LSI FY1999-2000 Proposal

Proposal ID: lsi99-10046

Title: A System for Functional Analysis and Engineering of Bacillus Subtilis

Research Field: Bioinformatics

Campus/Lab: UC Berkeley

Status: New

Period: September 15, 2000 - September 14, 2002

Private Sponsor: Genencor International, Inc.
Palo Alto, CA 94304

Participants

Principal Investigators:

Adam P. Arkin, Ph.D.
Assistant Professor
Bioengineering
MCL 144
MCL 144
Berkeley, CA 94720
Tel: 510-495-2366
Fax: 510-486-6059
Email: aparkin@lbl.gov

Private Sponsor Liaisons:

Donald Peter Naki
Sr. Research Associate
Bioinformatics
925 Page Mill Rd.
Palo Alto, CA 94304
Tel: 650-846-5825
Email: dnaki@genencor.com
Status: Technical Liaison

Tammy Choi
Finance Manager
Finance
925 Page Mill Rd.
Palo Alto, CA 94304
Tel: 650-846-7508
Email: tchoi@genencor.com
Status: Financial Liaison

Table of Contents

Participants	A-1
Table of Contents	B-1
Technical Abstract, Non-Technical Abstract, and Project Objective	C-1
First Year In-Kind (if applicable)	D1-1
First Year Budget	D2-1
Second Year In-Kind (if applicable)	E1-1
Second Year Budget	E2-1
Entire Project Period Budget	F-1
Budget Justification	G-1
Other Support - Current	H-1
Other Support - Pending	I-1
Resources	J-1
Competitive Renewals - Progress Report (if applicable)	K-1
Resubmitted Proposals (if applicable)	L-1
Private Sponsor Data	M-1
Biographical Sketches	N-1
Research Plan	O-1
Appendix	P-1

Abstracts

ABSTRACTS: These descriptions are meant to serve as succinct, accurate descriptions of the proposed work when separated from the application.

TECHNICAL ABSTRACT FOR EXPERT REVIEWERS:

This proposal is to create novel computational tools which, when combined with the appropriate data, will aid the in creating engineering-level understanding of bacterial pathways important to industrial production. In particular, we propose to create a knowledge base of Bacillus subtilis pathways relating to sporulation, competence and secretion. Linked to this knowledge base will be a database of expression data, biochemical data and localization data suitable for input to pathways analysis tools. The initial use of the database will be to systematize the large amount of data on these pathways in order to: 1) Generate a better understanding of the sporulation decision by allowing data to be displayed ordered onto functional pathways, 2) Generate predictions on how to control sporulation and secretion processes for industrial purposes. The latter goal will be achieved development of system modeling, analysis and simulation tools currently used for analysis of E. coli pathways. These databases will be linked to the existing subtilis databases so that proprietary information from this project may be cross-referenced with publicly available information. These databases and tools will serve as a central resource for the engineering of this industrially important organism.

NON-CONFIDENTIAL, NON-TECHNICAL ABSTRACT FOR PUBLIC INFORMATION OR PROGRAM PROMOTION: State the application's broad, long-term objcetives and specific aims, making reference to the potential public benefits of the project relevant to California. Do not include proprietary or confidential information. This may be distributed before the funding descision has been finalized.

Bacillus subtilis is an industrially important organism used in the production of cleansing agents, fabric softeners and pharmaceuticals. The advantage of subtilis for these processes is its secretion system that can be made to secrete small molecules and proteins into the surrounding medium. Secretion is, however, intimately linked into the life cycle of subtilis. This process is complex because under stress a population of this undergoes a set of developmental changes that most of the time ends in sporulation, but may also go through a process called competence or motility or a number of other possibilities. The population therefore becomes heterogeneous. These processes are all intimately linked into biosynthesis and secretion so that this heterogeneity leads to non-optimal production of desired chemicals. In this work, we propose building a system for organizing a great amount of knowledge and physical data about these processes sufficient for posing and testing strategies for the rational engineering of this organism for better yield and for understanding the molecular basis for the diverse developmental decisions.

PROJECT OBJECTIVE: Provide a one to two sentence, non-technical description of the project's overall objectives and key milestones.

This project will create a unique set of database, statistical data mining and display, and pathway modeling tools that will aid in understanding the complex response of Bacillus subtilis to environmental signals and in production of industrially relevant chemicals. The data/knowledge base (SubtleCor) and statistical analysis tools should be online in the first year and the modeling and pathway analysis tools should be complete the second.

In-Kind Contribution Summary

First Year Budget Period

NON-CASH / IN-KIND CONTRIBUTION ONLY: Itemize each budget item by category (e.g. equipment, software, supplies, etc.) and by unit cost. Justify on Form G (Budget Justification).

Category	**Description** (make, model, or other description)	**Cost** (List Price or Other Cost)	**Modified Value** (List Price / 2)
Supplies	Affynetrix (B. subtilis) GeneChips	40,638.30	20,319.15
Equipment		0.00	0.00
Other		0.00	0.00
	Total	40,638.30	20,319.15

DETAILED BUDGET FOR FIRST YEAR BUDGET PERIOD
DIRECT COSTS ONLY

FROM **9/15/2000** THROUGH **9/14/2001**

PERSONNEL (Applicant organization only) NAME	ROLE ON PROJECT	APPOINTMENT (Months)	(%)	% EFFORT ON PROJECT	INSTITU-TIONAL BASE SALARY	DOLLAR AMOUNT REQUESTED SALARY REQUESTED	BENEFITS REQUESTED	TOTALS
TBN TBN	Postdoctoral Trainee	12	100	50	41,334.00	20,667.00	3,513.00	24,180.00
TBN TBN	Postdoctoral Trainee	12	100	87	41,334.00	35,960.58	6,324.10	42,284.68
Arkin Adam	PI	12	100	10	105,804.00	10,580.40	973.73	11,554.13
				SUBTOTALS		67,207.98	10,810.83	78,018.81
CONSULTANT COSTS								0.00
SUPPLIES (Itemize by category)								
Software Licensing & Support Database								
						In-Kind	20,319.15	
						Cash	917.36	21,236.51
EQUIPMENT (Itemize)								
2 computer workstation for Postdocs								
						In-Kind	0.00	
						Cash	10,000.00	10,000.00
TRAVEL								
1 conference travel to discuss results of research								2,000.00
OTHER EXPENSES (Itemize by category)								
Photocopying								
						In-Kind	0.00	
						Cash	500.00	500.00

TOTAL DIRECT COSTS FOR FIRST YEAR BUDGET PERIOD **111,755.32**

TOTAL MODIFIED DIRECT COSTS FOR FIRST YEAR BUDGET PERIOD
(Excluding costs that are not subject to indirect costs) **81,436.17**

In-Kind Contribution Summary

Second Year Budget Period

NON-CASH / IN-KIND CONTRIBUTION ONLY: Itemize each budget item by category (e.g. equipment, software, supplies, etc.) and by unit cost. Justify on Form G (Budget Justification).

Category	**Description** (make, model, or other description)	**Cost** (List Price or Other Cost)	**Modified Value** (List Price / 2)
Supplies		0.00	0.00
Equipment		0.00	0.00
Other		0.00	0.00
	Total	0.00	0.00

DETAILED BUDGET FOR SECOND YEAR BUDGET PERIOD
DIRECT COSTS ONLY

FROM **9/15/2001** THROUGH **9/14/2002**

PERSONNEL (Applicant organization only) NAME	ROLE ON PROJECT	APPOINTMENT (Months)	(%)	% EFFORT ON PROJECT	INSTITU-TIONAL BASE SALARY	DOLLAR AMOUNT REQUESTED SALARY REQUESTED	BENEFITS REQUESTED	TOTALS
TBN TBN	Postdoctoral Trainee	12	50	50	42,161.00	21,080.50	7,167.37	28,247.87
Adam Arkin	PI	12	100	12	107,920.00	0.00	0.00	0.00
TBN TBN	Postdoctoral Trainee	12	100	100	42,161.00	42,161.00	7,167.37	49,328.37
				SUBTOTALS		63,241.50	14,334.74	77,576.24
CONSULTANT COSTS								0.00
SUPPLIES (Itemize by category)								
Software licensing and support Database								
						In-Kind	0.00	
						Cash	6,859.94	6,859.94
EQUIPMENT (Itemize)								
						In-Kind	0.00	
						Cash	0.00	0.00
TRAVEL								
3 conference travel to discuss results of research								6,000.00
OTHER EXPENSES (Itemize by category)								
Photocopying								
						In-Kind	0.00	
						Cash	1,000.00	1,000.00

TOTAL DIRECT COSTS FOR FIRST YEAR BUDGET PERIOD **91,436.18**

TOTAL MODIFIED DIRECT COSTS FOR FIRST YEAR BUDGET PERIOD
(Excluding costs that are not subject to indirect costs) **91,436.18**

Budget for Entire Proposed Project Period

A. Project Summary

Budget Category		Year1		Year2		Project Total
Salaries and Benefits			78,018.81		77,576.24	**155,595.05**
Consultant Costs			0.00		0.00	**0.00**
Supplies	In-kind	20,319.15		0.00		
	Cash	917.36	21,236.51	6,859.94	6,859.94	**28,096.45**
Equipment	In-kind	0.00		0.00		
	Cash	10,000.00	10,000.00	0.00	0.00	**10,000.00**
Travel			2,000.00		6,000.00	**8,000.00**
Other Expenses	In-kind	0.00		0.00		
	Cash	500.00	500.00	1,000.00	1,000.00	**1,500.00**
Total			**111,755.32**		**91,436.18**	**203,191.50**

B. Direct Costs Summary

	LSI Cash	Private Sponsor Cash	Private Sponsor Modified In-Kind	Total
Year 1 Direct Costs	50,797.87	40,638.30	20,319.15	**111,755.32**
Year 2 Direct Costs	50,797.88	40,638.30	0.00	**91,436.18**
Total	**101,595.75**	**81,276.60**	**20,319.15**	**203,191.50**

C. Maximum Allowable In-Kind Match

	LSI Total Direct Costs (from Section B)		In-Kind Ceiling		Maximum Allowable Modified In-Kind for Match	Total Modified Inkind (from Section B)	Additional In-Kind (Total Modified Less Maximum Allowable)
Private Sponsor	101,595.75	x	20%	=	20,319.15	20,319.15	0.00

D. Indirect Costs Summary

	Year 1	Year 2
1. Modified Direct Costs	81,436.17	91,436.18
2. Indirect Cost Rate	50.40%	50.40%
3. Private Sponsor Cash-to-Total Cash Ratio	44.4444%	44.44%
4. Indirect Costs Paid by Sponsor (D1 x D2 x D3)	**18,241.70**	**20,481.70**
Total Indirect Costs Paid by Sponsor (Years 1+2)		**38,723.41**

E. Contributions Summary

			LSI	Private Sponsor
Direct Costs:	Cash		101,595.75	**81,276.60**
	In-Kind	Match Amount	-0-	20,319.15
		Additional Amount	-0-	0.00 **20,319.15**
Indirect Costs			-0-	**38,723.41**
Voluntary Contribution			-0-	**0.00**
	Total		**101,595.75**	**140,319.16**

Budget Justification

(1) Postdoctoral Associates: Bioengineers To Be Named
(2) Travel: conference travel for PI and one postdoc scholar to discuss results of research.
(3) Publication costs: requested for photocopying and binding expenses related to publishing research findings.

