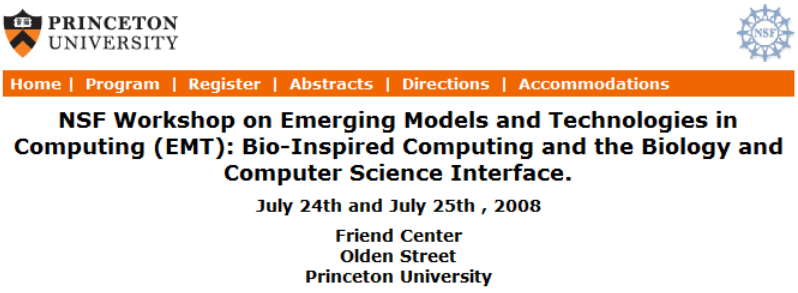


FINAL WORKSHOP REPORT

Sponsor: 004102 : Computing Research Association
Award Number: AGMT dtd 7-2-08

Title: NSF Workshop on Emerging Models and Technologies in Computing: Bio-Inspired Computing and the Biology and Computer Science Interface.

Report Prepared by:

Professor Ron Weiss
Departments of Electrical Engineering and Molecular Biology
Princeton University

Professor Laura Landweber
Departments of Ecology and Evolutionary Biology and Molecular Biology
Princeton University

Other Members of the Organizing Committee Include:

Professor Mona Singh
Departments of Computer Science and Molecular Biology
Princeton University

Professor Erik Winfree
Department of Computer Science
California Institute of Technology

Professor Mitra Basu
Department of Computer Science
Johns Hopkins University and the United States Naval Academy

Workshop Website: https://www.ee.princeton.edu/Events/emt/

Draft: 09/07/2008

Table of Contents

I. Executive Summary

I.1 Workshop Goals

The National Science Foundation, Division of Computer and Information Science and Engineering (CISE), Computing and Communications Foundations (CCF) sponsored a Workshop on Emerging Models and Technologies in Computing (EMT): Bio-Inspired Computing. This workshop brought together distinguished leaders in the fields of synthetic biology, bio-computing, systems biology, and protein and nucleic acid engineering to share their vision for science and research and to learn about the research projects that EMT has funded. The goal was to explore and to drive the growing interface between Biology and Computer Science. Discussions helped to define the opportunities and challenges for the future. In addition, current research efforts in the EMT Program were presented as oral and poster PI presentations.

Current PI's and leading researchers in the field considered the impacts of the EMT area at the interface between Biology and Computer Science on various fronts: research, technology transfer, undergraduate and graduate education, promotion of cross-disciplinary teaching and research activities, training of scientists, and societal benefits, such as encouraging enthusiasm among high school students and the general public.

An essential purpose of this workshop was to understand the perspectives and goals of the research community in order to provide NSF with information that will help them make a careful decision about the future of the EMT Program. A critical issue is whether to retain the EMT Program in its present form or to determine how the research it currently supports and the program's overall vision will fit into the re-clustering schemes within the CCF Division. A clear consensus emerged from the workshop discussions that several important grand challenges exist at the interface between Computer Science and Biology, and that the EMT program has been the cornerstone within CISE for funding such research. *The continuation of the EMT program or the creation of a new program to supplant its goals is therefore essential to fuel the growing and stimulating community at this research interface. Termination of the EMT program without suitable replacement will have a serious negative impact on CISE's role in supporting this unique niche between computer science and biology. Furthermore, because no other existing structure at NSF or NIH covers biologically-inspired computing and biological computation, the impact of loss of targeted support for this area that CISE has built up from scratch would be devastating.*

I.2 Summary of the Workshop Methodology and Findings

To achieve the workshop goals, the first day and a half consisted of presentations by world-renowned researchers not affiliated with the program, including Nobel laureate Eric Wieschaus who opened the program, and shorter presentations by EMT PI's. During the afternoon of the second day, workshop participants divided into five separate breakout groups, each of which had the task of with coming up with and

presenting three major grand challenges in their respective areas, including a general description, technical aspects, and broader impact. Each group also produced a written document after the workshop (contained in section III below). In addition, each group was asked to suggest a mechanism for community building and interdisciplinary growth.

The responses from the five working groups were extremely rich in variety, raising numerous critically important issues. Several grand challenges emerged (see below). Here we will provide a summary of the findings and recommendations. The individual group reports are also provided after this summary.

Grand challenges include:

1. To design biologically-inspired programs, models, hardware, and foundations for computing.

2. To achieve programmed control over molecules, cells, tissues, or organisms, using computer-science principles and interfacing with the external world.

3. To scale-up biological devices, systems, or models to greater levels of complexity.

4. To develop new tools and methods (experimental and computational) for observation, analysis and prediction of the behavior of biological systems from genomic or other sources of information.

In addition, an important recommendation that was also made more than once is the urgent need to nurture and to support the collaborative, interdisciplinary environment between computer scientists or engineers and biologists. Both ITR and EMT have done this very well in the past, fostering the growth of the current community of researchers that have excelled in many different areas. Funding for such interdisciplinary proposals should reflect the high cost of experimental research, to place emphasis on the validation of proposed models, devices or tools. Peer review panels need to bring together individuals with background in both computer science and biology, as only a specific EMT program can best provide. The community also repeatedly emphasized the need to provide hands-on biological training of individuals with computer science background, as well as computational training for biologists. This and other means to foster the growing interface of computer science and biology will strengthen its impact on society. On the other hand, without such deeply-committed levels and avenues of support as the EMT program has provided, this interdisciplinary field with tremendous potential for beneficial impact on society risks becoming or remaining a niche-discipline and also risks widening the gap between theory and experiments.

I.3 Recommendations

If CISE chooses not to retain the EMT Program in its present form, then we call for the creation of a new program that will provide a suitable home for the biologically-inspired side of computer science, a critical, multi-disciplinary research area that straddles the boundaries of individual research fields in biology, medicine, physics, chemistry, electrical engineering and computer science.

It is clear to us that the research EMT currently supports will become diluted or lost if it must be pigeon-holed into the current re-clustering schemes within CCF. Several important grand challenges exist at the interface between Computer Science and Biology (as described in the next section), and the EMT program (and before it ITR) has been the cornerstone within CISE for funding of such research. The continuation of this program or the creation of a new program to supplant its goals is therefore essential to fuel the growing and stimulating community at this research interface. Termination of the EMT program without suitable replacement will seriously deplete CISE's role in supporting this unique niche between computer science and biology. Furthermore, because no other existing structure at NSF or NIH covers biologically-inspired computing and biological computation, the impact of loss of targeted support for this area that CISE has built up from scratch would be highly destructive to this field.

If the EMT program is at its end, then we call for the creation of a new program. The grand challenges enumerated below and presented by the breakout groups can form the technical basis, rationale, and vision to support the creation of a new program within CISE.

