

Shared Principles Between the Computing and Biological Sciences

Report on an NSF-sponsored Workshop

While the biological and computing sciences have a long history of exchanging ideas, traditional cross-disciplinary research has generally shed light on a problem of greater interest to one field than the other. Due to recent developments, it is now possible to envision attacking research problems of interest to both communities using principles shared by natural and fabricated systems.

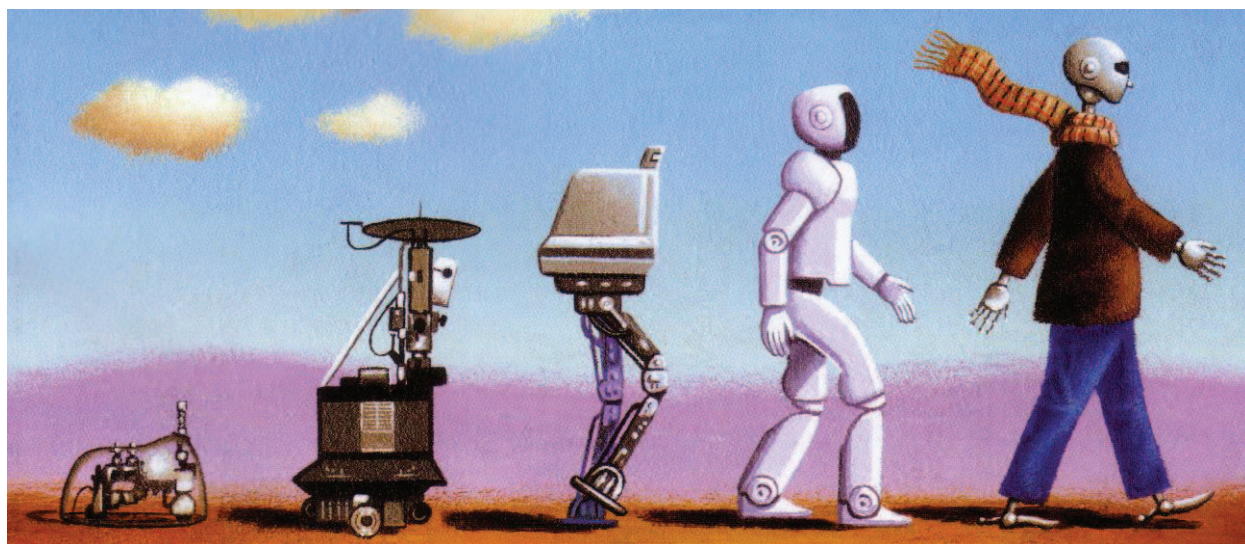


Illustration from the talk by Jeff Krichmar, "The Brain is Embodied and the Body is Embedded in the Environment."

As an important step towards ushering in a new era of research that transcends previously perceived boundaries between the computing and biological sciences, twenty-five leading investigators drawn from both fields met at a workshop held in Arlington Virginia on May 25-26, 2010. The current research of workshop participants embodies shared organizing principles and in many cases defies classification into one field or the other. This report presents the synthesis of ideas that emerged from the wide-ranging presentations and discussions.