Other Support - Current

PI: Adam Arkin

Source: Department of Energy

Amount: $ 37,500.00

Period: 10/1/99-9/30/00

Title: Structural/Functional Genomics and Pathways: D. radiodurans and B. subtilis

PI: Arkin/Holbrook/Gordon

Potential Overlap: This project is synergetic with the current proposal but focuses on bioinformatics for network deduction, DNA repair.

Aim: To combine sequence and structure analysis along with network deduction and network simulation techniques to create a functional understanding of D. radiodurans and B. subtilis stress response pathways.

Abstract: The sequencing of the complete genomes of a variety of microbes, the metazoan C. elegans, and the soon to be completed Drosophila and human genome projects, are a driving force for understanding biological systems at a new level of complexity. The goal of the computational structural and functional genomics initiative of the future is to link these sequencing efforts to a high-throughput program of annotation and modeling of both molecular structures and functional networks. 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; 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.

PI: Adam Arkin

Source: Department of Energy

Amount: $ 200,000.00

Period: 10/1/00-9/30/01

Title: Integrated Physiome Analysis

PI: Arkin

Potential Overlap: This project funds development of simulation tools and integration with other bioinformatic techniques. As such it will aid, but does not overlap with the current proposal.

Aim: The long-term objective of this project is to integrate and systematize the large amount of different types of biological data now being generated at LBNL and at sites all over the world. The purpose of this proposal is to: 1) promote tool development by the core investigators 2) provide core effort coordination and 3) set up the computation infrastructure.

Abstract: Analysis tool development will comprise: 1) advanced algorithms for stochastic and hybrid simulation of chemical and biochemical systems, 2) integration of new RNA analytical tools with other RNA structure prediction programs, 3) tools for protein

structure analysis and classification, 4) molecular profiling data analysis like that from gene microarrays. A set of linked databases will be defined to hold this and other heterogeneous biological data and a set of middle-ware brokers written to facilitate interface and complex querying of these databases. A number of application areas have been identified including: 1) prediction and analysis of Caulobacter cell cycle, and 4) integration of knowledge and data on C.Elegans development and prediction and modeling of its vulval development.

PI: Adam Arkin

Source: Office of Naval Research

Amount: $ 100,000.00

Period: 10/1/99-9/30/00

Title: Molecular Design Institute II

PI: Arkin

Potential Overlap: No overlap.

Aim: We will build a genetic flip-flop switch out of reusable components in Saccharomyces cerevisiae. The reasons for choosing to engineer a flip-flop into yeast are manifold: 1) First, it is a good test of molecular and network engineering skills. Multiple components must be designed by genetic manipulations such that their interactions are specific and there is no cross-talk among components. Small amounts of non-specific binding or transcriptional modulation can lead to circuit malfunction. 2) The circuit is a prime example of a genetic switch architecture, similar to those that underlie development and differentiation during the life-cycle of the cell. As such it provides a test bed for examining design criteria such as stability of the two states to genetic defects in any of the components (e.g. mutations in the binding sites, decreases in binding constant of the protein/protein interaction pairs) 3) It is a probe for stochastic gene expression in eukaryotes observed in the literature).

Abstract: One of the hallmarks of signal transduction circuit found in nature is asymmetric, asynchronous design. That is, there is little standardization of parts, e.g. all the promoters have different strengths and kinetics, transcription factors are designed to have different effects at different loci, and each enzymatic reaction has its own idiosyncratic mechanism and rates. In addition, all of the heterogeneous circuit elements are executing their functions concurrently and asynchronously; biological circuits are seemingly designed to deal with the fluctuating delays, different time-scales and energy requirements associated with each component process of the overall network. These factors also make design of novel biochemical circuitry from existent parts difficult to achieve. However, it is possible to achieve some level of homogeneity among the parts by exploiting the modular nature of yeast transcription factors and promoter organization. Using the yeast `two-hybrid` system technology (9) it is possible to rationally design a relatively complex circuit from molecularly engineered components. To demonstrate this, we will build a digital-like flip-flop, a switchable memory element, into a yeast cell

PI: Adam Arkin

Source: DARPA

Other Support - Current

Amount: $ 188,301.00

Period: 10/1/00-10/30/01

Title: Exquisite Detection

PI: Arkin

Potential Overlap: No overlap

Aim: This task has two goals: (1) extension of the Bio/Spice genetic regulatory network modeling system, now under development at the Lawrence Berkeley Laboratories, for application to analysis of the dynamical behavior of the Caulobacter cell cycle regulatory network, and (2) demonstration of custom engineering of a biological circuit within a living cell by building a flip-flop, a switchable memory element, into a yeast cell.

Abstract: Genetic regulatory networks control the development and behavior of all living cells. While great strides have been made in understanding the molecular mechanisms of individual components of the network in recent years, the understanding of the basis for their dynamical behavior is not nearly so advanced. The reason is that the networks are extremely complex and asynchronous, with multiple levels of negative and positive feedback, and the outputs of individual components have highly nonlinear responses to input signals. We have demonstrated the feasibility of modeling the molecular-level behavior of these networks using an object-oriented simulation approach, loosely analogous to the Spice tool used by electrical engineers to model asynchronous, nonlinear electrical circuits (Arkin et al 1998; McAdams & Arkin 1997). This task will extend the capabilities developed in these initial applications and further demonstrate use of the system to engineer a "custom" switch circuit within a living cell.

Other Support - Pending

PI: Adam Arkin

Source: DOD Office of Naval Research

Amount: $ 225,466.00

Period: 6/1/00-5/31/03

Title: Engineering Analysis of a Genetic Switch

PI: Arkin

Potential Overlap: None

Aim: This proposal, then, has four specific aims:

1. The creation of a detailed computer model of SOS-mediated prophage induction of l and analysis of this model to predict the determinants of the threshold steepness and set-point position.
2. Use of the model to prototype genetic experiments to test these predictions.
3. Experimental implementation of genetic constructs designed to test these hypotheses and the computer prediction. Quantitative analysis of the behavior of the modified phage to compare to model predictions. Using this comparison to point to modification to the model in step 1.
4. Following validation of the wild-type model and mutant models, the use of these models as a platform for designing induction switches with specified threshold gradients and set points.

It is only by completing all four tasks that one proves complete engineering understanding of the switch circuitry. The study serves as a model for the experimental/computational cycle that will be necessary for understanding more complex circuitry and will provide an example of how to combine experimental and computational approaches to engineer specific genetic network function. Finally, this project serves as a step for producing a detailed model of the entire viral life-cycle.

Abstract: We shall combine experimental and computational approaches to explore the systems behavior of a genetic switch in phage l. The molecular components of this switch are well understood. Though structurally simple, this small regulatory network displays a threshold response with a controlled set-point. We shall develop molecularly detailed, experimentally validated models of this process in order to divine the engineering principles of this circuitry. We will then use this understanding to design and implement novel "induction" switches with chosen threshold steepness and set-points to demonstrate that these principles have indeed been understood and perhaps to be used for biotechnology purposes.

PI: Adam Arkin

Source: National Science Foundation

Amount: $ 1,127,908.00

Period: 9/1/00-8/31/03

Title: Biocomplexity: Analysis, Rational Design, and Random Evolution of Complex Gene C

PI: Arkin

Other Support - Pending

Potential Overlap: None

Aim: The goal of this project is to develop a theoretical and experimental framework to characterize naturally occurring genetic control circuits and to assemble novel genetic control circuits from the characterized parts to meet a particular control strategy. To this end, the specific aims are:
1. to create a streamlined protocol for rapidly building experimentally validated, dynamical models of biological parts (e.g., protein kinetics, translational control, mRNA degradation, elongation, transcription initiation) of sufficient detail that the behavior of networks composed of interconnected sets of these part may be robustly predicted and engineered;
2. to create very detailed, experimentally validated models of cellular environmental sensing networks in which there are varying degrees of previous knowledge;
3. to develop a streamlined protocol for computer-aided design of gene expression networks with specified input/output behaviors from the previously measured biological components using these validated network models and the mathematical description of their components.
As model systems, we will study mathematically, computationally and experimentally, three "orthogonal" examples of genetic expression switches in E. coli: the chemosensing arabinose promoter system, the type-1C pili phase variation control network, and the OmpR mediated osmoregulatory system.
Achieving these goals requires the use of techniques from molecular biology, analytical biochemistry, statistical data analysis, database design and mathematical modeling and the project brings together a team of biologists, chemists and engineers to achieve these goals. Graduate, undergraduate, and post-doctoral students trained in this effort will be a new generation of `biological control` engineers to meet the needs of the 21st century biotechnology industry.