II. Grand Challenges by Topic

II.1 Breakout sessions

The workshop attendees chose to participate in one of five breakout groups. Each group was led by two highly respected group leaders:

Group 1: Systems Biology:
Dana Pe'er, Columbia University
Charles Ofria, Michigan State University

Group 2: Neuro-based Computing:
Alice Parker, University of Southern California
Daniel Bullock, Boston University

Group 3: Biomimcry:
Chien-Chung Shen, University of Delaware
Radhika Nagpal, Harvard University

Group 4: Biocomputing and Self-Assembly:
Milan Stojanovic, Columbia University
Ashish Goel, Stanford University

Group 5: Synthetic Biology:
Christina Smolke, Caltech
Niles Pierce, Caltech

II.2 Individual Group Reports

Each breakout group prepared their own individual statement regarding the stated objectives of the Workshop. Below are the individual break-out group reports.

Group 1: Systems Biology

Technology has altered the face of biology generating terabytes of genomic, proteomic and microscopy data. The role of the computer scientist has primarily been that of the service to the biologist: developing algorithms to assemble genomes, normalize microarray data and segment microscopic images. The gargantuan abundance of data and the scale of the biological system requires a novel quantitative approach. The future of computational biology is science driven by the engineer, asking systems-wide biology questions that can only be from a perspective of computational and mathematic questions. The EMT program at NSF should continue to empower the computational scientist to determine biological questions of interest and to develop algorithmic and statistical tools to solve them. For successful impact on the biological community, the computational scientist must be given the means, funding and mechanism to experimentally validate their work. Funding mechanisms, prioritized by computational biologists, are needed. Below are some of the questions and research areas we feel most urgent and promising:

- **1. GENOMES: Steps from genomes to programs that control organisms**
  - Interpreting genomes and their functional elements
  - Functional element interrelations
  - Epigenetic programs and re-interpretation of genome information
  - Biological network inference
  - Prediction of gene expression states, cellular states, organismal phenotypes
  - Comparative genomics and evolution
- **2. DYNAMICS: Moving from static to dynamic (temporal, spatial, and evolution)**
  - From static models to dynamical models
    - Time and dynamics
    - Space and imaging
    - Evolution and evolvability
  - Signal processing and information flow in biological networks.
  - Predictive models of dynamic cellular behavior
  - Cellular decision making
  - Emergence of biological complexity from evolutionary processes
- **3. INTEGRATIVE MODELS: Methodology for integration across models and informing biological principles**
  - Meta-modeling and integrating across
    - data types (metabolic, regulatory, PPI)
    - scales/resolution (molecules, proteins, complexes, compartments, cells, organisms). Interconnections between layers
    - model organisms (e.g. integrating networks of yeast, flies, worms, human)
    - modeling approaches (e.g. probabilistic, kinetic, stoichiometric, biophysical)
  - Foundational frameworks and languages for integration
- **Community building and interdisciplinary growth: support for joint studies**
  - Joint support for computationally-driven wet-lab biological experiments
  - Interdisciplinary centers for experimental validation within universities.
  - Validation at many levels of resolution
  - Experimental training lab rotation for computational students

Group 2: Neuro-based Computing:

Neuro-based computing is a broad research area with a diverse set of challenges, but with significant payoff as the challenges are met. For example, neuro-based robotic vehicle vision systems could be invaluable, providing navigation capabilities that are robust in the presence of adverse environmental and lighting conditions. Facial recognition is a second application of neuro-based computing. Major progress has been made in modeling the brain mathematically, simulating small portions of the cortex and constructing biomimetic neural circuits. While progress is dramatic, significant challenges remain.

One could view the **problem space** spanned by neuro-based computing along several separate and distinct dimensions. The first dimension is the *level of focus and integration* in the research approach - from the micro- to meso- and macro-levels. This includes bottom up modeling and design of individual neurons and cellular mechanisms to abstraction and modeling of emergent group behavior of neural circuits at a physiologically defined intermediate scale to modeling, engineering and understanding the function of the cortex as a whole. The second dimension is the choice of selected *brain functions to emulate*, from perception, through intention, to actuation, culminating in task-responsive neuromorphic perception. The third dimension is the range of *building blocks* needed to construct mathematical models and simulation tools for portions of the brain, from mathematical models of microscopic cellular function and mesoscopic behavior, to databases containing inventories of components. These dimensions lead to three major challenges.

The **first Grand Challenge** is to build a brain-like device, or brain-inspired hardware and software with a large number of nodes ($>10^9$) and essential feedback loops – sufficient for emulation of complex tasks. To meet this challenge, researchers need to produce *in silico* biophysical mathematical models of neurons and assemblies (a parts inventory), as well as develop physical neural electronic circuits. It is important to provide means of functional integration between real (physiological) excitable cells and tissues, hardware renditions and *in silico* implementations of neurons and neural circuits for future hybrid devices, including prostheses.

The **second Grand Challenge** is to understand the mechanisms for integrating the micro- meso- and macroscopic levels of brain modeling. This involves identification and simulation of neurodynamic order parameters regulating neuron groups, determining the minimum (critical) size of populations required for generating an order parameter, and emulation of neuron network devices to provide multi-layered abstraction and generalization. An essential validating step in this challenge is to demonstrate functional agreement and equivalent behavior of neural networks built by bottom-up approach (from individual neurons) to integrated mesoscopic or macroscopic level mathematical models of the same.

The **third Grand Challenge** is to emulate brain function with the purpose of achieving adaptive behavior in autonomous robots or for more efficient biomimetic distributed computing for solving complex tasks. The emulation should contain perception behavior, should capture the essence of intention and attention, and the latter should feed back on how the perception is carried out in any given environment. The emulation should contain task-responsive neuro-robotic control, and should implement learning and memory at all levels of study. The outcomes of this challenge should not only help move towards recreating brain function but also should inspire and advance areas of computation with brain-like levels of parallelism and efficiency.

Group 3: Biomimcry:

Background. Biomimicry is to study nature's best ideas and then imitate these designs and processes to solve human problems. To some extent, several other subjects, such as neuron-based computing, bio-computing, synthetic biology, etc., could be considered part of biomimicry.

In the current state-of-the-art, there are ample examples of applying the practice of biomimicry to computer networks and robotics applications. For instance, the foraging behavior of ants inspired the metaphor of swarm intelligence, and motivated the development of the Ant colony optimization (ACO) meta-heuristics that can be used to find approximate solutions to difficult optimization problems, such as the traveling salesman problem (TSP). Recently, swarm intelligence had inspired several unicast and multicast routing protocols for mobile ad hoc and wireless sensor networks. Other examples include: applications of pattern formation and self-assembly to subjects of modular (cellular) robotics and smart materials; applications of swarming protocols (bird flocking, fish schooling, etc.) to networked robotic systems; application of artificial immune systems to computer and network security.

