentally validated model of the λ-phage lysis lysogeny decision suffient for prediction took over two years). The combined research program will deliver a useful new bioinformatics tool (Bio/Spice) to the biomedical community and will yield insight into two industrially and medically important pathways. Once these pathway models are validated against experimental data (from our own laboratory and our collaborators), we will explore the means of pharmaceutical control of these processes (e.g. to force 100% sporulation of a *B. subtilis* population) and test how well such interventions fare when parameters change due to mutation.

A. Specific Aims

The biochemical and genetic pathways that underlie a microbe's response to a particular antibiotic and its ability to evade and evolve away from that antibiotic are complex, nonlinear networks of chemical reactions. Further, since these pathways are often regulated by small numbers of molecules, the behavior of a particular cell is often a stochastic process, and a population of microbes can be extremely heterogeneous. Under these conditions it is difficult to predict rationally how a set of pharmaceutical challenges to bacterial population will affect that population's long-term survival. This inability both betrays a lack of fundamental understanding of cellular function and a lack of enough quantitative measurements of the relevant cellular processes to make a good model. The central goal of our laboratory and this proposal is to develop a data-driven theoretical and computational framework onto which genomic, biochemical and molecular profile (e.g. gene microarray) data can be hung in order to explore dynamic pathway function and response to pharmaceutical challenge. The value of this approach is two fold. The first is the systematization of the vast amount of data on genomes, pathways, and temporal patterns of cellular concentrations and localizations; a task made necessary by the sheer volume and complexity of the data now being generated. Even the process of checking for consistency among the different measurement types and replications of those measurements is almost impossible without such a framework, let alone asking complex question of the data such as which protein is likely to control which gene. The second value is the ability to check whether the standing knowledge on a particular pathway is sufficient to explain the observed data or whether additional knowledge is necessary. If additional pieces are needed, then the framework provides a basis on which to build and test hypotheses of cellular function. Collectively, we call the suite of data systematization, analysis and simulation tools, Bio/Spice. Our specific aims involve the improvement and completion of the current Bio/Spice tool, application of this tool to a particular set of model microbial systems of medical and industrial import and development of new experimental techniques to rapidly measure the concentrations of key proteins and metabolites:

1) We will produce a biologist-friendly biological network analysis toolkit called Bio/Spice geared toward organizing the existing data about an organism into a form suitable for functional analysis. We will populate the database part of this tool initially, with data from *Escherichia coli* (different pathogenic and non-pathogenic strains) and *Bacillus subtilis*.
2) We will use Bio/Spice to produce experimentally validated models of the control and temperature sensitivity of type-1 pili phase variation in uropathic *E. coli* and the sporulation initiation/competence pathways in *B. subtilis*.

B. Background and Significance

Specific Aim 1: The goal of any genome project, beyond simply producing the list of putative genes that compose the genome, is to produce a basis from which the mechanisms of development, disease, environmental sensing, biosynthesis and control of metabolism can be deduced. In order to achieve this goal at least four fundamental tasks must be completed: 1) the identification of gene and genetic regulatory elements coded for in the genomic DNA; 2) the determination of the encoded protein and RNA functions; 3) deduction of the interactions amongst the enzyme, transcription factors, structural proteins, RNA, DNA and other cellular components that implement the mechanisms above, and 4) determination of the temporal pattern of activity, i.e., the kinetics, the resultant biochemical networks.