Abstract: The production of some high-value molecules using metabolic engineering and probing metabolic and genetic function in organisms necessitate complex genetic control systems. Unlike the situation in electronic circuits, where individual elements of a circuit are well characterized and can be easily assembled from a catalogue of parts to perform a particular control operation, there are few choices in the types of genetic control circuits available to execute a particular operation. Furthermore, many genetic control circuits are incompletely characterized so that they do not give expected results when used under a new set of environmental conditions. In order for one to design a particular genetic control system for any type of control strategy, a set of well-characterized genetic "parts" is essential.

PI: Adam Arkin

Source: Stanford University

Amount: $ 1,221,085.00

Period: 8/1/00-7/31/05

Title: Whole Proteome Affinity Reagents and Devices as Data Collection Platforms for cel

PI: Arkin

Potential Overlap: not available

Aim: The goal of this project is to develop an automated method and computational tools for the high-throughput identification of small molecular weight compounds that

interact with unique gene products in the yeast Saccharomyces cerevisiae. As an end product, Stanford DNA Sequencing and Technology Development Center and Molecular Sciences Institute proposes to identified over 6000 low molecular weight compounds that functionally interact with the approximately 6000 known proteins encoded in the yeast genome. The compounds identified by this proposal could be useful as tools in studies of genetically less tractable organisms such as the pathogenic yeast Candida albicans, and other human pathogens, and potentially in the treatment of human diseases such as cancer, as many genes and biochemical pathways are universally conserved.

Abstract: Until recently, research in the biological sciences usually consisted of the detailed and painstaking analysis of a single gene product, which often took the full-time effort of several graduate students for their entire time spent as a student. With the advent of genomic DNA sequencing projects and DNA chips, it has now become possible to perform genome-wide scale investigations with roughly the same amount of effort. In other words, with the right tools, an investigator can study biological causes and effects on a much larger scale. However, this cannot be done with the standard tool set that traditionally has been available to biologists; new tools are needed. The development and use of such tools can only be accomplished at the intersection of multiple disciplines, including biology, information technology, and microsystems technology.

The goal of this project is to develop an automated method and computational tools for the high-throughput identification of small molecular weight compounds that interact with unique gene products in the yeast Saccharomyces cerevisiae. As an end product, Stanford DNA Sequencing and Technology Development Center and Molecular Sciences Institute proposes to identified over 6000 low molecular weight compounds that functionally interact with the approximately 6000 known proteins encoded in the yeast genome. The compounds identified by this proposal could be useful as tools in studies of genetically less tractable organisms such as the pathogenic yeast Candida albicans, and other human pathogens, and potentially in the treatment of human diseases such as cancer, as many genes and biochemical pathways are universally conserved.

Resources

FACILITIES: Specify the facilities to be used for the conduct of the proposed research. Indicate the performance sites and describe capacities, pertinent capabilities, relative proximity, and extent of availability to the project. Under "Other," identify support services such as machine shop, electronics shop, and specify the extent to which they will be available to the project.

Laboratory:

The Berkeley site has three laboratory wet benches and access to full molecular biological facilities (warm rooms, shakers, microscopes, hood, cold rooms, spectrophotometers, ph meters, centrifuges, etc.).

Clinical:

N/A

Animal:

N/A

Computer:

The laboratory has 10 NT workstations, 2 dual-processor Silicon Graphics Octane IRIX systems, and a 4 node (8 processor) IBM SP2 supercomputer. In addition, the laboratory has access to the National Energy Research Supercomputer Centers technical support and supercomputer farm including top-end Cray T3E`s. In addition, the laboratory is outfitted with mathematical analysis, simulation, graphing, and composition software, two color printers and 1 dedicated, and three shared laser printers.

Office:

The laboratory has three rooms (8 workspaces) dedicted to computational and dry office space. Xerox machines, fax machines and office supplies are close available.

Other:

None

MAJOR EQUIPMENT: List the most important equipment items already available for this project, noting the location and pertinent capabilities of each.

None

Private Sponsor Data

Private Sponsor: Genencor International, Inc.
925 Page Mill Rd.
Palo Alto, CA 94304
Phone: 650-846-5825
Email: dnaki@genencor.com

Nature Of California Business: Biotechnology

California Employees: 208

US Employees: 542

Worldwide Employees: 1,025

Employees in R&D: 194

Benefit to Company: The proposed research will benefit Genencor International by providing access to computational methods and software useful for the rational engineering of metabolic, signaling and regulatory pathways of Bacillus subtilis, which is used as a host in the large-scale production of industrially important biomolecules. The proposed collaboration with Dr. Adam Arkin`s lab will potentially lead to the development of an improved strain with higher product yields.

Benefit to California: The proposed research will benefit California by promoting the development of environmentally favorable industrial products. Bacillus subtilis is an important host organism used for the production of industrial enzymes which are used in a wide range of products and processes, including laundry detergents, textile processing, and animal feed production, to name a few. These environmentally friendly enzymes can replace chemicals that produce hazardous waste and/or consume large amounts of energy. The proposed research would help to make the production of enzymes more cost competitive with chemicals. In addition, the project will fund the training of students in the area of Bioinformatics, an important new discipline with a severe shortage of qualified talent.

Biographical Sketch

Provide the following information for all Principal Investigator(s) and Co-Principal Investigator(s).

NAME	POSITION TITLE
Adam P. Arkin	Assistant Professor

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE (*if applicable*)	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

RESEARCH AND PROFESSIONAL EXPERIENCE: Include in a list, in chronological order, the titles, all authors, and complete references to all publications during the past three years and to representative earlier publications pertinent to this application. If the list of publications in the last three years exceeds two pages, select the most pertinent publications. **DO NOT EXCEED TWO PAGES.**

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

Honors

M.I.T. Technology Review Top 100 Young Innovator Award Recipient.

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.

Research Plan
(Limit items a-e to fifteen pages)

a. Specific Aims

This grant is dedicated to developing an understanding of the *Bacillus subtilis* cross-talked sporulation, competence and secretion pathway sufficient for understanding the behavior of bacterial populations in controlled bioreactors and for engineering the organism for better product yield. The scope of this project includes the systematization of a great deal of the published *subtilis* data into a pathway knowledgebase and the creation of a novel heterogeneous database suitable for associating Genecor, International's microarray and other biochemical data with this knowledge. These databases will be incorporated into a set of developing data quality control and statistical analysis programs and a suite of pathway analysis tools designed for prediction of cellular function. To close the experiment/theory loop we plan to use these analytical tools to explain experimentally measured behavior of wild-type and mutant strains of *subtilis* under laboratory and industrial conditions and will use the tools developed to aid in designing strain modifications to increase yield. This collaboration will yield models bacterial sensing, development and secretion of both pure scientific and industrial interest. Because these models require multiple cycles of analysis and experiment, the students and scientists involved with data analysis and modeling will be tightly integrated with the core experimental efforts at Genencor. The specific aims are, therefore:

1) To create a integrated knowledgebase of *B. subtilis* pathway information and related data (sequence, structure, kinetics, models, etc.) interrelated to a database of experimental measurements that include molecular profiling data (RNA expression arrays, protein concentration data, images, growth curves, production rates, etc.). This system (SubtleCor) will form the basis for the analysis and serve a dual purpose as a laboratory information management solution.
2) To create an integrated suite of statistical analysis tools for use with Genencor data. These tools will be used for data quality control by providing quantitative measures of reproducibility and by creating statistical models of measurement error. These latter measures are fundamental to detecting cellular responses (gene expression, etc.) significantly above measurement error. These tools are central to creating quantitative predictions of strain behavior.

3) To create a suite of pathway display and prediction tools that take data deposited in Subtlecor and relate them to changes in known pathway behavior and predict new regulatory links in signaling networks. This software will also display the results of the statistical analysis in Aim 2.

4) To adapt a notebook and network analysis/simulation software suite (Bio/Spice) (currently in beta development) for use by the Genencor scientists to predict the dynamic activity of *B. subtilis* cellular signaling networks. To use the analytical tools

to create an quantitative understanding of the sporulation decision and the secretion pathways sufficient for explaining gene expression and bioreactor behavior measurement and, in the longer term, sufficient for aiding in re-engineering the bacterium for higher yield of particular products under specified conditions. **b.**

Relevance

This work proposed herein represents a model for a type of collaboration that will ultimately become common between biotechnology industry and academia. Currently, the directed data gathering ability of industry outstrips the ability of academic labs to collect genome wide data. This data is currently very complicated and expensive to collect. In order to get a true functional view of cellular function, information at all levels of organismal behavior must be collected. Industrial sites have developed linked production-line data gathering facilities that produce measurement at all levels including: sequence, gene expression, protein expression, kinetics, growth curves and product production efficiencies all directed towards solving a small set of biological problems. It is rare for academic laboratories to be able to produce data at some many levels, collected under similar conditions and towards a single goal. On the other hand, academic labs are often the development sites for cutting edge measurement and analysis technology. In this particular case, Genencor has collected produced a large library of well-defined strains of *B. subtilis* and collected a great deal of functional genomic data. The Berkeley team has pioneered the application of computational tools to analyze genetic and development networks in microbial systems in order to generate "engineering" principles for the systems under study. The project will allow the honing of these tools into industrial grade solutions for tracking, statistical analysis and modeling of industry data in order to aid in the engineering of new strains capable of better production of biochemical. On the research side, this project will go a long way towards producing an excellent functional genomic understanding of an important model organism.

Because this work involves molecular biology, biochemistry, mathematics, and computation students trained under this project will be uniquely able to approach the plethora of multidisciplinary problems now facing biological science. In addition, they will be trained to work in both an academic and industrial setting. The department of Bioengineering at Berkeley is creating a "molecular bioengineering" curriculum to span bioinformatics, cell and tissue engineering and more traditional areas of biomedical engineering. The training provided to the students during the course of this research will help to develop courses and mentorship programs to be included in this curriculum.