Grand challenges. Although current practices have been proven successful, greater challenges exist which could be categorized into two classes.

- **Foundations**: to develop sound design methodology that allows systematic imitation of nature's best designs and processes to solve human problems. Such methodology would include innovative programming paradigm and rigid mathematical theory. In particular, the design methodology should address the issue of translating global objectives to localized interaction rules so that the designed systems exhibit the desired self-* properties of self-configurable, self-organizing, self-healing, self-repairing, *etc*. The mathematical theory shall model the complexity of the designed systems to quantify their adaptability, scalability, and stability. The mathematical foundations developed will also have the broader impact of benefiting biology and eco-systems, *e.g.*, synchronization, swarming, *etc*.
- **Applications**: to design artificial systems that may work like and/or interact with biological systems. Example systems include *intelligent building* which exists like a human body or plant that regulates itself in terms of energy, temperature, humidity, pollution, stress integrity, *etc*.; *artificial ant colony* that lives with real ant colony in the garden to monitor and maintain the environment; self-assembled organs that live within a human body. The intrinsic features of these applications, such as being distributed, heterogeneous, concurrent, and asynchronous, challenge their design objectives of being robust, scalable, secure, cost-effective, and personal. Such applications shall have the broader impact of sustaining a healthier planet.

Mechanisms for community building and interdisciplinary growth. To confront these grand challenges, NSF should provide funding for conferences on biomimicry that attract both biologists and computer scientists, *e.g.*, IEEE Conference on Self-Adaptive and Self-Organizing Systems (SASO), and funding for exchange programs for students to study complementary discipline(s). NSF could also propose a specific Grand Challenge on Biomimicry (*e.g.*, intelligent building) so people could compete for sound solutions. However, most fundamentally, educational curricula (from K-12 through college to post-graduate) should be extended to promote the transfer of ideas inspired by nature to the design of our world, for a more sustainable, healthier planet.

Group 4: Biocomputing and Self-Assembly:

This field aims to: (1) solve outstanding computer science questions taking advantage of unique properties of molecules; (2) build and design self-assembled information-rich molecular devices, circuits, and systems, and study their interactions with living and non-living matter; (3) gain understanding of, build new, or re-engineer existing chemical and biological systems, using computer science principles; (4) build abstract models grounded in real word describing molecular behaviors that allow testable predictions of behaviors of complex molecular mixtures and enable abstract design and analysis of molecular systems.

We mention only a few example of progress achieved through CCF funding over the past funding cycle: programmed self-assembly of 2D and 3D objects, several developments of new molecular scale computing media, constructions of the first molecular functional systems and circuits, the first examples of molecular devices that could lead to molecular robotics with rich and complex behaviors, explanations of fundamental biological mechanisms, and development of theoretical models, analysis techniques, and algorithms. Building on this progress, we propose several grand challenges and new directions, bringing us closer to fulfilling some grand visions.

1. **Defining physical and chemical principles that would allow persistent increase in scale and complexity of molecular systems constructed by the bottom-up approach**. Issues of scale and complexity include: (i) Size of individual devices, interconnecting lithographic (top-down) and self-assembling (bottom up approaches) approaches. One possible milestone of a program could be achieving a reliable design of 2D asymmetric systems on 10-μm scale with an arbitrary positioning of elements with 5 nm resolution (possibly expanding current origami techniques); (ii) Complexity of systems, and increase in number of compatible individual devices in a single solution without compartmentalization, while developing efficient inter-compartment communications. One possible milestone could be the first demonstrations of adaptable or trainable molecular systems, or integration of up to 1000 molecular devices (improving on and incorporating the current molecular circuits and automata); (iii) Expanding principles used in molecular computing design to meso- and macro scales, and from 2D systems to 3D systems, with one possible milestone being algorithmic asymmetric 3D systems (building on the current 3D crystals).
2. **Interfacing molecular computing with the external world**. The challenge would be to develop multiple technologies that would allow reliable integration of multiple inputs, reliable sensing, reliable actuation using molecular devices, and reliable incorporation of the output of molecular scale devices into other technologies. This is a key step en route to practical applications, and includes connections to materials (eg. Silicon), electronic devices, enzymatic systems, and cells and tissues.
3. **New theoretical models, algorithms, analysis, and software design and simulation techniques**. It would be very useful to develop general models that allow us to reason about a range of molecular systems in a unified way. These models could be used to develop systems for design, analysis, and performance prediction of molecular systems. Some key goals would be to understand information processing in 3-D and to develop a general theory of robust systems spanning between the micro and molecular scale.
4. **Community building and broad impact**. Given the significance and the novelty of this field and its highly interdisciplinary nature, it is important to develop curricula and provide hands-on interdisciplinary opportunities for students at the undergraduate level or even earlier. It is also important to provide opportunities for chemists, biologists, and computer scientists to work in a collaborative fashion.

Group 5: Synthetic Biology:

Synthetic biology is an emerging research field in which programmable molecules are used to design, construct, and evolve novel regulatory, synthesis, and developmental circuitry for biotechnological applications and elucidation of biological principles. Synthetic biology has the potential to transform the scale and reliability with which we program biological function, addressing societal challenges in energy and food production, environmental quality, medicine, and global health. The interface between synthetic biology and computer science is critical to realizing this promise because many of the conceptual advances that will be required are centered on achieving robust and scalable information processing using synthetic biological circuits. Here we propose three grand research challenges and suggest mechanisms for fostering a research community at the intersection of computer science and synthetic biology.

Grand challenge #1: Circuit evolvability XOR robustness. Elucidate the principles of evolvable versus robust circuit architectures. Use this theoretical understanding to design and experimentally demonstrate synthetic network architectures that exhibit either evolvable or robust phenotypes under identical selection pressures and time constraints. For example, $E_1 \rightarrow E_1$ (robust under selective pressure) but $E_2 \rightarrow F$ (evolvable under same selective pressure); where E_1, E_2 = thermostat; F = phototaxis.

Grand challenge #2: Compilers and simulators for programming biological function. Develop theory, abstractions, models, CAD, programming languages, analysis and design algorithms, composition rules, and molecular algorithms for programming biological function. Compilers should accept as input, modular high-level specifications of the desired function, and provide as output, a molecular executable that implements the program. Compile and experimentally demonstrate a functional system with complexity arising both from the global architecture of the circuitry (requiring an understanding of scalability) and from the local functional complexity of one or more components (requiring atomically precise design of an active site, etc).

Grand challenge #3: Genome-based manufacturing of macroscale objects. Design and experimentally demonstrate synthetic developmental circuits that produce geometric complexity at the macroscale. For example, design the components, circuit architectures, and environment so that it is possible to 'grow' a heart or a shelter using synthetic circuits. This project raises the challenge of coupling developmental feedback between macroscale mechanics and molecular information processing (and vice versa).