I. Introduction

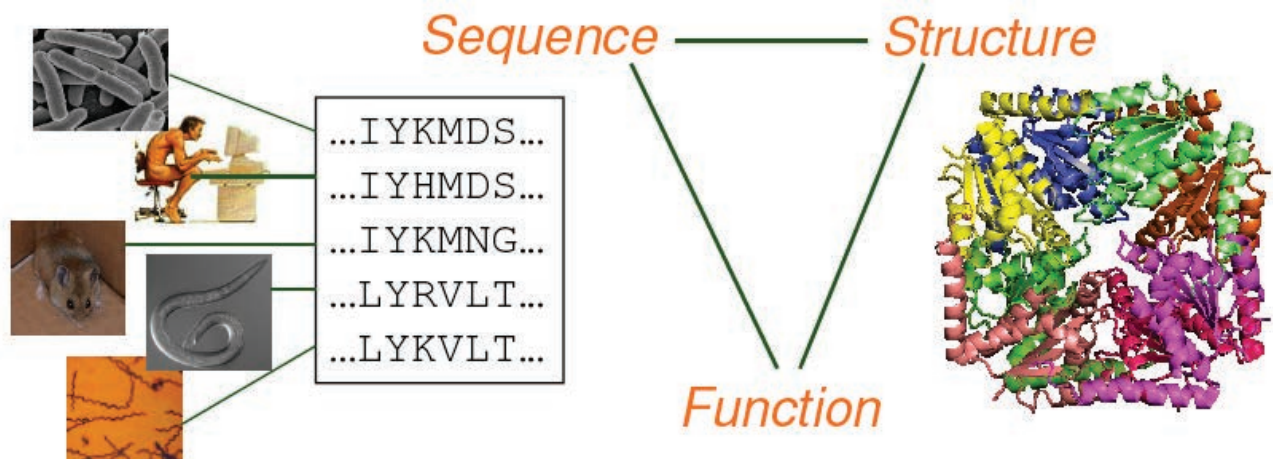
The purpose of this workshop was to bring together leading biologists and computational scientists interested in exploring shared principles between computing and biology. These two communities have a shared history of mutual influence where the transfer of concepts and tools has resulted in significant advances in both fields. Typically, the computing community develops techniques useful to the biological sciences such as mathematical models or search algorithms, while the biological sciences community develops hypotheses, which inspire new computational techniques such as swarm theory, genetic algorithms, and learning theories. This workshop was intended to take the discussion further, and ask to what extent there are larger organizing principles that are shared between computer science and biology, and whether the exploration of such shared principles might enable new and deeper collaborations between the two communities. The ultimate goal of the workshop was to yield of set of a and recommendations that will be helpful to NSF program officers in identifying new directions, designing new cross-disciplinary programs, and facilitating collaborations.

II. Major Themes of the Workshop

A. Information and Computation as a Framework for Advancing Theory in Biology

One of the key differences between computing and biology is the role of abstraction and the ability to define functionality independently of mechanism. In this realm, computing is permeated with abstract notions of functionality, whereas biology is almost entirely driven by empirical observations. Some examples of the extent to which computing is a "science of abstraction" include Boolean gates, formal languages, Chomsky hierarchy of machines/complexity, high-level programming languages, virtual machines, programmability, symbolic systems, syntax vs. semantics, compositionality, executability, and encapsulation.

Biology lacks a framework for a theory based on principles. The few theories that have made true contributions, most notably Darwin's theory of evolution by natural selection, the Mendelian theory of the gene or Burnet's theory of clonal selection in the immune system, and to a much lesser extent the contemporary theories of consciousness



Proteins: From sequence to structure to function. Illustration from the talk with the same title by Chris Bailey-Kellogg.

(e.g., Edelman's "neural Darwinism," Baars' and Changeux' "global workspace," Damasio's "core self-perception," or immune system, and to a much lesser extent Penrose's somewhat eccentric "quantum microtubule" theory), have all involved a mechanistic description which is then superseded, upon being proven, by the mechanistic description itself. That is, biological research is so heavily mechanism-driven that the theoretical aspect of these problems declines in significance in direct proportion to how much of the detailed mechanisms are worked out. Biology needs a language of "functionality" to describe the 'function' of the computation rather than the mechanics of biochemistry or electrophysiology (e.g., as exemplified in the work of Luca Cardelli).

Existing models for such a theoretical base include information theory and control theory (e.g., top-down and bottom-up modes of control), but care must be taken not to apply these approaches too literally. In other words, there is a major need for theory development.