The progress of many genome projects as well as many decades of conventional genetics and biochemistry has admirably succeeded in compiling the parts lists and partial wiring diagrams of several organisms. In some special cases there is a great deal of physical data on interaction strengths, chemical mechanisms and kinetics. The challenge is now to combine all these results in such a way that it is possible to predict and control cellular function from these data. However, biological regulatory networks are highly nonlinear, execute asynchronously, under a wide range of environmental conditions and, with a significant level of internally generated noise due to stochastic fluctuations in signal protein concentrations and rates of gene expression. The conventional intuitive approach to predicting network behavior is simply inadequate for these networks that involve many genes, nonlinear feedback interactions, and stochastic elements. The problem is analogous to that faced by engineers in understanding how a complex computer chip works given a list of parts and a partial wiring diagram. Even with full knowledge of the elementary device physics for the circuit components and even given complete wiring diagrams, engineers still rely on a sophisticated suite of analytical tools, circuit simulators and synthesizers, and electronic probes in order to design, diagnose and understand complex circuits. Biologists, on the other hand, have few such tools despite the fact that the 'circuits' they study are far more complex and less characterized than their electronic analogs.

The goal of this laboratory is to create a theoretical and computational framework for the systematization and integration of genetic and biochemical data into validatable dynamic models of cellular function. The suite of database, analysis and simulation tools are to be packaged into an integrated framework analogous to the SPICE tools used by electrical engineers. Thus, we call this tool set Bio/Spice (see below). Design of the Bio/Spice tool involves a number of sub-tasks:

1) Design and population of heterogeneous molecular biological databases that integrate data generated as a result of the genome projects, structural analyses and advances in multivariate measurement of gene expression, protein concentration/localization, and small molecule concentrations.
2) Development of network "reverse engineering" tools that produce testable hypotheses about unknown genetic/biochemical network structure and kinetics from analyses of perturbation/time-series of concentrations and predictions from other bioinformatic tools such as sequence and structure homology programs, upstream-sequence alignment tools for prediction of regulatory sites in co-expressed genes, etc.
3) Development of robust mechanistic models for cellular processes including gene expression, enzymatic regulation of scaffolded multiprotein complexes, molecular diffusion and cell growth. Models must be formulated at multiple levels of abstraction ranging from detailed physical models of stochastic molecular interaction to deterministic kinetic formalisms to (if absolutely necessary) Boolean models of gene expression so that analysis can be accomplished in systems for which there are varying levels of knowledge for various of the subsystems.
4) Development of regulatory motif searching algorithms for the identification and analysis of regulatory motifs that recur across and within organisms. (A futile cycle is an example of such a motif).
5) Development of static (steady-state, bifurcation) and dynamic (simulation) analyses of cellular function based on models, mechanisms and parameters from above. This involves sophisticated numerical technologies for multi-resolution and multiscale

analysis that are proving necessary not only in biology but also mechanical engineering and electronics.

The laboratory has made progress and published in each of these areas.

Specific Aim 2: In parallel with this effort we are using the developed technology to study signalling and developmental pathways in various prokaryotic systems. The two systems under particular study are the control of type-1 pili phase variation in urinary tract invasion by uropathic *E. coli* and the sporulation initiation/competence/motility pathways in *Bacillus subtilis.* Each of these systems is of a different level of complexity, state of knowledge and involves different mechanisms of control and thus provide excellent focusing problems for the design and evaluation of the Bio/Spice suite.

Type-1 fimbrial phase variation in *E. coli*

Random phase variation is a process in which individual cells in a population alternate between a piliated state, in which protein structures (pili) are built densely on the outside of the cell, and a non-piliated state [1-4]. This process in *E. coli* presents an excellent and relatively simple model of a random genetic decision that increases the pathogenic potential of a population of infectious bacteria. Pili, or fimbriae, expressed on the surface of the *E. coli* recognize and adhere to target mannose-containing receptors in, for example, human buccal cells, proximal tubular cells of the kidney, epithelial cells in the bladder, lung, intestine, and various inflammatory cells. It is thought that phase variation *in vivo* allows infections to spread more efficiently by allowing adherent bacteria to desorb and migrate (and perhaps chemotax) to other sites. In addition, bacterial pili are also known to attach to phagocytic cells in a mannose-dependent manner leading to lectinophagocytosis. However, other studies imply that pili may also aid in protecting *E. coli* from phagocytic killing even while promoting binding to those cells. Thus, presence of pili in the absence of phagocytes is advantageous in that it promotes adherence whereas, at sites at which phagocytes are present, pili would be a distinct disadvantage. Though a strong clinical correlation has been found, for example, between type-1 pili and the potential to cause cystitis and urethritis, no clear mechanism for this potentiation is given.