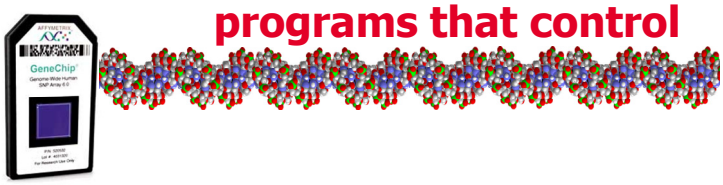
Building a community at the interface of Computer Science and Synthetic Biology. This interdisciplinary research area often mandates research combining theory and experiment. Funding levels should reflect the high cost of experimental research. Care should be taken so that the review process does not dismiss proposals for crossing disciplinary boundaries.

The synthetic biology community is doing a good job of engaging researchers at a young age through the iGEM competition; iGEM projects provide an ideal opportunity for targeted funding of students investigating computer science aspects of synthetic biology. Interdisciplinary hands-on bootcamps provide an intense and valuable mechanism for introducing computer science students to synthetic biology and synthetic biology students to computer science.

III. Breakout Group Presentations

Group 1: Systems Biology

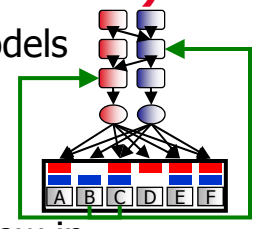
GENOMES: Steps from genomes to programs that control



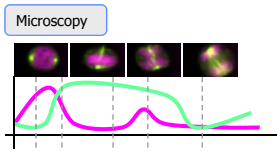
- Interpreting genomes and their functional elements
- Functional element interrelations
- Epigenetic programs and re-interpretation of genome information
- Biological network inference
- Prediction of gene expression states, cellular states, organismal phenotypes
- Comparative genomics and evolution

DYNAMICS: Moving from static to dynamic (temporal, spatial, and evolution)

- From static models to dynamical models
  - Time and dynamics
  - Space and imaging
  - Evolution and evolvability



- Signal processing and information flow in biological networks.
- Predictive models of dynamic cellular behavior
- Cellular decision making
- Emergence of biological complexity from evolutionary processes



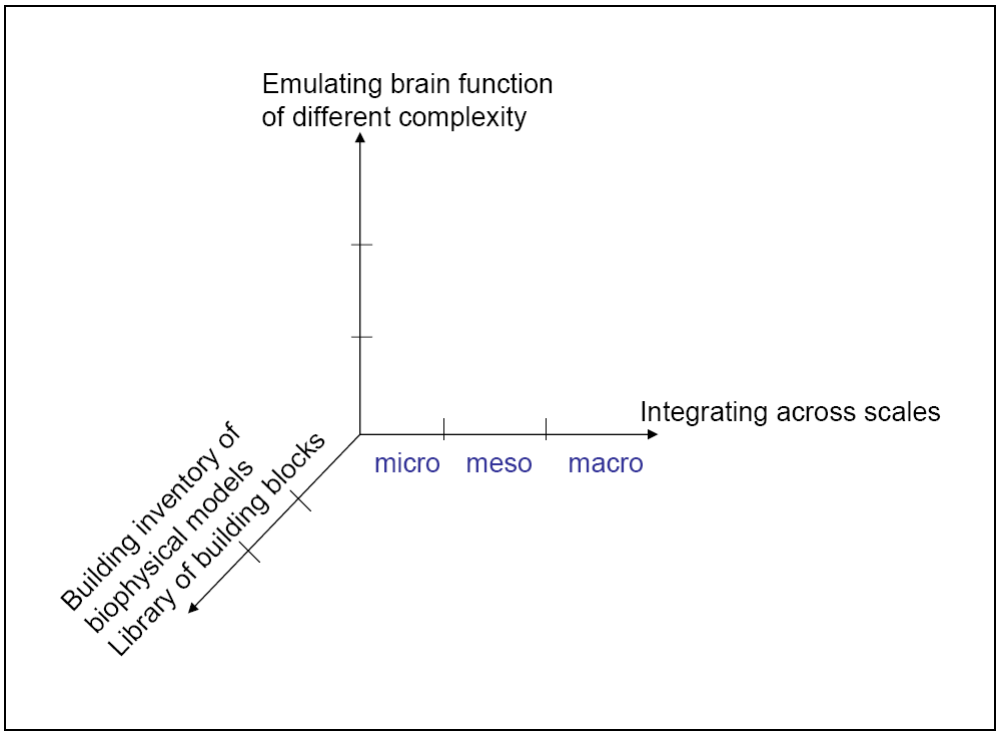
INTEGRATIVE MODELS: Methodology for integration across models and informing biological principles

- Meta-modeling and integrating across:
  - data types (metabolic, regulatory, PPI)
  - scales/resolution (molecules, proteins, complexes, compartments, cells, organisms). Interconnects between layers
  - model organisms (e.g. integrating networks of yeast, flies, worms, human)
  - modeling approaches (e.g. probabilistic, kinetic, stoichiometric, biophysical)
- Foundational frameworks and languages for integration
- Applying systems biology ideas to engineering
  - E.g. error-tolerant integration, reliable processing in a noisy environment, highly-distributed decision making
  - Ability to evolve as a design paradigm (precedent: genetic algorithms)

Community building and interdisciplinary growth: support for joint studies

- Joint support for computationally-driven wet-lab biological experiments
- Interdisciplinary centers for experimental validation within universities.
- Experimental training lab rotation for computational students

Group 2: Neuro-based Computing:



Grand Challenge # 1

Build a brain-like device

- Brain-inspired hardware with >10^9 nodes
- Biophysical models of neurons and assemblies (parts-inventory)
- Connecting neuron networks to mesoscopic behavior of the brain

* Bridge the gap to show that Micro-level networks implement Meso-level mathematical models

Grand Challenge # 2

Understand mechanisms for integrating
Micro- Meso- Macroscopic levels

- Understanding and simulating neurodynamic order parameters regulating neuron groups
- Determining the minimum size of populations required for generating an order parameter
- Emulation of the creation of network devices for multi-layered abstraction and generalization

Grand Challenge # 3

Emulate brain function that controls adaptive behavior in autonomous robots

- Perception in the context of Intention
- Task-responsive neuro-robotic control
- Implementation of learning and memory at all levels of dynamics

Methods for building community

- Funding for education of researchers in formulating their models at multiple levels
- Workshops for reconciling across disciplines – engineering, biology, mathematics, ethics
- Web community with shared data and software

Group 3: Biomimcry:

What is Biomimicry?

- **Bios** means life
- **Mimesis** means to imitate
- Biomimicry is to study nature's best ideas and then imitates these designs and processes to solve human problems
- To some extend, most other subjects (*e.g.,* neuro-based computing, bio-computing, synthetic biology, *etc.*) could be part of biomimicry!!!