Finally, most of the work and tools produced during this work will be made public and published in the professional literature and made available online (pursuant to intellectual property agreements with Genencor). Thus, this project fulfills all the major goals of the Life Sciences Initiative.

c. Background, Significance, and Preliminary Studies

Cellular differentiation in all organisms requires a defined program of physiological and morphological changes that result in the formation of an alternate cell-type. Such morphological changes range from the construction of new surface structures such as a bacterial flagellum, to conversion to a dormant state such as a bacterial endospore or cyst, chorion assembly in an

insect, or terminal differentiation of pluripotent stem cells in mammalian development. Orchestrating such changes in a cell is an underlying program of gene expression exquisitely coordinated over a period of time by the incorporation of a number of regulatory factors that govern and modulate differential gene expression. In order for cells to coordinately make the decision to make the right cell type at the right time, they must correctly read a plethora of external and internal signals. Once the decision to execute a decision is made, successful completion of any program of cell differentiation requires accurate synthesis and assembly of macromolecules within the cell. To ensure a high fidelity in the synthesis and assembly of biological structures the cell has developed a vast network of chemical reactions implementing sophisticated feedback control, switches, amplifiers and filters that detect signals sent among cells and its own cellular state and select one of a number of possible developmental sequences while suppressing the other pathways. The hallmark of these pathways is their complexity. Typically, tens of different biochemical pathways, each composed of tens to hundreds of interacting proteins themselves, will be coupled together at various points to exchange information.

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. In fact, all bacteria of the genera *Bacillus*, *Clostridium*, *Sporosarcina* and *Thermoactinomyces* possess the ability to form highly resistant structures called endospores (Slepecky & Leadbetter, 1984). Endospore forming bacteria are important agents in human, animal and insect disease (Crowther & Baird-Parker, 1983; Walker, 1985). For example, *C. botulinum* is the causative agent of botulism, as a result of the production of an extracellular toxin (Baird-Parker, 1969). The spores of *C. botulinum* are able to survive, germinate and grow under commonly used food preservation regimes (Blocher & Busta, 1985). Some *B. cereus* and *Clostridium* strains can also cause brief, but severe, gastroenteritis due to ingestion of food contaminated with either spores or vegetative cells. Spore-forming bacteria are also the causative organisms of more familiar diseases such as tetanus and anthrax (Walker, 1985). In all these cases, spores play an important role in the etiology of the disease. Spore-forming bacteria are also important to the biotechnology, agricultural and pharmaceutical industries (Levinson et al., 1978). The crystal toxins formed during sporulation by *B. sphaericus* and *B. thuringiensis* are widely used as microbial insecticides, since they have high specificity, are biologically degradable, and host resistance appears to be rare (Ellar et al., 1985). Various bacilli, including *B. subtilis*, secrete extracellular enzymes such as subtilisin, α-amylase, β-amylase, penicillin amidase and a neutral protease, that are of considerable industrial importance (Takasuki, 1989; Metz et al., 1988; Pelletier & Sygusch, 1990; Suga et al., 1990; Quinn et al., 1989; Saxena et al., 1987). Certain species of the bacilli such as *B. megaterium, B. polymyxa, B. circulans* and *B. brevis* are known to produce antibiotics (Kohlbrenner et al., 1990; Shimada et al., 1990). Interest in these practical uses of bacilli has grown over the last ten years due to the great increase in the understanding of the biochemistry and genetics of the model organism *B. subtilis* (Piggot & Hoch, 1985). This understanding has resulted in complex pathway charts and a great deal of disparate data about pathway function, macromolecular expression, sporulation and growth behaviors, etc. However, the very

complexity of this information has precluded all but a very qualitative understanding of sporulation process as summarized here.

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 [1, 2]. This process lasts some 8 hours and uses approximately 125 developmental genes [3]. 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 [1, 4, 5]. 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.

Sporulation, however, is not the only response to nutrient stress. Some fraction of the cells never goes into sporulation at all. Some cells undergo competence, a process by which exogenous DNA is taken up into the cell and possible incorporated into the genome, and then may or may not enter sporulation. Other processes such as cell motility, antibiotic synthesis, respiration and, importantly, secretion are also coupled in and feedback on the sporulation and competence processes. All together over 140 gene products have been identified that are involved with the decision to execute or ramp up these various processes. 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, 3) the sheer number and complexity of the individual chemical interactions preclude a qualitative description for how this system functions, and 4) the measurement of the integrated functioning of these interactions have come from different laboratories under different conditions, with varying strains and with indivdual goals. This latter fact makes it difficult to interrelate data from these various sources in a rigorous way. In order to facilitate the understanding of this model system we propose to begin a quantitative analysis of these pathways. This will require efforts at all levels of biological data analysis on quality-controlled data sets. Genecor has been producing such data sets directed towards understanding the secretory behavior of *B. subtilis* in fermentors and shake-flasks and the Arkin laboratory has been developing data management and analysis tools for predicting and understanding the function of complex cellular networks. We

have already begun creation of the *Bacillus subtilis* knowledge base. The current pathway information stored in the database is shown diagrammatically below.

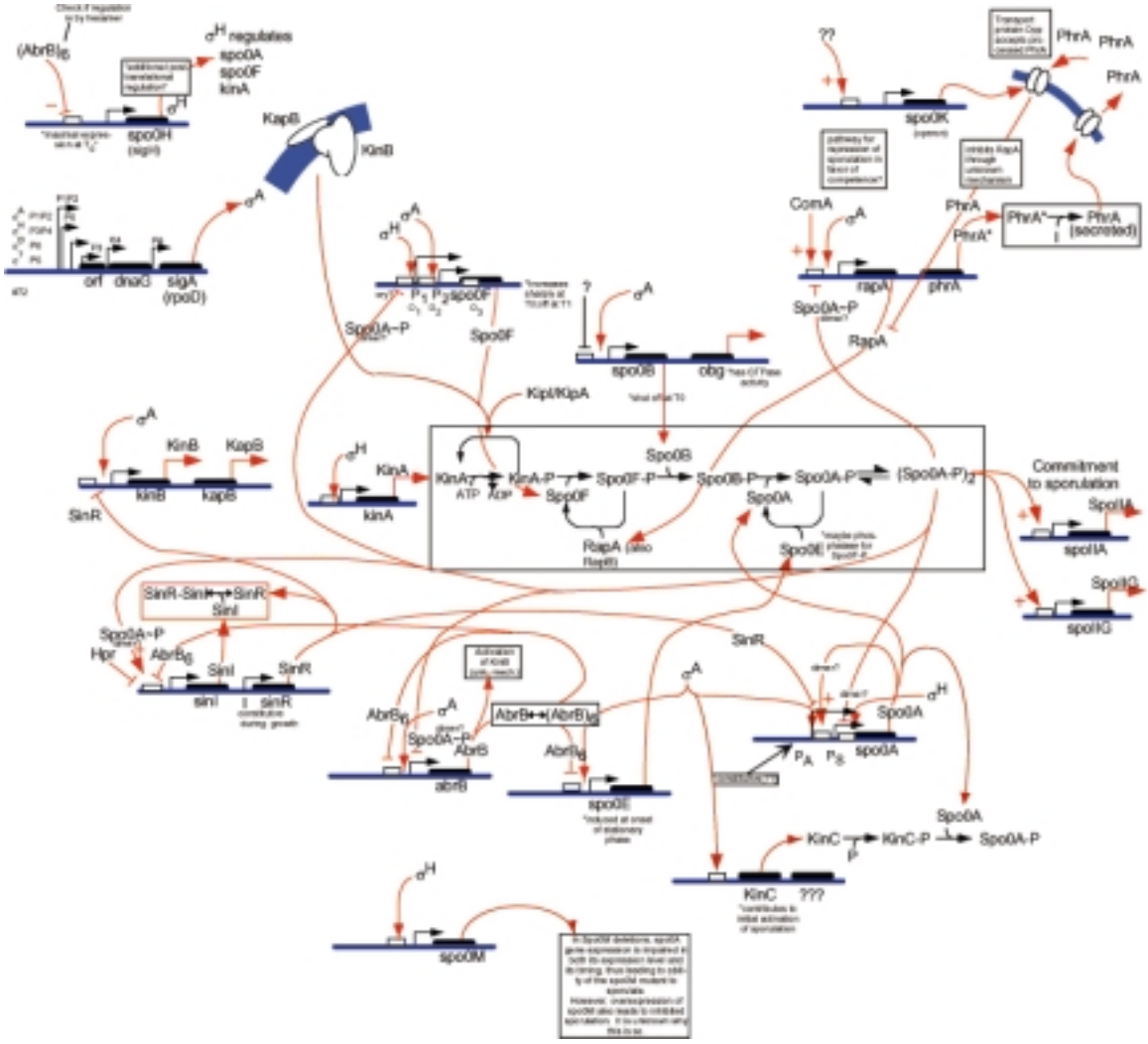


Figure 1. A fraction of the sporulation initiation decision network. Numerous other pathways including competence, motility, respiration, antibiotic production, and intermediary metabolism impinge upon and are in turn regulated by members of this pathway. The text quality is higher on the screen than when printed. However, the diagram gives a flavor for the complexity of the network.

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 and medical system.

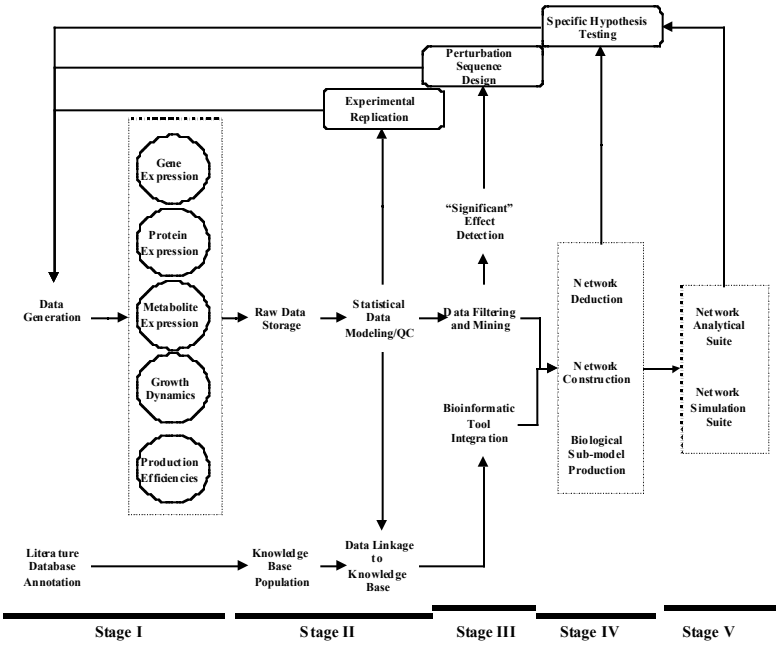
Genecor, International has a long history of engineering and adapting *B. subtilis* for industrial purposes and in performing the basic experimental research for answering specific functional questions about *B. subtilis* behavior [6-11].