B. Computing as a Source of Modeling Tools

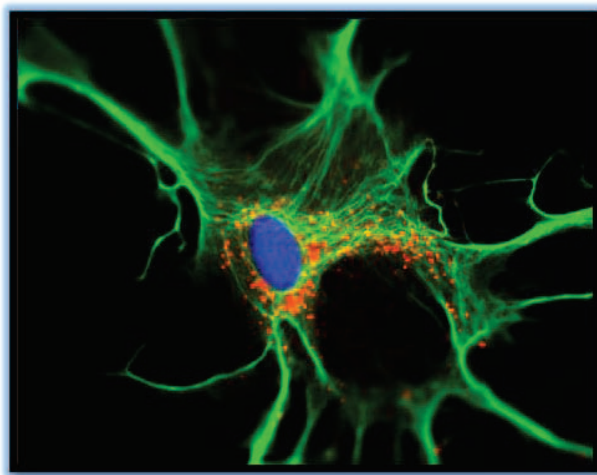
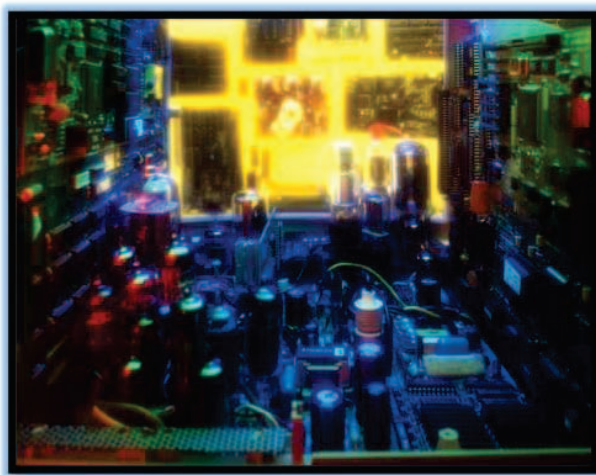
In some sense, the most fundamental shared principle between computing and biology is the need to deal with complex, dynamic, large-scale, asynchronous real-time systems. The science of computing has made huge progress in developing hardware and software tools for modeling, building, controlling, and analyzing such systems. Computer scientists can thus offer essential tools—both conceptual and practical—for building large-scale, high-performance models of biological systems. Many of the same methods developed for computational modeling and analysis in other fields (physics, chemistry, materials science, and so on) can be adapted to the needs of

biological modelers, who often lack well-designed modeling platforms at the systems level. Computational and information theoretical principles can also contribute to issues of coding, redundancy, error correction, and the design of experiments that arise in building and using very large-scale models.

C. Biology Versus Traditional Computer Science

As noted above, computer scientists have developed a number of new methods and algorithms that were inspired by biological systems, such as neural networks, genetic algorithms, artificial immune systems, ant-colony optimization and amorphous computing, among others. However, as was pointed out several times at the workshop, computer science has barely begun to mine the possibilities for inspiration that biology has to offer.

There exist vast differences between present-day computing systems and what we know about biological information processing. For example, current computing systems are typically centralized, localized, serial, deterministic, and digital/symbolic, whereas biological information processing is decentralized, distributed, massively parallel, stochastic, non-symbolic, and able to seamlessly switch between digital and analog modes. In biology, information is often encoded as rates or as concentrations, which must be physically sampled over time to be "read". Moreover, biological systems seamlessly integrate information over multiple spatial and temporal scales; no such capability exists in even the most sophisticated computing systems. In contrast to current computing systems, which require synchronization in many aspects of their processing, biological systems often operate with asynchronous components. Available computing systems also require components



Crosstalk and insulation: Computers versus the brain. Illustration from the talk by Hana El-Samad, "Design Principles for Cellular Organization."

to be reliable, with very low error probabilities, whereas biological systems operate with unreliable components that undergo frequent failures. Current computing systems require precise, finite calculations, whereas biological information processing thrives on approximation and open-endedness. In present-day computing systems there is typically a one-to-one relationship between “algorithm” and “function”, whereas biological systems are intensely *pleiotropic*, meaning that components are used and reused for multiple functions. Moreover, in biological systems, *function* is often defined by physical structure (e.g., binding between two molecules or connectivity structure in a network) instead of via a serial *algorithm*. As one additional example, biological systems automatically learn relevant features and associations from the environment. Learning and adaptation is integral to the system and is continual. In computer science, learning is an “add-on component,” if present at all.

It is still an important open question to what extent and in what situations these features of biological systems might be beneficial in the computing systems of the future, which many believe will need to become more adaptive, flexible, and “life-like,” and may

well be implemented with biomolecular or other molecular-scale “hardware”. This question can be resolved only via true collaborations between biologists and computational scientists.

III. Specific Topics for Collaborations

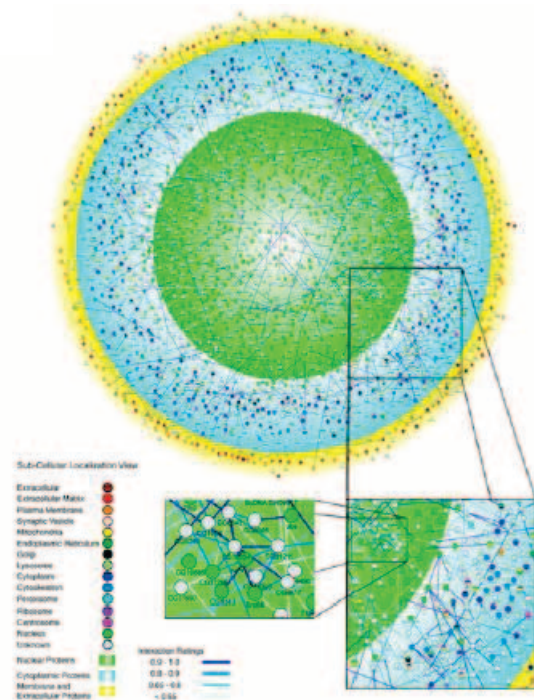
During the workshop, the participants were asked to discuss specific problems related to common principles that would benefit from collaborations between biological and computational scientists over the next five-to-ten years. The list below represents only a sampling of such topics. It is by no means exhaustive but is meant to spur further discussion and brainstorming.