State of the Art

- Examples of biomimicry practices in networking and robotics applications
  - Applications of **swarm intelligence** to network routing protocols
  - Applications of **pattern formation** and **self-assembly** to modular (cellular) robotics and smart materials
  - Applications of **swarming protocols** (flocking, schooling, etc.) to networked robotic systems
  - Application of **artificial immune system** to security systems
  - Early efforts of mathematical modeling and analysis

Grand Challenge #1 - Foundations

- General description: development of design methodology, programming paradigm and mathematical theory
- Technical aspects
  - Global to local translation
  - Mathematical modeling and analysis
  - Self-* properties (self-configurable, self-organizing, self-healing, self-repairing, etc.)
  - Adaptability
  - Complexity
  - Scalability
  - Stability
- Broader impact
  - Mathematical foundations will also benefit biology and eco-systems, *e.g.* synchronization, swarming, etc.

Grand Challenge #2 - Applications

- General description – how to design artificial systems that may work like and/or interact with biological systems. For instance,
  - **Intelligent building** (like human body or plants) that self-regulates itself, including energy, temperature, humidity, pollution, stress integrity, etc.
  - **Artificial ant colony** that lives with real ant colony in the garden to monitor and maintain the environment.
  - **Self-assembled organs** that live within human body.
- Technical aspects
  - Foundations and systematic study of biological systems
  - Intrinsic features - distributed, heterogeneous, concurrent, and asynchronous
  - Desired properties - robustness, scalability, security, cost-effectiveness, personalization
- Broader impact
  - Sustainable, healthier planet

Suggestions to NSF

- Mechanisms for **community building** and **interdisciplinary growth**
  - Funding for **conferences** on biomimicry that attract both biologists and computer scientists, *e.g.*, IEEE Conference on Self-Adaptive and Self-Organizing Systems (SASO)
  - Funding for **exchange programs** for students to study complementary discipline(s)
  - **NSF Grand Challenge** on Biomimicry (*e.g.*, intelligent building)
  - **Educational curricula** (K-12, college, and post-graduate) to promote the transfer of ideas inspired by nature to the design of our world, for a more sustainable, healthier planet.

Group 4: Biocomputing and Self-Assembly:

Breakout session

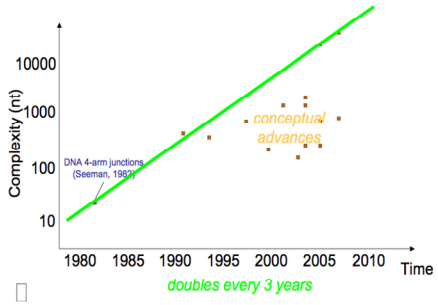
Self assembly and bio-computing

Achievements:

- Highly interdisciplinary
- Great progress!

Challenge: Scale and complexity

How do we continue the extraordinary gain in scale and complexity of molecular self-assembly and bio-computing over the last three decades?



•100nm ⇨10 μm scale:
Landmark for optical & lithographic access

•Increasing functional complexity

•Fault tolerance is key
•Conceptual: programming languages, design tools, systems design understanding
•Experimental: techniques for characterization and debugging

Challenge: Interfacing molecular computing

Towards applications: Inputs, Outputs, Sensors, Actuators

- With materials, substrates (lithographic scale would help!), and chemistries
- With hardware technology
- With chemical and biological systems
  - Therapeutic applications

- Considerable preliminary progress
  - Inductors
  - Decorated assemblies
  - Nano-wires

Challenge: Unified models and theories for molecular models

- The tile assembly model has served as a unifying paradigm
  - Theoretical and experimental advances
- Need similar models for
  - Bio-computing, Molecular robotics, and devices
- Information processing in 3-D
  - Combine geometry with combinatorics and dynamics
- Algorithms and mathematics for robustness
- Mathematical theories and techniques

Community building

- Involve undergraduates
- More avenues for computer scientists, chemists, and biologists to interact
  - Workshops and conferences
  - Funding
- Resources
  - Eg. Many simulators and sequence design codes already available
  - Experimental incubators?

Group 5: Synthetic Biology:

Computer Science in Synthetic Biology

A draft definition for synthetic biology:

Exploiting programmable molecules to design, construct, and evolve novel regulatory, synthesis, and developmental circuitry for technological applications and elucidation of biological principles.

Circuit evolvability XOR robustness

- Elucidate the principles of evolvable versus robust circuit architectures
- Experimental demonstrations of synthetic network architectures with evolvable versus robust phenotypes under identical selection pressures:

  E1 -> E1 (under selective pressure)

  E2 -> F (under same selective pressure)
- For example

  E1, E2 = thermostat; F = phototaxis

Compilers and simulators for programming biological function

- Theory, abstractions, models, CAD, programming languages, analysis and design algorithms, and molecular algorithms
- Exploit modularity, programmability, stochasticity
- Composition rules for scaling and for component complexity

Genome-based manufacturing of macroscale objects

- Experimental demonstrations of synthetic developmental circuits that produce geometric complexity
- The circuits should for example: grow a heart or a shelter
- Exploit developmental feedback between information and mechanics

Community Building

- Interdisciplinary hands-on bootcamps to introduce CS students to Synthetic Biology and vice versa (expensive due to experimental component)
- Support and expand CS component of iGEM
- Funding avenues for joint theory and experimental research (need for bigger budgets due to experimental components)
- Review process that doesn't dismiss such proposals for crossing disciplinary boundaries

Appendix A1 Agenda

Thursday, July 24, 2008

8:00	8:30	Registration and Breakfast			
8:30	9:00	Welcoming Remarks and Introductions			
		8:30	8:35	Michael Foster, Div. Director, CCF/CISE	
		8:35	8:45	Joanne Tornow, Div. Director of MCB, Bio Directorate	
		8:45	9:00	Pinaki Mazumder, Program Director, EMT/NSF	
9:00	12:30	Systems and Synthetic Biology			
		Chair: Laura Landweber, Ecology and Evolutionary Biology & Ron Weiss, Electrical Engineering & Molecular Biology, Princeton University			
		9:00	9:50	Eric Wieschaus, Princeton University: PLENARY TALK: Positional Information, Morphogen Gradients and Local Changes in Cell Shape During Embryonic Development in Drosophila: Opportunities for Quantitative Analyses and New Unanswered Questions	
		10:00	10:30	Drew Endy, MIT: On the Essentiality of Computer Science in Embracing, Managing, or Deleting Biological Complexity	
		10:30	11:00	BREAK	
		11:00	11:30	Kobi Benenson, Harvard University: Molecular automata: from concepts to applications	
		11:30	12:00	Christina Smolke, Caltech: Programming RNA Devices to Control Cellular Information Processing	
		12:00	12:30	Tony Forster, Vanderbilt University Medical School: Programmable Cell-free Replicating Systems	
12:30	1:30	LUNCH - Convocation Room, Friend Center			
1:30	3:20	Computational Systems and Evolutionary Biology			
		Chair: Mona Singh, Computer Science, Princeton University			
		1:30	2:20	David Haussler, UC Santa Cruz: PLENARY TALK: 100 Million Years of Evolutionary History of the Human Genome	
		2:20	2:50	Dana Pe'er, Columbia University: Genetic Regulatory Complexity: Lessons from Yeast and Cancer	
		2:50	3:20	Michal Zochowski, University of Michigan: Understanding Network Correlates of Neural Computation	
		3:20	3:40	BREAK	
3:40	5:50	Parallel PI Oral Presentations, Friend Center			
		Session I - Systems Biology (Friend Center Room 004)			
		Chair: Reka Albert, Penn State Univeristy			
			3:40	Wei Wang, University of North Carolina at Chapel	Hill Mining Patterns from Protein Structures
			3:55	Yaohang Li, North Carolina A&T State University	Multi-scoring Functions Sampling in Protein Loop Structure Prediction