In *E. coli* an eight gene cluster (*fimA-H*) is involved in the synthesis, assembly and regulation of type 1 pili. Phase variation results from the inversion of a 314-bp DNA fragment containing the promoter for *fimA*, the structural gene for the major subunit of the pilus as well as downstream structural genes. Inversion of this region turns off production of *fimA* allowing cells to switch between piliated and non-piliated states. Inversion is mediated by two other *fim* genes, *fimB* and *fimE*. The *fimB* gene effects inversion of the segment in both directions with nearly equal propensity and *fimE* sets the promoter in the off position. The ability of the bacterium to phase-vary has been shown to depend on cellular growth conditions (such as viscosity, osmolarity, and nutrient availability) and, in some clinical isolates, contact interactions (with agar and other cells). One or more of these conditions could trigger regulatory signals that alter the expression or activity of *fimB* and *fimE*. In broth culture conditions, however, variation of these signals at each bacterial cell in a population can be minimized. Thus, random phase-variation inherent in the unicellular biochemistry and genetics can be studied both theoretically and experimentally while leaving open the ability to expand the study to incorporate signal-transduction and external control of genetic states, cell-cell and cell-substrate interactions.

A good fraction of the control circuitry for this phase-inversion is know and yet some very basic questions go unanswered. Why is the design of the switch such that both the fraction of piliated cells in a population and the individual switching rates of a given cell can be controlled (through modulation of Fim E and Fim B)? What is the role of switch based control of FimE expression? How and why is the maximum rate of switching achieved at body temperature? What is the role of population heterogeneity in evading immune response and promoting spread and organ colonization of these uropathic bacteria?

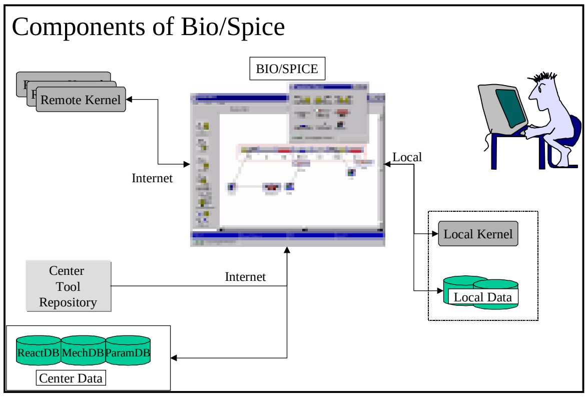
Sporulation Initiation in *Bacillus Subtilis*

The prokaryote *Bacillus subtilis* has provided a model example of unicellular differentiation, which because of the excellent genetics available has provided an excellent system to understand the basic mechanisms of regulating developmental gene expression. This industrially important gram positive bacterium, is part of a close clad of bacteria including *Bacillus cereus*, also used in plant biocontrol and a food poisoning agent, and *Bacillus anthracis*, or anthrax, the a deadly pathogen. As a response to nutrient exhaustion much of a population of *B. subtilis* enters a program of irreversible differentiation that results in the formation of a dormant life form termed the endospore or spore [5, 6]. This process lasts some 8 hours and uses approximately 125 developmental genes [7]. A number of well defined physiological and morphological changes occur during this process but the landmark event is the formation of two compartments within the developing cell or sporangium, termed the forespore and mother cell. The forespore is the germ line cell destined to become the mature spore but is made first by a process of membrane invagination which encases one of the two chromosomes present in the developing cell in two layers of phospholipid membrane. In both spore chambers separate programs of temporal gene expression occur, each driven by RNA polymerase bound to one of a number of alternate sigma factors [5, 8, 9]. These transcription factors appear at different times during development and, by recognizing different promoters, ensure expression of a unique regulon of genes during development. Four of these sigma factors are active only in the forespore or mother cell chambers providing spatial control of gene expression as well as temporal control. In addition, a number of DNA-binding proteins can further modulate gene expression.