The Arkin laboratory has pioneered the application of computational tools to the analysis of genetic/developmental networks and signal processing in microbial systems [12-21].

d. Research Design and Methods (limit to ten pages)

Research and Design Overview

The long term goals of this project are to develop a system that allows biological engineering of industrial microbial processes. Because of the complexity of the networks controlling *B. subtilis* development, a tight cycle of experiment and data analysis must be produced. The proposed long-term workflow is shown in the diagram below.



Each stage represents a level of data transformation. Stage I is collection of data about *B. subtilis* both from Genecor experiments and other relevant data. Stage II includes storage of raw data, basic statistical analysis for initial estimation of parameters and quality control, construction of the SubtleCor knowledge base, and linkage of the knowledge base to the data sets. At this stage, quality control and/or statistical analysis might demand replication of particular experiments to obtain reliable results. Stage III involves basic statistical mining of data in which statistically significant cellular responses to perturbation are detected and logged. These findings will then be used to design further experiments. All data will be combined with bioinformatically filtered sections of the knowledge base to relate cellular responses to their biochemical and genetic context. This intermediate goal is the basis for initial qualitative pathway maps. It will also guide the choice and design of the experiments best suited to provide the quantitative data needed for the modeling effort in Stages IV and V (where deemed appropriate). Stage IV analyses take these latter data and organize them into molecular signaling networks. Some of these analytical tools will also create testable hypotheses about missing pieces of the network: reaction lists, connectivity, stoichiometry, kinetics. Stage V is construction of detailed quantitative models of

the signal transduction cascades at various levels of detail. It is here that the deepest insights about the cellular function of the individual network components can be generated, but it also the most data-intensive part of the process and is thus a late stage process. The research goals of the proposal represent a modest fraction of this long term goal: Creation of the data and knowledgebase, development of basic statistical analysis tools and the early beginnings of dynamical modeling directed towards organizing the Genenocor data in a manner readily queryable by Genencor scientists and to begin to understand the basis of the sporulation decision.

Data Development (Stage I)

Genenor, International has been generating data on *Bacillus subtilis* as a central part of their program in producing industrially important enzymes such as subtilisin and α-amylase. They have developed a library of mutants in various parts of the signal transduction and secretion pathways and have begun collected gene expression, protein expression and other biochemical data on the behavior of these strains under various conditions in culture. As part of this proposal, an agreement between Genencor and UC will be reached in which significant parts of this data will be released for analysis many results of which will be made publicly available. The Berkeley team, will aid in creating of the database that will allow efficient systematization of this data.

To start with, mutants in the degU and scoC locus will be compared to wild-type under various shake-flask and fermentation conditions. These loci are involved in the protease pathways, competence and in sporulation thus form a good opening set for the analytical tools. Mutants will be of both of the loss of function and gain of function type (always on). Microarray data will be taken for whole genome expression as a function of growth phase, pH, oxygen tension, and shake-flask or fermentor conditions. These experiments, along with other data on sporulation efficiency and other biochemical measured will be compiled and related to pathway information. Initial models will be built to explain the response. These analyses will most likely succeed to the point wherein specific experiments could be suggested to measure critical parameters necessary for improving model prediction. If the cycle proves successful, then other mutants that are relevant and non-proprietary may be tried included changes in KinA, KinB or SpoIIE.

Databasing and Statistical Data Modeling

Because of the variety of data available on *Bacillus subtilis* development of the relational database for storage of this information will by an ongoing process. For "array" type gene expression data, for example, our current implementations are based on the GATC (http://www.gatconsortium.org/) and the pan-bacterial microarray database (http://bugarrays.stanford.edu/) standards. We have custom database schema for storing proteomic data, and generic "series" data in which the same measurement is performed at different time points, under different environmental conditions or different bacterial strains. A currently simple laboratory information management (LIM) schema is implemented along side these tables in order to relate measurements to particular samples, experimental protocols, experimenters, etc. The current implementation of the Berkeley database is in MySQL, a freeware database that has stable implementation in various UNIX and LINUX platforms as well as in Windows.

The combination of a LIM system with the database is necessary if a statistical model of measurement error is to be created. This model is necessary in order to attach significance to changes in behavior observed under different conditions. For each type of data produced we will design experimental protocols so that the error introduced at each stage of the experiment may be tracked. For example, for the nylon membrane-based expression measurements there are the RNA prep, the PCR, the labeling and hybridization stage, the membrane batch, the wash protocol and the image analysis, all of which can lead to difference in otherwise identical measurements. Replicates passed through each stage of the experiment will help to determine the major sources of error.

Once a statistical error model for each data type is derived, then results across all data types from experiments with the same strain under different conditions or different strains under the same conditions may be brought up by SQL query from the database and cross-compared and grouped by standard statistical methods such as factor analysis **(ref)**, clustering **(ref)** and multidimensional scaling **(ref)**. As the data sets grow and the criteria for classification becomes more complex (e.g. strain grows in 12 liter bioreactors but not 100 liter bioreactors and shows heat shock response) more sophisticated statistical classification methods such as Support Vector Machines may be used to classify new measured behaviors.

Knowledge-basing

Ultimately, in order to understand cellular function, the data generated during the course of experiment, must be made consistent with the compiled knowledge of the biochemical and genetic pathways controlling the behavior. Figure 1 above is a representation of part of that knowledge compiled from over two hundred papers in the field. Such charts are difficult to construct and thus the ability to systematize literature information and field knowledge into a database of pathways and models ensures that once compiled the information will not be lost, can be modified in a controlled way and may be directly associated with experimental data. The combination of pathway representations and data opens the door for validated modeling of cell function either through simulation, static analysis or logical deduction tools (see Bio/Spice below). We propose to leverage ongoing efforts in the construction of pathway information and model knowledgebase at Berkeley and specialize it for use with *B. subtilis* secretion and biosynthetic pathways of use to Genencor and in-line with the Berkeley efforts to understand the sporulation decision. Schema for the knowledgebase are under development:

Data in this database will be linked to outside databases of sequence, structure and function such as GenBank, PDB, Subtlist and the WIT database recently purchased by Genencor, Inc for *subtilis*. The database described in the last section combined with the *subtilis* specific knowledgebase will form the SubtleCor database proposed in this project.

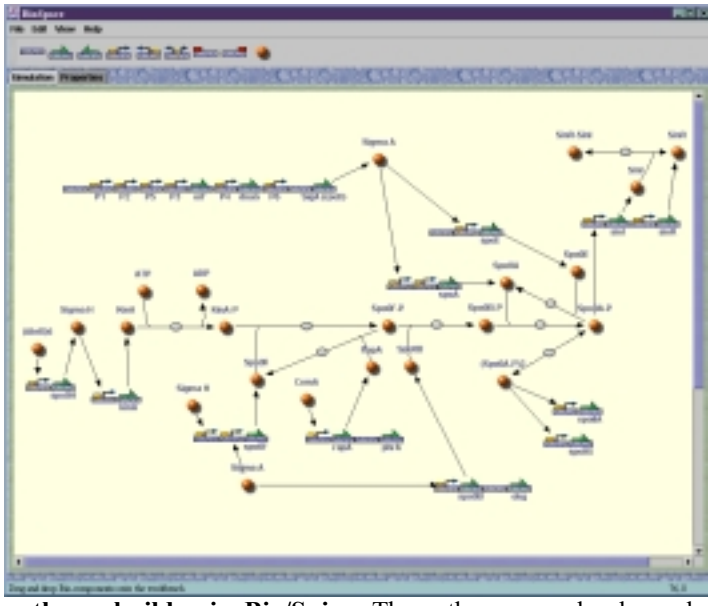
Modeling

Though much insight in cellular function can be gained simply by identifying the isolated molecular function associated with the protein to be expressed, the cellular implication for the change in this specific activity can only be understood in the context of the regulatory network. Metabolic modeling provides the analytical tools that facilitate analysis in this context. Metabolic modeling, both steady-state methods such as flux-balance analysis and metabolic control analysis as well as full dynamic simulation, have long been used both in academia and industry as a tool for understanding biological function. However, the data necessary for robust,

predictive modeling of *in vivo* processes has been lacking. In addition, more modern techniques of molecular biology have called into question many of the assigned functions of metabolic proteins or assigned them multiple such activities. Further, the role of genetic regulatory interactions has been largely ignored or treated with inappropriate mathematical models. However, now, with the new wealth of data generated with high-throughput molecular profiling methodology, it is possible and paramount to create predictive *dynamic* models of metabolic function so that cells can be better engineered for industrial purposes and so that metabolic defects can be better understood and controlled.

The chemistry underlying most interesting cellular signaling behavior is complicated enough to require sophisticated systematization and analysis. In the face of regulatory systems as complex as these, biological intuition fails, and it is necessary to develop simulation and analysis tools to aid in dissecting the networks. As circuits go, cellular circuits are of a theoretically difficult sort. They are asynchronous, most generally analog (rather than digital or Boolean), and may have stochastic elements because of the low number of molecules involved (especially gene expression) [13, 14, 16]. *The very complexity of cellular systems and their measurement, which is often cited as a reason that modeling cannot be done, is actually the very reason why it should.* There is no question that realistic models of these processes can and must be made. What remains is to design the tools that can scale up these simulations to deal consistently and accurately with larger systems that must be modeled with multiple levels of abstraction where knowledge is more or less elaborated.