4:10	Ying Xu, University of Georgia	Barcodes for Genomes and Applications
4:25	Samantha Kleinberg, Courant Institute of Mathematical Sciences, NYU	Systems Biology via Redescription and Ontologies
4:40	Eric Xing, Carnegie Mellon University	Nonparametric and Hierarchical Bayesian Methods for Genetic Interface
4:55	Reka Albert, Pennsylvania State University	Discrete Dynamic Modeling of Signal Transduction Networks
5:10	Jesus Izaguirre, University of Notre Dame	Regulation of Motility and Cellular Dynamics in Myxobacteria
5:25	Mona Singh, Princeton University	Function and Topology in Cellular Interaction Networks
Session II - Neuro-based Computing (Friend Center Room 006)		
Chair: Walter Freeman, University of California at Berkeley		
3:40	Jeff McKinstry, The Neurosciences Institute	Computation with Spikes
3:55	Bradley Hughes, University of California, Riverside	DNA Based Neural Networks
4:10	Robert Kozma, AFRL/RYHE Sensory Directorate Neuropercolation	Phase Transitions in Large-Scale Random Networks
4:25	Daniel Bullock, Boston University	Biomimetic Cortical Nanocircuits
4:40	Michael Erickson, University of California, Riverside	Biocomp: Biologically Inspired Computational Model for Perception
4:55	Alice Parker, University of Southern California	Neural Circuits and Computations in Active Vision
5:10	Walter J. Freeman, University of California at Berkeley	Emulation of Large-scale Systems Dynamics of Brains Using ODE and Random Graph Theory
5:25	Pradeep Shenoy, University of Washington	Brain-Computer Interfaces for Control and Computation
Session III - Biocomputing and Assembly (Friend Center Room 008)		
Chair: John Reif, Duke University		
3:40	Hongbin Yu, Arizona State University	Designed Spiral DNA Structures for Electronic Applications
3:55	Bernard Yurke, Boise State University	Connecting the Nanoworld to the Macroworld
4:10	Milan Stojanovic, Columbia University	Some Examples of EMT-sponsored Molecular Computing Projects
4:25	Evgeny Katz, Clarkson University	Signal-responsive Materials Integrated with Enzyme-logic Systems
4:40	Vladimir Privman, Clarkson University	Noise Reduction and Scalability in Biochemical Computing
4:55	Deborah Fygenson, UCSB	Silver Atom Clusters: Squence Dependent Fluorophores that Selfassemble on ssDNA
5:10	Ashish Goel, Stanford University	Design and Analysis of Molecular Algorithms
5:25	Laura Landweber, Princeton University	RNA-mediated Epigenetic Programming and Re-programming of Cellular DNA Rearrangements

6:30 - 8:30 Reception and PI Poster Presentations

Friday, July 25, 2008

8:00	8:30	Registration and Breakfast		
8:30	8:45	Welcoming Remarks - Sampath Kannan, Div. Director, CCF/CISE		
8:45	11:00	Computational Biology and Programmable Molecular Systems		
		Chair: Erik Winfree, Computer Science, Caltech		
		8:45	9:30	Luca Cardelli, Microsoft: PLENARY TALK: Molecules as Automata
		9:30	10:00	Niles Pierce, Caltech: Biomolecular Choreography
		10:00	10:30	Paul Rothemund, Caltech: Fabricating with Structural DNA Nanotechnology: The Next Steps
		10:30	11:00	Manolis Kellis, MIT: Regulatory Genomics of Drosophila and Mammalian Species
		11:00	11:15	BREAK
11:15	12:15	Parallel PI Oral Presentions, Friend Center		
		Session I - Systems Biology (Friend Center Room 004)		
		Chair: Animesh Ray, Keck Graduate Institute		
			11:15	Animesh Ray, Keck Graduate Institute — Mining Protein Networks for Synthetic Gene Interactions
			11:30	Cecilia Clementi, Rice University — Modeling Protein Dynamics at Multiple Resolution
			11:45	Itsik Pe'er, Columbia University — Identity by Descent between Purported Unrelateds
			12:00	Mohammed Zaki, RPI — A Mechanistic Model to Study the Effect of Topology on Protein Unfolding Pathways
		Session II - Modeling (Friend Center Room 006)		
		Chair: Liu Yang, Johns Hopkins University		
			11:15	Andrei Paun, Louisiana Tech University — Discrete Nondeterministic Modeling of Cellular Pathways
			11:30	Liu Yang, Johns Hopkins University — Modeling Cell Shape Changes Using Level Set Methods
			11:45	Mark Alber, University of Notre Dame — Computational Model of Swarming Bacteria
			12:00	Emilia Entcheva, Stony Brook University — Modeling Excitable Tissue with Hybrid Automata
		Session III - Biocomputing and Assembly (Friend Center Room 008)		
		Chair: Ned Seeman, New York University		
			11:15	Radhika Nagpal, Harvard University — Programmable Myriads: Self-assembly in Robotics and Epithelial Tissues
			11:30	John Reif, Duke University — Programmable DNA Nanodevices
			11:45	Nadrian Seeman, NYU — A Designed 3D Nucleic Acid Array
			12:00	Erik Winfree, Caltech — Molecular Programming: DNA Circuits and Self-assembly

12:15	1:15	LUNCH - Convocation Room, Friend Center
1:15	3:30	Parallel Breakout Sessions
		Systems Biology, CS Room 302 Chairs: Dana Pe'er, Columbia University and Charles Ofria, Michigan State University
		Neuro-based Computing, CS Room 301 Chairs: Alice Parker, University of Southern California and Daniel Bullock, Boston University
		Biomimcry, CS Room 401 Chairs: Chien--Chung Shen, University of Delaware and Radhika Nagpal, Harvard University
		Biocomputing and Self-Assembly, CS Room 402 Chairs: Milan Stojanovic, Columbia University and Ashish Goel, Stanford University
		Synthetic Biology, Deans Conference Room, Friend Center Chairs: Christina Smolke, Caltech and Niles Pierce, Caltech
3:30	3:45	BREAK
3:45	4:45	General Discussions of Final Document
4:45	5:00	NSF Summary: Future of EMT Program, Pinaki Mazumder, Program Director, EMT/NSF