Other responses, most of which are regulated by the initiation of sporulation, include the synthesis of several extracellular enzymes and secondary metabolites, the development of competence, motility etc. Because of this regulatory program the cell population diversifies into a number of subpopulation with different activities. In fact, in a wild type strain, only 10 to 30% of the cells will complete the program that produces a heat-resistant spore. The reminder of the population will pursue other activities that have not been characterized. All together over 140 gene products have been identified that are involved with the decision to execute or ramp up these various processes. Although sporulation has been studied for almost 40 years, the development of these populations is of recent finding and it is poorly understood. One of the problems is that, due to the complexity of the system, some of the tools necessary to address this problem, i.e. total RNA analysis and cell sorters, have become available only recently. The completion of the *B. subtilis* genome project (http://pbil.univ-lyon1.fr/nrsub/nrsub.html), a good deal of annotation, the initiation of a proteome project (http://microbio2.biologie.uni-greifswald.de:8880/) and a wide array of genetic and biochemical studies has yielded and extraordinary amount of direct and suggestive information of how these chemical processes are accomplished in the cell.

However, the exact control of the sporulation/competence switches and the role of competence in cell nutrition and DNA repair has not yet been identified. This is partly because: 1) the molecular parts have not all been fully identified, 2) the modes of regulation of these parts, transcriptionally, translationally and post-translationally, have not been completely elucidated, and 3) the sheer number and complexity of the individual chemical interactions preclude a qualitative description for how this system functions. In order to facilitate the understanding of this model system we propose to begin a quantitative analysis of these pathways. The direct questions we are asking are: 1) Where exactly is the sporulation initiation switch? That is, what components are absolutely necessary in order to definitively kick off the SpoO II cascade. Where is the bifurcation in control that allows part of the population to avoid sporulation? How can we force 100% decisions either towards or away from sporulation using targeted drugs? What protocol is most likely to disallow *B. subtilis* to mutate away from this attack? Results from this project will apply also to *B. cereus* and *B. anthracis* since the sporulation programs are highly conserved.