Much of biochemical modeling has focused on the stationary state of the system of interest and on small perturbations around this state. Methods such as Metabolic Control Theory and Flux Balance Analysis can, when used with good stoichiometric models of mass-flow in pathway, be powerful tools for analyzing populations of cells in bioreactors or other highly controlled or homeostatic situations [22-25]. However, when the metabolic state of the system is dynamic enough to cause changes in even gene expression, then many of the assumption behind the steady-state methodologies above are violated. Moreover, many of the regulatory interactions that are not active or are functionally hidden at steady-state come into play. Under these situations full dynamical models of the network must be made. To make such dynamical models detailed mechanistic hypotheses for every step in the pathway must be made, parameterized and validated by experimental data. We have been developing methods in chemical reaction network deduction, mechanistic hypothesis testing, and detailed simulation of complex genetic and biochemical systems [13-19, 21]. The model deduction, analysis and simulation tools built during the course of this work have been retooled into a package called Bio/Spice.



One view of the active pathway builder in Bio/Spice. The pathways can be drawn by the biologist or loaded from the database if it is stored there. Data may be linked to this pathway and displayed. In addition, this pathway is interactive and can be made to link out to other databases. These pathways may be associated with mathematical models as well that may be submitted to the Bio/Spice analytical kernel.

The Bio/Spice project goals are to produce a biological network analysis suite as useful to biologists as the SPICE simulation tool is to electrical engineers [26-28]. Bio/Spice is a pathway dynamics project management and analysis suite that can hook into data warehouses of molecular profiling information and relate this information to a pathway database. These may then be combined with the novel model database developed here that allows the rapid static (stoichiometric network analysis, bifurcation analysis) and dynamic (simulation) analysis of the pathway. The software allows modeling at multiple levels of abstraction, allows the user to try out different hypotheses of regulatory network structure in order to explain a set of molecular profiling data, and can display simulation results, pathways, and pathways adorned with expression and simulation data. Currently, this software is somewhat primitive and only 'professional' users can easily work with it. Part of this proposal will go to making this software robust enough for industrial users to employ and count on. Further, this project will allow this Bio/Spice team to compile pathway knowledge bases for *Bacillus subtilis* systems of interest to Genencore, link Bio/Spice to Genencor data, develop mathematical models appropriate for understanding *subtilis* function, and train students and scientific users in the use of the tool. Bio/Spice is optimized for use with prokaryotic and very simple eukaryotic systems. However, there are problems specific to sporulating bacteria that the system that are currently beyond the capabilities of the Bio/Spice tool, in part because of the simulation difficulties for spatially distributed systems (strong spatial effects, complex mechanisms of gene expression, etc.) and in part because of the lack of well-validated physical-chemical models for the dynamics of the network components (*e.g.,* the kinetics of genome translocation into the forespore). One of the central research topics will be how to create experimentally validatable models of these important network components. Successful completion of these modules will also go long way towards

better modeling of eukaryotic systems. For more information on the Bio/Spice program goals see: http:/genomics.lbl.gov/~biospice.

Closing the cycle

Once the database is significantly populated and the Bio/Spice system is brought up the specifications for the *subtilis* system it can be used to search the database with complex queries, visualize complex data in functional context (on pathways), test mechanistic hypotheses for how different reactor conditions, feedstocks, pathway modifications can change pathway and ultimately secretion and growth behavior. Going through the process of building validatable models of cellular function will rapidly point out where data will have to be better or replicated. With the data fitting capabilities of Bio/Spice, when observed data cannot be fit by the current model, the hypothesis testing module can aid in identifying where missing interactions or components might be brought into play to match that data. This in turn suggests experiments to find these components.

e. Timetable and Milestones

This proposal calls for two years of funding. We break the period up into six months segments so that progress may be tracked in a formal way. Because of the need for close contact between teams, we will call for monthly design and data explanation sessions and semiannual progress report meeting. The expected milestones for each of the six month periods are:

Months 1-6: Creation of the initial database from transferred Genencor data, finalization of database schema for the knowledgebase and the data/knowledge relations. Population of the knowledgebase with sporulation, competence and secretion pathways.

Month 7-12: Development of statistical data analysis tools, development and exchange of quality control/statistical validation procedures for data, development of user/database interfaces sufficient for Stage I and Stage II use of SubtleCor database. Development of data visualization for multiple-types of data onto functional, annotated pathways. Creation of initial explanatory models for shake-flask and fermentor data in the context of gene expression patterns of non-proprietary mutants. Generation of hypothesis of pathway interactions that lead to differences in mutant and wild-type behaviors under different conditions.

Months 13-18: (*Conditional on sufficient progress in the first year. Genecor may decide to rescind funding if they deem the project is not in their interest)* Development of multivariate analysis tools that relate growth conditions and strain genotype to growth, sporulation and secretion phenotypes reliably. Directed experiments to obtain better data for such analysis for phenotypes of Genencor interest. Initial dynamical models of *subtilis* pathways constructed and tested against SubtleCor data. These will be the basis for understanding the molecular basis of the multivariate statistical data models and for detecting when our pathway knowledge is not sufficient for understanding the behavior of different strains of *subtilis* under different conditions.

Months 19-24: Use of the Bio/Spice tool to create mechanistic hypotheses to explain Genecor data. Experimental testing of the hypotheses. Development of hypotheses of how to change growth conditions and/or pick or make mutant strains to best fit production goals. Experimental testing of model predictions.

It should be noted that Genencor, International is already engaged in experimental functional genomic research on *B. subtilis* and thus the first year of development is largely on the Berkeley side while the tools for collaborative analysis are developed and tested (with Genencor help). In addition, the work proposed at Berkeley is leveraged off a number of different projects that require development of pathway knowledgebases, databases for different forms of data, and development of Bio/Spice (see Other Support). The postdoctoral fellow and graduate student requested for this project will be trained in both the experimental and computational ends of this project thus facilitating the integration of experiment and theory.

f. Literature Cited

References for Sporulation[1-5, 29-109]
References for Competence[48, 52, 54, 58, 60, 63, 64, 66-72, 76, 79, 81-83, 85-87, 89-91, 94, 95, 110-126]
References for Motility [48, 54, 67-71, 75, 85, 118, 127, 128]
References for DNA Repair [73, 78, 115, 117, 119, 120, 123]
References for Antibiotic production [54, 70, 73, 75, 90, 97, 101, 107, 110]
References for metabolism[105]
References for Pho Regulation [43]

1. Errington, J., *Determination of cell fate in Bacillus subtilis.* Trends Genet, 1996. **12**(1): p. 31-4.
2. Errington, J., Bacillus subtilis sporulation: regulation of gene expression and control of morphogenesis. Microbiol Rev, 1993. **57**(1): p. 1-33.
3. Stragier, P. and R. Losick, *Molecular genetics of sporulation in Bacillus subtilis.* Annu Rev Genet, 1996. **30**: p. 297-41.
4. Piggot, P.J., *Spore development in Bacillus subtilis.* Curr Opin Genet Dev, 1996. **6**(5): p. 531-7.
5. Jenal, U. and C. Stephens, *Bacterial differentiation: sizing up sporulation.* Curr Biol, 1996. **6**(2): p. 111-4.
6. Miller, B.S. and M.R. Diaz-Torres, Proteome analysis of biofilms: growth of Bacillus subtilis on solid medium as model. Methods Enzymol, 1999. **310**: p. 433-41.
7. Shafikhani, S., et al., Generation of large libraries of random mutants in Bacillus subtilis by PCR-based plasmid multimerization. Biotechniques, 1997. **23**(2): p. 304-10.
8. Ruppen, M.E., G.L. Van Alstine, and L. Band, *Control of intracellular serine protease expression in Bacillus subtilis.* J Bacteriol, 1988. **170**(1): p. 136-40.
9. Winters, P., et al., The ampS-nprE (124 degrees-127 degrees) region of the Bacillus subtilis 168 chromosome: sequencing of a 27 kb segment and identification of several genes in the area. Microbiology, 1996. **142**(Pt 11): p. 3033-7.
10. Miller, B.S., et al., Solid medium labeling applied to two-dimensional gel electrophoresis. Anal Biochem, 1997. **245**(2): p. 245-7.
11. Ferrari, E., et al., Transcription of Bacillus subtilis subtilisin and expression of subtilisin in sporulation mutants. J Bacteriol, 1988. **170**(1): p. 289-95.
12. McAdams, H.H. and A.P. Arkin, *Genetic regulation at the nanomolar scale: It's a noisy business!* Trends in Genetics, 1999: p. Accepted.
13. McAdams, H.H., Arkin, A.P., Simulation of Prokaryotic Genetic Networks. Annu. Rev. Biophys. Biomol. Struct, 1998. **27**: p. 199-224.
14. McAdams, H. and A. Arkin, *Stochastic Mechanisms in Gene Expression.* Proceedings of the National Academy of Sciences, USA, 1997. **94**: p. 814-819.
15. 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.
16. Arkin, A., J. Ross, and H.H. McAdams, Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected E. coli cells. Genetics, 1998. **149**(4): p. 1633-1648.
17. 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.
18. Arkin, A.P. and J. Ross, *Computational functions in biochemical reaction networks.* Biophys. J., 1994. **67**: p. 560-578.
19. 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.
20. Arkin, A.P., et al., Frequency Filtering and Decoding by Chemical and Biochemical Systems. In preparation, 1996.
21. Swanson, C., A. Arkin, and J. Ross, *An Endogenous Calcium Oscillator May Control Early Embryonic Division.* Proceedings of the National Academy of Sciences, USA, 1996. **94**: p. 1194-1199.
22. Varma, A. and B.O. Palsson, Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. Appl Environ Microbiol, 1994. **60**(10): p. 3724-31.
23. Mandalam, R.K. and B.O. Palsson, Elemental balancing of biomass and medium composition enhances growth capacity in high-density Chlorella vulgaris cultures. Biotechnol Bioeng, 1998. **59**(5): p. 605-11.
24. Kacser, H. and J.A. Burns, Symp. Soc. Exp. Biol., 1973. **7**: p. 65-104.
25. Heinrich, R. and T.A. Rapoport, European Journal of Biochemistry, 1974. **42**: p. 89-95.
26. Maxfield, C. and G. Moretti, *Introduction to analog simulation.* Printed Circuit Design, 1995. **12**(10): p. 32-5.
27. von Sosen, H. Advanced modeling techniques for subthreshold and full chip simulation. 1995.
28. Puhan, J., T. Tuma, and I. Fajfar, *SPICE for Windows 95/98/NT.* Elektrotehniski Vestnik, 1998. **65**(5): p. 267-71.
29. Grossman, A.D., Integration of developmental signals and the initiation of sporulation in B. subtilis. Cell, 1991. **65**(1): p. 5-8.
30. Kunkel, B., Compartmentalized gene expression during sporulation in Bacillus subtilis. Trends Genet, 1991. **7**(5): p. 167-72.
31. Trach, K., et al., Control of the initiation of sporulation in Bacillus subtilis by a phosphorelay. Res Microbiol, 1991. **142**(7-8): p. 815-23.
32. Seror, S.J., A. Levine, and F. Vannier, Post-initiation control of chromosomal replication in Bacillus subtilis: a mechanism for limiting over-replication or for duplicating key growth and sporulation genes? Res Microbiol, 1991. **142**(7-8): p. 861-7.
33. Margolis, P., A. Driks, and R. Losick, Differentiation and the establishment of cell type during sporulation in Bacillus subtilis. Curr Opin Genet Dev, 1991. **1**(3): p. 330-5.
34. Losick, R. and P. Stragier, Crisscross regulation of cell-type-specific gene expression during development in B. subtilis. Nature, 1992. **355**(6361): p. 601-4.
35. Errington, J. and N. Illing, Establishment of cell-specific transcription during sporulation