Appendix A2. Participant List

Organizers

Professor Ron Weiss
Department of Electrical Engineering and Molecular Biology
Princeton University
rweiss@princeton.edu

Professor Laura Landweber
Department of Ecology and Evolutionary Biology
Princeton University
lfl@princeton.edu

Professor Mona Singh
Department of Computer Science
Princeton University
mona@cs.princeton.edu

Professor Erik Winfree
Deparment of Computer Science
Caltech
mailto:winfree@centrosome.dna.caltech.edu

Professor Mitra Basu
Department of Computer Science
Johns Hopkins University and the United States Naval Academy
mailto:basu@usna.edu

Participants from Academia

Mark Alber
University of Notre Dame
malber@nd.edu

Reka Albert
Pennsylvania State University
rza1@psu.edu

David Anderson
Georgia Institute of Technology
dva@ece.gatech.edu

Natalie Andrew
Harvard Medical School
Natalie_Andrew@hms.harvard.edu

Ernesto Andrianantoandro
Princeton University
eandrian@princeton.edu

Mitra Basu
United States Naval Academy
basu@usna.edu

Kobi Benenson
Harvard University
kbenenson@cgr.harvard.edu

Leonidas Bleris
Harvard University
lbleris@cgr.harvard.edu

Mark Borodovsky
Georgia Tech
borodovsky@gatech.edu

Daniel Bullock
Boston University
danb@cns.bu.edu

Marcus Calhoun-Lopez
The University of New Mexico
mcalhoun@cs.unm.edu

Yang Cao
Virginia Tech
ycao@cs.vt.edu

Zhixiong Chen
Mercy College
zchen@mercy.edu

Cecilia Clementi
Rice University
cecilia@rice.edu

Shuguang Cui
Texas A&M University
cui@ece.tamu.edu

Russell Deaton
University of Arkansas
rdeaton@uark.edu

Chris Dorey
Rutgers University
cbdorey@gmail.com

Rhonda Dzakpasu
Georgetown University
dzakpasu@physics.georgetown.edu

Drew Endy
MIT
endy@mit.edu

Emilia Entcheva
Stony Brook University
emilia.entcheva@sunysb.edu

Michael Erickson
University of California at Riverside
erickson@ucr.edu

Tony Forster
Vanderbilt University Medical School
a.forster@vanderbilt.edu

Walter J Freeman
University of California at Berkeley
dfreeman@berkeley.edu

Deborah Fygenson
UC Santa Barbara
deborah@physics.ucsb.edu

Jhenny Galan
University of the Sciences in Philadelphia
j.galan@usp.edu

Ashish Goel
Stanford University
ashishg@stanford.edu

Ralph Greenspan
The Neurosciences Institute
greenspan@nsi.edu

Radu Grosu
Stony Brook University
grosu@cs.sunysb.edu

David Haussler
UC Santa Cruz
haussler@soe.ucsc.edu

Danny Ho
University of Maryland Baltimore County
dho1@umbc.edu

Bradley Hughes
University of California at Riverside
bradley.hughes@email.ucr.edu

Jesus Izaguirre
University of Notre Dame
izaguirr@gmail.com

Natasha Jonoska
University of South Florida
jonoska@math.usf.edu

Evgeny Katz
Clarkson University
ekatz@clarkson.edu

Savas Kaya
Ohio University
kaya@ohio.edu

Manolis Kellis
MIT
manoli@mit.edu

Samantha Kleinberg
New York University
samantha@cs.nyu.edu

Rao Kosaraju
Johns Hopkins University
kosaraju@cs.jhu.edu

Robert Kozma
AFRL/RYHE Sensory Directorate
robert.kozma.ctr@hanscom.af.mil

Laura Landweber
Princeton University
lfl@princeton.edu

Daniel Lee
University of Pennsylvania
ddlee@seas.upenn.edu

Yaohang Li
North Carolina A&T State University
yaohang@ncat.edu

Zhiwei Liu
University of the Sciences in Philadelphia
z.liu@usp.edu

Ting Lu
Princeton University
tinglu@princeton.edu

James McCollum
Virginia Commonwealth University
james.m.mccollum@gmail.com

Jeff McKinstry
The Neurosciences Institute
jeffmckinstry@nsi.edu

Melanie Moses
University of New Mexico
melaniem@unm.edu

Chris Myers
Cornell University
crm17@cornell.edu

Radhika Nagpal
Harvard University
rad@eecs.harvard.edu

Rohit Nambisan
Department of Cognitive and Neural Systems
rohitnambisan@gmail.com