C. Methods and Procedures

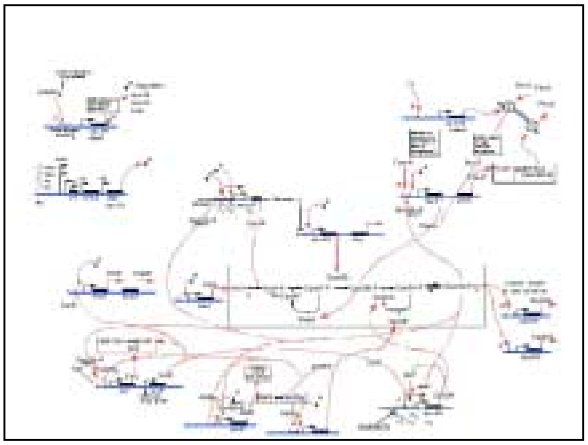


Bio/Spice: The Bio/Spice project is an ongoing project in this laboratory. However, we have recently increased the scope of the package and have begun a much more intricate and involved database design. Bio/Spice consists of four integrated packages: A heterogeneous biological database, molecular profiling data analysis tools and network deduction software, a multiresolution/multiscale biochemical/genetic network cellular simulation kernel and a biologist friendly graphical user interface (GUI). The database contains sequence, structure, pathway, kinetic parameter and mechanism, microarray and other molecular profiling data connected by a large set of relations that allow complex queries among all data types. This package is under intense development using the freeware database MySQL, a set of specially design middleware query engines, and C/C++ and Java interfaces. Three separate databases have been populated both by hand and automatically through the use of specially designed web-spiders: 1) **ReactDb**,. a database of biochemical/Genetic reactions has been populated by web spiders written for automatic consolidation and standardization of data from EcoCyc (Pangea Systems), Kegg (Kyoto University) and WIT (Argonne National Labs) and by manual data entry for the three prokaryotic networks described above. 2) **MechDb**. a database of well-known and novel biochemical mechanisms including arbitrary polynomial and rational reaction mechanisms, more than 60 parameterizable enzyme mechanisms, three models of prokaryotic cell growth dynamics, three models of prokaryotic transcription initiation, six models of elongation control, two models of mRNA translation/degradation and one model of diffusion and compartmentation, and 3) **ParamDB,** a database of particular kinetic parameters that currently include kinetic parameters for λ-phage lysis/lysogeny, *E. coli* metabolism and phase variation, and *B. subtilis* sporulation. One important future focus will be to augment MechDb with better models of eukaryote-specific mechanisms. These databases have been linked to a Java-based, biologist-friendly, model building/simulation notebook. 2) The notebook can interact with remote and local simulation and analysis kernels distributed on different computers. The notebook organizes the different tasks describes above. Optimization of the kernel for the various tasks is a collaborative project with scientists at Stanford and Naval Air Warfare Centers Weapons Division at China Lake. :The reverse engineering algorithms are being tested and improved in collaboration with scientists at Stanford and UC Davis. Publications on these topics and related work are [10-17]. There are major needs for development in the heterogeneous biological database and the specialized queries that are necessary for creation and validation of dynamic simulations from data. In addition, a great deal of work is necessary to create the hypothesis generation and validation tools as well as creation of a biologist friendly interface to these tools. The goal is to produce an industry-grade product for free use of the academic and medical communities.

Genetic Switch Analysis: Both the *E. coli* type-1 pili phase variation and the *Bacillus subtilis* sporulation initiation pathway are examples of genetic switches that drive "development" decisions. Genetic switches are of particular interest since they are major control points in the cell and thus are likely target for pharmaceutical intervention assuming the switch is importance in virulence. Also, the switching architectures are often shared across wide classes of organisms though implemented by different set of molecules in each. We are already along in a pilot study of the type-1 pili switching dynamics. A molecular level kinetic model including the effects of FimE, FimB, H-NS , IHF and LRP is under initial testing. We are nearing completion of the initial model for stochastic type-1 pili phase variation that explains a number of seemingly contradictory observations as well as the H-NS mediated temperature sensitivity of the circuit. The mechanism for achieving robust peak switching at 37 C is a simple but powerful architecture that we think may also be found in other "environmental" sensing circuits such as the osmoregulation network. The design of models for bacterial population/host interactions is now underway. In order to proceed further we need to measure the concentration of these various factors under a number of conditions (particularly temperature) and measure switching rates as well. We have designed but not yet implemented a number of deletion strains, and integratable vector constructs that allow us control over these various components of the circuit (e.g. constructs with inducible FimE in its appropriate upstream position from the FimA invertible locus driving green fluorescent protein). We propose to use these strains to get detailed kinetics of the switch circuit under well-controlled conditions. From these a detailed, experimentally validated model of switch function may be produced. This model will then form a basis for the investigation of the role of type-1 pili in the dynamics of tissue colonization and transmission.

We are in the data-compilation phase of the *B. subtilis* (part of the data is shown to the bellow). The main questions here are where is the essential switch that tells the organism to sporulate under population/nutritive stress? Why does only a fraction of the population undergo competence, sporulation and motility? How the population partitioning controlled?