in Bacillus subtilis. Mol Microbiol, 1992. **6**(6): p. 689-95.
36. Strauch, M.A. and J.A. Hoch, *Sporulation in prokaryotes and lower eukaryotes.* Curr Opin Genet Dev, 1992. **2**(5): p. 799-804.
37. Yudkin, M., *Spore formation in Bacillus subtilis.* Sci Prog, 1993. **77**((Pt 1-2)): p. 113-30.
38. Hoch, J.A., Regulation of the phosphorelay and the initiation of sporulation in Bacillus subtilis. Annu Rev Microbiol, 1993. **47**: p. 441-65.
39. Hoch, J.A., Regulation of the onset of the stationary phase and sporulation in Bacillus subtilis. Adv Microb Physiol, 1993. **35**: p. 111-33.
40. Hoch, J.A., The phosphorelay signal transduction pathway in the initiation of Bacillus subtilis sporulation. J Cell Biochem, 1993. **51**(1): p. 55-61.
41. Strauch, M.A. and J.A. Hoch, *Signal transduction in Bacillus subtilis sporulation.* Curr Opin Genet Dev, 1993. **3**(2): p. 203-12.
42. Duncan, L., S. Alper, and R. Losick, Establishment of cell type specific gene transcription during sporulation in Bacillus subtilis. Curr Opin Genet Dev, 1994. **4**(5): p. 630-6.
43. Hulett, F.M., The signal-transduction network for Pho regulation in Bacillus subtilis. Mol Microbiol, 1996. **19**(5): p. 933-9.
44. Shapiro, L. and R. Losick, *Protein localization and cell fate in bacteria.* Science, 1997. **276**(5313): p. 712-8.
45. Stephens, C., *Bacterial sporulation: a question of commitment?* Curr Biol, 1998. **8**(2): p. R45-8.
46. Arigoni, F., et al., The SpoIIE phosphatase, the sporulation septum and the establishment of forespore-specific transcription in Bacillus subtilis: a reassessment [In Process Citation]. Mol Microbiol, 1999. **31**(5): p. 1407-15.
47. 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.
48. Bai, U., I. Mandic-Mulec, and I. Smith, SinI modulates the activity of SinR, a developmental switch protein of Bacillus subtilis, by protein-protein interaction. Genes Dev, 1993. **7**(1): p. 139-48.
49. Chung, J.D., et al., Gene expression in single cells of Bacillus subtilis: evidence that a threshold mechanism controls the initiation of sporulation. J Bacteriol, 1994. **176**(7): p. 1977-84.
50. Chung, J.D., S. Conner, and G. Stephanopoulos, *Flow cytometric study of differentiating cultures of Bacillus subtilis.* Cytometry, 1995. **20**(4): p. 324-33.
51. Cosby, W.M. and P. Zuber, Regulation of Bacillus subtilis sigmaH (spo0H) and AbrB in response to changes in external pH. J Bacteriol, 1997. **179**(21): p. 6778-87.
52. Cosby, W.M., et al., Altered srf expression in Bacillus subtilis resulting from changes in culture pH is dependent on the Spo0K oligopeptide permease and the ComQX system of extracellular control. J Bacteriol, 1998. **180**(6): p. 1438-45.
53. 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.
54. Dubnau, D., *Genetic competence in Bacillus subtilis.* Microbiol Rev, 1991. **55**(3): p. 395-

424.
55. Dunn, G., Statistical analysis of the patterns of spore formation in Bacillus subtilis. J Gen Microbiol, 1980. **121**(1): p. 113-6.
56. 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.
57. Fujita, M. and Y. Sadaie, Feedback loops involving Spo0A and AbrB in in vitro transcription of the genes involved in the initiation of sporulation in Bacillus subtilis. J Biochem (Tokyo), 1998. **124**(1): p. 98-104.
58. Gasc, A.M., et al., Organization around the dnaA gene of Streptococcus pneumoniae. Microbiology, 1998. **144**(Pt 2): p. 433-9.
59. Gober, J.W., Compartmentalized transcription and the establishment of cell type during sporulation in Bacillus subtilis. Bioessays, 1992. **14**(2): p. 125-8.
60. Grossman, A.D., Genetic networks controlling the initiation of sporulation and the development of genetic competence in Bacillus subtilis. Annu Rev Genet, 1995. **29**: p. 477-508.
61. Harwood, C.R., Bacillus subtilis and its relatives: molecular biological and industrial workhorses. Trends Biotechnol, 1992. **10**(7): p. 247-56.
62. Kroos, L., et al., Control of sigma factor activity during Bacillus subtilis sporulation [In Process Citation]. Mol Microbiol, 1999. **31**(5): p. 1285-94.
63. Lazazzera, B.A., J.M. Solomon, and A.D. Grossman, An exported peptide functions intracellularly to contribute to cell density signaling in B. subtilis. Cell, 1997. **89**(6): p. 917-25.
64. LeDeaux, J.R., J.M. Solomon, and A.D. Grossman, Analysis of non-polar deletion mutations in the genes of the spo0K (opp) operon of Bacillus subtilis. FEMS Microbiol Lett, 1997. **153**(1): p. 63-9.
65. Levin, P.A. and R. Losick, Transcription factor Spo0A switches the localization of the cell division protein FtsZ from a medial to a bipolar pattern in Bacillus subtilis. Genes Dev, 1996. **10**(4): p. 478-88.
66. Liu, L., et al., Plasmid-amplified comS enhances genetic competence and suppresses sinR in Bacillus subtilis. J Bacteriol, 1996. **178**(17): p. 5144-52.
67. Liu, J. and P. Zuber, A molecular switch controlling competence and motility: competence regulatory factors ComS, MecA, and ComK control sigmaD-dependent gene expression in Bacillus subtilis. J Bacteriol, 1998. **180**(16): p. 4243-51.
68. Mandic-Mulec, I., et al., Sin, a stage-specific repressor of cellular differentiation. J Bacteriol, 1992. **174**(11): p. 3561-9.
69. Mandic-Mulec, I., L. Doukhan, and I. Smith, *The Bacillus subtilis SinR protein is a repressor of the key sporulation gene spo0A*. J Bacteriol, 1995. **177**(16): p. 4619-27.
70. Marahiel, M.A., M.M. Nakano, and P. Zuber, *Regulation of peptide antibiotic production in Bacillus.* Mol Microbiol, 1993. **7**(5): p. 631-6.
71. Msadek, T., et al., ClpP of Bacillus subtilis is required for competence development, motility, degradative enzyme synthesis, growth at high temperature and sporulation. Mol Microbiol, 1998. **27**(5): p. 899-914.
72. Mueller, J.P., et al., Identification of Bacillus subtilis adaptive response genes by subtractive differential hybridization. Res Microbiol, 1991. **142**(7-8): p. 805-13.