Charles Ofria
Michigan State University
ofria@msu.edu

Alice Parker
University of Southern California
parker@eve.usc.edu

Andrei Paun
Louisiana Tech University
apaun@latech.edu

Dana Pe'er
Columbia University
dpeer@biology.columbia.edu

Itsik Pe'er
Columbia University
itsik@cs.columbia.edu

Niles Pierce
Caltech
niles@caltech.edu

Christopher Poirier
Johns Hopkins University
cpoirier@jhu.edu

Joshua Porter
Johns Hopkins University
josh.porter@jhu.edu

Vladimir Privman
Clarkson University
privman@clarkson.edu

Cil Purnick
Princeton University
ppurnick@princeton.edu

Animesh Ray
Keck Graduate Institute
aray@kgi.edu

John Reif
Duke University
reif@cs.duke.edu

Fred Roberts
Rutgers University
froberts@dimacs.rutgers.edu

Paul Rothemund
Caltech
pwkr@dna.caltech.edu

Giovanni Samaey
K.U. Leuven
giovanni.samaey@cs.kuleuven.be

Herbert Sauro
University of Washington
hsauro@u.washington.edu

Nadrian Seeman
New York University
ned.seeman@nyu.edu

Chien-Chung Shen
University of Delaware
cshen@cis.udel.edu

Pradeep Shenoy
University of Washington
pshenoy@cs.washington.edu

Mona Singh
Princeton University
mona@cs.princeton.edu

Christina Smolke
Caltech
smolke@cheme.caltech.edu

Darko Stefanovic
University of New Mexico
darko@cs.unm.edu

Milan Stojanovic
Columbia University
mns18@columbia.edu

Wei Wang
University of North Carolina at Chapel Hill
weiwang@cs.unc.edu

Ron Weiss
Princeton University
rweiss@princeton.edu

Eric Wieschaus
Princeton University
efw@princeton.edu

Erik Winfree
Caltech
winfree@caltech.edu

Zhen Xie
Harvard University
zxie@cgr.harvard.edu

Eric Xing
Carnegie Mellon University
epxing@cs.cmu.edu

Ying Xu
University of Georgia
xyn@bmb.uga.edu

Liu Yang
Johns Hopkins University
liuyang@jhu.edu

Hongbin YU
Arizona State University
yuhb@asu.edu

Bernard Yurke
Boise State University
BernardYurke@boisestate.edu

Mohammed Zaki
RPI
zaki@cs.rpi.edu

Elena Zaslavsky
Princeton University
elenaz@princeton.edu

Yanqing Zhang
Georgia State University
yzhang@cs.gsu.edu

Michal Zochowski
University of Michigan
michalz@umich.edu

Participants from Federal Agencies

Sampath Kannan
National Science Foundation
kannan@cis.upenn.edu

Pinaki Mazumder
National Science Foundation
pmazumde@nsf.gov

Loverine Taylor
National Science Foundation
ltaylor@nsf.gov

Joanne Tornow
National Science Foundation
jtornow@nsf.gov

Tatsuya Suda
National Science Foundation
suda@ics.uci.edu

Participants from Industry

Jeff Stuart
Lockheed Martin
jastuart23@gmail.com

Robert Nachbar
Merck Research Laboratories
nachbar@merck.com

Luca Cardelli
Microsoft
luca@microsoft.com

Joseph Mannino
SiCortex, Inc.
jmannino@sicortex.com

Appendix A3. Abstracts of Invited Presentations

Programmable Cell-free Replicating Systems

Anthony Forster
Vanderbilt University Medical School

We have developed a simplified, purified, bacterial translation system to facilitate mechanistic studies and enable new applications. One envisioned application is programming evolution *in vitro* of small molecule ligands and drug candidates. Towards this goal, we have redesigned the genetic code for the synthesis and display of polymers containing unnatural amino acids. A long term goal is achieving a better understanding of biological replication by reconstituting it *in vitro*. Of the 151 genes we postulate to be necessary for self-replication from small molecules, protein synthesis constitutes 96%.

Biomolecular Choreography

Niles Pierce
Caltech

In nature, self-assembling and disassembling complexes of proteins and nucleic acids bound to a variety of ligands perform intricate and diverse dynamic functions. By contrast, attempts to rationally encode structure and function into synthetic amino acid and nucleic acid sequences have primarily focused on engineering molecules that self-assemble into prescribed target structures without explicit concern for transient system dynamics. To design systems that perform dynamic functions without human intervention, it is necessary to encode within the biopolymer sequences the reaction pathways by which self-assembly occurs. This talk will describe the use of mechanisms, abstractions and algorithms to program diverse nucleic acid self-assembly and disassembly pathways, yielding molecular executables implementing a variety of dynamic functions.

Fabricating with Structural DNA Nanotechnology: The Next Steps

Paul Rothemund
Caltech

Structural DNA nanotechnology has advanced greatly since its conception by Ned Seeman in the early 1980s. Much work over the last 25 years has focused on understanding how to use DNA as a geometric building block, what motifs are possible and how they can be composed to form larger structure. Now, two-dimensional crystals and small two-dimensional structures of arbitrary shape or pattern are routine; three-dimensional crystals and finite structures have been created and will similarly be routine. Work focused on DNA geometry will continue to provide interesting results for years to come (for example curved geometries have barely been considered) but there is now great interest in making the powerful DNA geometries we currently understand into a practical fabrication methodology. This will require, among other advances,

scaling up the complexity of current geometries, functionalization of DNA nanostructures with active devices, and integration of DNA nanostructures with conventional microfabrication. We will discuss our current efforts at addressing these challenges.

Programming RNA Devices to Control Cellular Information Processing

Christina Smolke
Caltech

The engineering of biological systems is critical to developing effective solutions to many societal challenges including energy and food production, environmental quality, and health and medicine. Modest levels of programmed cellular computation, logic, and control are needed to engineer complex biological systems. Recent progress has been made in the construction of RNA devices that process and transmit molecular input signals to regulated protein level outputs, linking computation and logic to gene expression patterns and thus cellular behavior. A first-generation composition framework that supports the programming of RNA devices exhibiting diverse device function from well-characterized components without complex device redesign will be described. The extensibility of this framework has been demonstrated for the implementation of higher-order cellular information processing operations. Coupled with technologies that enable the *de novo* generation of new RNA sensor components, RNA devices allow researchers to construct various user-programmed information processing operations in living systems and highlight the potential of synthetic biology strategies to support the rapid engineering of cellular behavior. The resulting improvements in our ability to transmit information to and from living systems, and implement control within cells themselves, will transform how we interact with and program biology.

Positional Information, Morphogen Gradients and Local Changes in Cell Shape During Embryonic Development in Drosophila

Eric Wieschaus
Princeton University

During embryonic development in Drosophila, cells are assigned to specific developmental fates based on the spatial distribution of gene products whose concentration varies continuously between neighboring cells. These initial concentration differences must be accurately read, and must then be translated into mechanical properties that govern specific cell behaviors at gastrulation. The input/output relationships that govern these developmental transitions are remarkably precise and provide an excellent opportunity to investigate how cells convert quantitative information into discrete developmental programs.

Part of our work has focused on the role of the Bicoid and Dorsal proteins in controlling anterior/posterior and dorsal ventral pattern. Using in vivo imaging techniques, we have measured nuclear concentrations of Bcd during late cleavage divisions and have and have correlated Bcd levels with transcription of its downstream targets. The extraordinary accuracy of gradient establishment, as well as the accuracy of the transcriptional readout we observe, raises

questions ranging from the scaling of morphogen gradients in the context of variable egg size and evolution, the effects of protein degradation on gradient profiles, the relationship between absolute Bcd concentration and Hb transcription and the physical limits to precision set by random arrivals of Bcd molecules at their target sites in the Hb promoter.

The information provided by Bicoid and Dorsal are translated into visible changes in cell shape through its effects on the cytoskeleton and cell adhesion. We have followed this transition in the ventral cells that respond to high concentrations of Dorsal protein. These cells accumulate high levels of apically localized Myosin II that undergoes periodic pulsating constrictions. We investigate genetic linkage between the positional information provided by Dorsal, the periodic pulsating behavior of MyosinII, and the ultimate change in cell shape associated with cell fate choice response.

Understanding Network Correlates of Neural Computation

Michal Zochowski
University of Michigan

The advent of new experimental techniques that allow for monitoring the activity of many neurons simultaneously provides hope for a clearer understanding of distributed dynamics involved in the brain computation. This at the same time requires formulation of novel tools for characterization of functional neuronal interactions during information processing. In this talk I will present results from the approaches undertaken in my laboratory to analyze and model neural activity observed during the memory consolidation processes in freely behaving animals. Here, we focus on structural and functional network mechanisms underlying hippocampal dynamics during various memory tasks. Our results point to a coherent picture of network modifications taking place in the hippocampus during these processes.

Appendix A4. Presentations

Provided directly to NSF representative.