Answering these questions will require efforts at all levels of biological data analysis. Genome sequence analysis will need to be used in order to identify further members of the operons expressed during these various processes. For example, operons have been identified that coordinately express proteins necessary for both competence and DNA repair thus implying coupling between these processes [18]. Also identification of upstream regulatory sequences and sigma factor binding sites will prove invaluable. Further, the identification of RNA players, an area much neglected in this field, will be undertaken. Recent data indicates that transcript secondary structure in the leader sequences to certain sporulation genes is an important aspect of their regulation [19, 20]. Further, there are implications that small functional RNAs may play a



critical role in the sporulation initiation process [21, 22]. Identification of structural motifs is also proving important. For example, it has been shown that post-translational modification of certain proteins is central to signal detection [23] and sporulation [24]. Data on this latter process led to the hypothesis of a sporulation specific ADP-ribosylating protein that has not yet been found. Structural prediction and analysis may point to possible genes that encode such proteins. Comparison with known and predicted operons may provide a clue to its regulation. Other structural problems include prediction of the multimerization state of transcription factors such as Abr (that may be a hexamer) and the basis of sigma factor specificity. Finally, network analysis can be brought to bear to integrate output from these proceeding analyses and data from the literature to produce dynamical models of the sporulation/competence/motility processes in order to understand the factors that lead a particular cell to choose to execute a particular combination of pathways. We have begun a collaboration with computational structural genomics, and RNA identification laboratories at Lawrence Berkeley National Laboratory to provide the missing bioinformatic expertise and have initiated a collaboration with Genencor, International to provide specialized strains and gene array data on the whole of the *B. subtilis* genome under a wide range of host conditions. In order to integrate all these data and analysis results, to develop the tools to use the biochemical data, and construct a model of the sporulation pathway will require a good deal of man-power. The particular goals of this project will be to construct the validated model of sporulation/competence/other switching and then to design interventions that force the maximum fraction of the population down a specified path in such a way that there is minimal chance of escape via mutation.

Successful completion of this project will not only provide an unprecedented understanding of complex, cross-talking signal transduction pathways but will also provide a research model of how to database heterogeneous biological data and integrate multilevel analytical tools (from sequence to cell function) towards the solution of an important industrial, medical and military system.

D. Long Term Objectives

The scope of the proposed program is ambitious but is a model for an integrated program for dealing with complex regulatory networks. The building of the computational and theoretical infrastructure is essential if we are to move our understanding of biological function to level that exists in engineering fields. Thus, one long term goal is to develop these "dynamical genomic" tools to the stage that analogous engineering and analysis tools have achieved in electrical engineering. This means producing an industrial strength modelling system of sufficient detail and realism to predict the molecular level and functional effect of pharmaceutical challenge to microbial systems. This requires development of high-end heterogeneous database, new numerical mathematical procedures for simulation, well-designed data mining tools, and a biologist friendly user interface. Bio/Spice is being developed to address these needs.

A second important long-term goal is to really understand the molecular basis of microbial pathogenesis in a number of systems. We have chosen here to study parts of two systems that have medical and industrial importance. These systems are of sufficient complexity that it is extremely difficult if not impossible to reason about the response to perturbation without a quantitative model. We wish to be able to expand the models of these systems around these pathways to understand more and more the subtleties of cellular function.

E. Other Support

The development of the Bio/Spice tool is supported in part by a grant (N00014-99-F-0458, Instant Cell Analysis, Bio/Spice, Cellular Device and Exquisite Detection: Towards and Interactive Biology) from the Defense Advanced Research Projects Agency (DARPA) from 9/1/99-12/31/01 at an average level of $105,000/yr and by a Laboratory Directed Research Grant from Lawrence Berkeley National Laboratory for one year at $200,000. These projects provide support for the development of the Bio/Spice technology, for use with experiments on *Saccharomyces cerivisiae* (in this and other labs), *C. elegans*, and *Caulobacter crescentus* (both collaborations with Stanford). Another Laboratory Directed Research Grant from Lawrence Berkeley National Laboratory is providing partial funding for a combined *Deinococcus radiodurans* and *Bacillus subtilis* computational project ($150k for 1 year over four principle investigators). There is currently no support at all for the type-1 pili phase variation project. The award of this grant will allow the focused development of the database on *Bacillus subtilis* and *E. coli*, and the necessary experiments and modeling for the type-1 pili project and to aid in the analysis of the experimental data we are receiving on *Bacillus subtilis*.