73. Nicholson, W.L., L. Chooback, and P. Fajardo-Cavazos, Analysis of spore photoproduct lyase operon (splAB) function using targeted deletion-insertion mutations spanning the Bacillus subtilis operons ptsHI and splAB. Mol Gen Genet, 1997. **255**(6): p. 587-94.
74. Nishiguchi, M., et al., Structural requirements of Bacillus subtilis small cytoplasmic RNA for cell growth, sporulation, and extracellular enzyme production. J Bacteriol, 1994. **176**(1): p. 157-65.
75. Nishihara, T. and E. Freese, *Motility of Bacillus subtilis during growth and sporulation.* J Bacteriol, 1975. **123**(1): p. 366-71.
76. Ogura, M. and T. Tanaka, Bacillus subtilis DegU acts as a positive regulator for comK expression. FEBS Lett, 1996. **397**(2-3): p. 173-6.
77. 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.
78. Pedraza-Reyes, M., F. Gutierrez-Corona, and W.L. Nicholson, Temporal regulation and forespore-specific expression of the spore photoproduct lyase gene by sigma-G RNA polymerase during Bacillus subtilis sporulation. J Bacteriol, 1994. **176**(13): p. 3983-91.
79. Perego, M. and J.A. Hoch, Cell-cell communication regulates the effects of protein aspartate phosphatases on the phosphorelay controlling development in Bacillus subtilis. Proc Natl Acad Sci U S A, 1996. **93**(4): p. 1549-53.
80. Predich, M., G. Nair, and I. Smith, Bacillus subtilis early sporulation genes kinA, spo0F, and spo0A are transcribed by the RNA polymerase containing sigma H. J Bacteriol, 1992. **174**(9): p. 2771-8.
81. Roggiani, M. and D. Dubnau, ComA, a phosphorylated response regulator protein of Bacillus subtilis, binds to the promoter region of srfA. J Bacteriol, 1993. **175**(10): p. 3182-7.
82. Rudner, D.Z., et al., The spo0K locus of Bacillus subtilis is homologous to the oligopeptide permease locus and is required for sporulation and competence. J Bacteriol, 1991. **173**(4): p. 1388-98.
83. Siranosian, K.J. and A.D. Grossman, Activation of spo0A transcription by sigma H is necessary for sporulation but not for competence in Bacillus subtilis. J Bacteriol, 1994. **176**(12): p. 3812-5.
84. Smirnov, V.V., et al., [The growth and sporulation of Bacillus subtilis under different aeration conditions]. Mikrobiol Zh, 1993. **55**(3): p. 38-44.
85. Smith, I., I. Mandic-Mulec, and N. Gaur, *The role of negative control in sporulation.* Res Microbiol, 1991. **142**(7-8): p. 831-9.
86. Solomon, J.M., et al., Convergent sensing pathways mediate response to two extracellular competence factors in Bacillus subtilis. Genes Dev, 1995. **9**(5): p. 547-58.
87. Solomon, J.M., B.A. Lazazzera, and A.D. Grossman, Purification and characterization of an extracellular peptide factor that affects two different developmental pathways in Bacillus subtilis. Genes Dev, 1996. **10**(16): p. 2014-24.
88. Stephens, C., M. Singer, and L. Shapiro, *Bacterial sporulation. An ATP/ADP switch.* Curr Biol, 1994. **4**(7): p. 630-2.
89. Turgay, K., et al., Competence in Bacillus subtilis is controlled by regulated proteolysis of a transcription factor. Embo J, 1998. **17**(22): p. 6730-8.

90. van Sinderen, D., et al., Characterization of the srfA locus of Bacillus subtilis: only the valine-activating domain of srfA is involved in the establishment of genetic competence. Mol Microbiol, 1993. **8**(5): p. 833-41.
91. van Sinderen, D., R. Kiewiet, and G. Venema, Differential expression of two closely related deoxyribonuclease genes, nucA and nucB, in Bacillus subtilis. Mol Microbiol, 1995. **15**(2): p. 213-23.
92. Vold, B.S., et al., Transcriptional analysis of Bacillus subtilis rRNA-tRNA operons. I. The tRNA gene cluster of rrnB has an internal promoter. J Biol Chem, 1988. **263**(28): p. 14480-4.
93. Webb, C.D., et al., Use of green fluorescent protein for visualization of cell-specific gene expression and subcellular protein localization during sporulation in Bacillus subtilis. J Bacteriol, 1995. **177**(20): p. 5906-11.
94. Weinrauch, Y., et al., A Bacillus subtilis regulatory gene product for genetic competence and sporulation resembles sensor protein members of the bacterial two- component signal-transduction systems. Genes Dev, 1990. **4**(5): p. 860-72.
95. Yoshikawa, H., et al., Temperature-sensitive sporulation caused by a mutation in the Bacillus subtilis secY gene. J Bacteriol, 1993. **175**(11): p. 3656-60.
96. Zhang, B. and L. Kroos, A feedback loop regulates the switch from one sigma factor to the next in the cascade controlling Bacillus subtilis mother cell gene expression. J Bacteriol, 1997. **179**(19): p. 6138-44.
97. 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.
98. Lord, M., D. Barill, and M.D. Yudkin, Replacement of Vegetative sigmaA by Sporulation-Specific sigmaF as a Component of the RNA Polymerase Holoenzyme in Sporulating Bacillus subtilis. J Bacteriol, 1999. **181**(8): p. 2346-2350.
99. Perego, M., et al., Characterization of the gene for a protein kinase which phosphorylates the sporulation-regulatory proteins Spo0A and Spo0F of Bacillus subtilis. J Bacteriol, 1989. **171**(11): p. 6187-96.
100. Strauch, M.A., et al., The transition state transcription regulator abrB of Bacillus subtilis is a DNA binding protein. Embo J, 1989. **8**(5): p. 1615-21.
101. Han, W.D., et al., A novel sporulation-control gene (spo0M) of Bacillus subtilis with a sigmaH-regulated promoter. Gene, 1998. **217**(1-2): p. 31-40.
102. Grimshaw, C.E., et al., Synergistic kinetic interactions between components of the phosphorelay controlling sporulation in Bacillus subtilis. Biochemistry, 1998. **37**(5): p. 1365-75.
103. Tzeng, Y.L. and J.A. Hoch, Molecular recognition in signal transduction: the interaction surfaces of the Spo0F response regulator with its cognate phosphorelay proteins revealed by alanine scanning mutagenesis. J Mol Biol, 1997. **272**(2): p. 200-12.
104. Bird, T.H., et al., The Bacillus subtilis response regulator Spo0A stimulates transcription of the spoIIG operon through modification of RNA polymerase promoter complexes. J Mol Biol, 1996. **256**(3): p. 436-48.
105. Ireton, K., et al., Krebs cycle function is required for activation of the Spo0A transcription factor in Bacillus subtilis. Proc Natl Acad Sci U S A, 1995. **92**(7): p. 2845-9.

106. Asayama, M. and Y. Kobayashi, Signal transduction and sporulation in Bacillus subtilis: autophosphorylation of Spo0A, a sporulation initiation gene product. Mol Gen Genet, 1993. **238**(1-2): p. 138-44.
107. Healy, J., et al., Post-transcriptional control of a sporulation regulatory gene encoding transcription factor sigma H in Bacillus subtilis. Mol Microbiol, 1991. **5**(2): p. 477-87.
108. Weir, J., et al., Regulation of spo0H, a gene coding for the Bacillus subtilis sigma H factor. J Bacteriol, 1991. **173**(2): p. 521-9.
109. Strauch, M., et al., The SpoOA protein of Bacillus subtilis is a repressor of the abrB gene. Proc Natl Acad Sci U S A, 1990. **87**(5): p. 1801-5.
110. 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.
111. Gomer, R.H., Intercellular signalling. Knowing that you're among friends. Curr Biol, 1994. **4**(8): p. 734-5.
112. Dubnau, D., et al., Two-component regulators and genetic competence in Bacillus subtilis. Res Microbiol, 1994. **145**(5-6): p. 403-11.
113. Kunst, F., et al., The DegS/DegU and ComP/ComA two-component systems are part of a network controlling degradative enzyme synthesis and competence in Bacillus subtilis. Res Microbiol, 1994. **145**(5-6): p. 393-402.
114. Hahn, J., et al., Inactivation of mecA prevents recovery from the competent state and interferes with cell division and the partitioning of nucleoids in Bacillus subtilis. Mol Microbiol, 1995. **18**(4): p. 755-67.
115. Kruger, E., et al., The Bacillus subtilis clpC operon encodes DNA repair and competence proteins. Microbiology, 1997. **143**(Pt 4): p. 1309-16.
116. Lazazzera, B.A. and A.D. Grossman, *A regulatory switch involving a Clp ATPase.* Bioessays, 1997. **19**(6): p. 455-8.
117. Radany, E.H., et al., Transfection enhancement in Bacillus subtilis displays features of a novel DNA repair pathway. I: DNA base and nucleolytic specificity. Mutat Res, 1997. **384**(2): p. 107-20.
118. Rashid, M.H., A. Tamakoshi, and J. Sekiguchi, Effects of mecA and mecB (clpC) mutations on expression of sigD, which encodes an alternative sigma factor, and autolysin operons and on flagellin synthesis in Bacillus subtilis. J Bacteriol, 1996. **178**(16): p. 4861-9.
119. Redfield, R.J., Evolution of natural transformation: testing the DNA repair hypothesis in Bacillus subtilis and Haemophilus influenzae. Genetics, 1993. **133**(4): p. 755-61.
120. Redfield, R.J., Genes for breakfast: the have-your-cake-and-eat-it-too of bacterial transformation. J Hered, 1993. **84**(5): p. 400-4.
121. Turgay, K., et al., Biochemical characterization of a molecular switch involving the heat shock protein ClpC, which controls the activity of ComK, the competence transcription factor of Bacillus subtilis. Genes Dev, 1997. **11**(1): p. 119-28.
122. van Sinderen, D. and G. Venema, comK acts as an autoregulatory control switch in the signal transduction route to competence in Bacillus subtilis. J Bacteriol, 1994. **176**(18): p. 5762-70.
123. Wojciechowski, M.F., M.A. Hoelzer, and R.E. Michod, DNA repair and the evolution of

transformation in Bacillus subtilis. II. Role of inducible repair. Genetics, 1989. **121**(3): p. 411-22.
124. Dahl, M.K., et al., The phosphorylation state of the DegU response regulator acts as a molecular switch allowing either degradative enzyme synthesis or expression of genetic competence in Bacillus subtilis. J Biol Chem, 1992. **267**(20): p. 14509-14.
125. Hahn, J. and D. Dubnau, Growth stage signal transduction and the requirements for srfA induction in development of competence. J Bacteriol, 1991. **173**(22): p. 7275-82.
126. Raymond-Denise, A. and N. Guillen, Identification of dinR, a DNA damage-inducible regulator gene of Bacillus subtilis. J Bacteriol, 1991. **173**(22): p. 7084-91.
127. Shapiro, J.A., *The significances of bacterial colony patterns.* Bioessays, 1995. **17**(7): p. 597-607.
128. Mirel, D.B., V.M. Lustre, and M.J. Chamberlin, An operon of Bacillus subtilis motility genes transcribed by the sigma D form of RNA polymerase. J Bacteriol, 1992. **174**(13): p. 4197-204.

Appendix

APPENDIX: Twenty (20) collated sets of all appendix material must be submitted through the Contracts and Grants Office with other required forms so as to be postmarked no later than the proposal deadline. Identify each item with the name of the Principal Investigator. Do not use the appendix to circumvent the page limitations of the research plan. An application that does not observe these limitations may be returned.

not applicable