F. References

1. Klemm, P., *Two Regulatory fim Genes, fimB and fimE, Control the Phase Variation of Type 1 Fimbriae in Escherichia coli.* The EMBO Journal, 1986. **5**(6): p. 1389-1393.
2. Klemm, P., L.B. Jensen, and S. Molin, *A Stochastic Killing System for Biological Containment of Escherichia coli.* Applied and Environmental Microbiology, 1995. **61**(2): p. 481-486.
3. Schwann, W.R., H.S. Seifert, and J.L. Duncan, *Growth Conditions Mediate Differential Transcription of fim Genes Involved in Phase Variation of Type 1 Pili.* Journal of Bacteriology, 1992. **174**(7): p. 2367-2375.
4. Orndorff, P.E. and S. Falkow, *Organization and Expression of Genes Responsible for Type 1 Piliation in Escherichia coli.* Journal of Bacteriology, 1984. **159**(2): p. 736-744.
5. Errington, J., *Determination of cell fate in Bacillus subtilis.* Trends Genet, 1996. **12**(1): p. 31-4.
6. Errington, J., *Bacillus subtilis sporulation: regulation of gene expression and control of morphogenesis.* Microbiol Rev, 1993. **57**(1): p. 1-33.
7. Stragier, P. and R. Losick, *Molecular genetics of sporulation in Bacillus subtilis.* Annu Rev Genet, 1996. **30**: p. 297-41.
8. Piggot, P.J., *Spore development in Bacillus subtilis.* Curr Opin Genet Dev, 1996. **6**(5): p. 531-7.
9. Jenal, U. and C. Stephens, *Bacterial differentiation: sizing up sporulation.* Curr Biol, 1996. **6**(2): p. 111-4.
10. Arkin, A.P. and J. Ross, *Computational functions in biochemical reaction networks.* Biophys. J., 1994. **67**: p. 560-578.
11. Arkin, A.P. and J. Ross, *Statistical Construction of Chemical Mechanisms from Measured Time-Series.* Journal of Physical Chemistry, 1995. **99**(3): p. 970-979.
12. Arkin, A.P., *et al.*, *Frequency Filtering and Decoding by Chemical and Biochemical Systems.* In preparation, 1996.
13. Arkin, A.P., .Shen, P.-D., Ross, J., *A Test Case of Correlation Metric Construction of a Reaction Pathways from Measurements. Science*, 1997. **277**(5330): p. 1275.
14. Arkin, A., *Signal Processing by Biochemical Reaction Networks*, in *Self-Organized Biodynamics and Nonlinear Control*, J. Walleczek, Editor. 1999, Cambridge University Press: Cambridge. p. accepted.
15. McAdams, H. and A. Arkin, *Stochastic Mechanisms in Gene Expression.* Proceedings of the National Academy of Sciences, USA, 1997. **94**: p. 814-819.
16. McAdams, H.H., Arkin, A.P., *Simulation of Prokaryotic Genetic Networks. Annu. Rev. Biophys. Biomol. Struct*, 1998. **27**: p. 199-224.
17. McAdams, H.H. and A.P. Arkin, *Genetic regulation at the nanomolar scale: It's a noisy business!* Trends in Genetics, 1999: p. Accepted.
18. Kruger, E., *et al.*, *The Bacillus subtilis clpC operon encodes DNA repair and competence proteins.* Microbiology, 1997. **143**(Pt 4): p. 1309-16.
19. Decatur, A., *et al.*, *Translation of the mRNA for the sporulation gene spoIIID of Bacillus subtilis is dependent upon translation of a small upstream open reading frame.* J Bacteriol, 1997. **179**(4): p. 1324-8.

20. Asayama, M., K. Saito, and Y. Kobayashi, *Translational attenuation of the Bacillus subtilis spo0B cistron by an RNA structure encompassing the initiation region.* Nucleic Acids Res, 1998. **26**(3): p. 824-30.
21. Okamoto, K. and B.S. Vold, *Activity of ribosomal and tRNA promoters of Bacillus subtilis during sporulation.* Biochimie, 1992. **74**(7-8): p. 613-8.
22. Fink, P.S., *et al.*, *Expression of small RNAs by Bacillus sp. strain PS3 and B. subtilis cells during sporulation.* FEMS Microbiol Lett, 1997. **153**(2): p. 387-92.
23. Kleerebezem, M., *et al.*, *Quorum sensing by peptide pheromones and two-component signal- transduction systems in Gram-positive bacteria.* Mol Microbiol, 1997. **24**(5): p. 895-904.
24. Huh, J.W., J. Shima, and K. Ochi, *ADP-ribosylation of proteins in Bacillus subtilis and its possible importance in sporulation.* J Bacteriol, 1996. **178**(16): p. 4935-41.

5. FACILITIES

Laboratory: Projects requiring the use of molecular biological/biochemical facilities have access to a number of resources at LBNL and University of California, Berkeley. Cold rooms, warm rooms, fermentors, fumes hoods, PCR machines, centrifuges, culture rooms, microscopy and cytometry facilities, and other equipment and support are provided as a resource to the project on site.

Clinical: N/A

Animal: N/A

Computer: In addition, to a local network of high-end workstations and servers (Silicon Graphics and Intel-based machines) the laboratory has a large array of software tools for code development, research, publishing, etc. These will have to be expanded for the current proposal.

Office: Scientific staff have offices in the Calvin Laboratory on the University campus. Various administrative staff are in close proximity to these offices. New staff members will have spaces assigned appropriate to their role.

Other: Users may apply for time at LBNL's National Energy Research Scientific Computing Center (NERSC), the world's most powerful unclassified computing resource. NERSC will also provide infrastructure support for the Center. Further information can be obtained at: http://www.nersc.gov/.

In addition, the Univerity of California, Berkeley, and the Lawrence Berkeley National Laboratory provide a rich molecular biological and computational biological environment. There is a vibrant, collaborative community of microbiologists, genetics and molecular biologists at Berkeley in the Departments of Moelcular and Cell Biology, Plant and Microbial Biology and the Public Health. In addition, the Chemistry Department has an excellent cadre of analytical and physical biochemists and the Department of Engineering has top scientists in Database design, simulation and model deduction. LBNL is also the home to a Human Genome Project, the *Drosophila*Genome Project, and the Center for Computational Genomics at NERSC. This biological community is also enhanced by the extremely close proximity of ther San Francisco Bay Area schools and biotechnology industry.

6. BUDGET

This project will require at least one half-time post-doctoral researcher to manage the biological data, maintain and design specialized queries for the *B. subtilis* and *E. coli* databases, and to help build and maintain the actual models of the pathway. A second full-time post-doctoral researcher should be supported for pursuing the *E. coli* experimental work and interfacing with our *B. subtilis* experimental collaborators as well as constructing the pathway models ($35k a piece). Desktop computers terminal will have to be provided for both post-docs and outfitted with the proper software and licenses ($6k/post-doc). Laboratory materials specific for this experimental work will also need to be provided ($5k).

Total Yearly Budget

Item	Number	Unit cost	Total Cost
Post-doctoral Researchers	1.5	$35,000	$52,500
PC Workstations/Software	2	$6,000	$12,000
Laboratory Materials	--	$5,500	$5,500
Total			$70,000