- **Functions of stochasticity.** What are the sources and functions of stochasticity or “noise” in biological systems? There are myriad examples of stochasticity at all levels in living systems, such as stochastic expression of genes, noisy neural firing, generation of “random” immune cell receptor shapes, and unpredictable variability in heart rate, to name but a few examples. There have been numerous hypotheses concerning the sources and functions of such

stochasticity. Are there any unifying principles underlying these phenomena, including those that relate to information processing? Can research on the benefits of stochasticity and randomness in the field of computing be useful in understanding the functions of biological stochasticity?

- **Better understanding of network properties of biological systems, and connections between network structure and function.** There was much discussion at the workshop concerning the relevance of network science to the study of biological networks. A number of major questions were raised that would benefit from collaborations among computational scientists, theoretical network scientists, and biologists, including the following: What are the functionally crucial network properties of biological systems, e.g., structure and dynamics, hubs, etc.? Which ones matter most to system output, and which ones reflect different functional system states? To what extent is function grounded in network structure, and how does this function come about from underlying structure? How can structure (including modularity) in such networks be best uncovered from data? How does functional activity in a living system affect underlying network structure? What drives signaling in networks, and what network perturbations drive failures? What topological properties of signaling networks allow them to adapt—that is, reset themselves after responding to a stimulus? How do biological networks allow efficient navigation without global knowledge of the network structure?
- **Neural code.** How is information embodied and transferred in the nervous

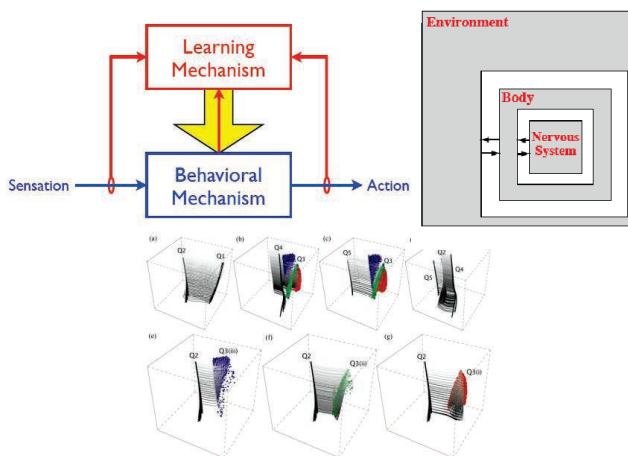
system? Is there a true neural code, in the sense that the activities of individual neurons or small groups of neurons can encode information useful to the system as a whole? If so, how would we go about decoding it? What and where are its coders and decoders? If not, then what are the alternatives (e.g., combinatorial interactions among brain regions)? Similarly, how and to what extent is the wiring of the brain embodied in genes?



Protein interaction networks. Illustration from the talk by Luis Amaral, "Networks and My Other Playgrounds."

- **Plasticity versus stability.** How can we understand the interactions and tradeoffs between the necessity for plasticity and the need for stability? How do biological systems maintain functional integrity while allowing basic components to continually change? Similarly, how do these systems maintain homeostasis in the face of the continual fluctuations with which they are confronted in daily life?

- Multiscale analysis.** Adaptation and integration occur over many scales, both spatial (e.g., molecules, cells, networks, organisms, ecosystems), as well as temporal (e.g., milliseconds to millennia). What is “the state” of an organism when you consider all spatial and temporal scales simultaneously? How can this be treated mathematically? A typical approach is to separate the scales, but this sacrifices many important properties. Is there a better way? Where are the joints that allow for spatial or temporal decomposition, given the broad distribution of activities? Can computer science generate relevant abstractions? What is the nature of the computations that are being done? Can we develop appropriate computational abstractions (e.g. formal languages) in 2-, 3-, and 4-dimensions?



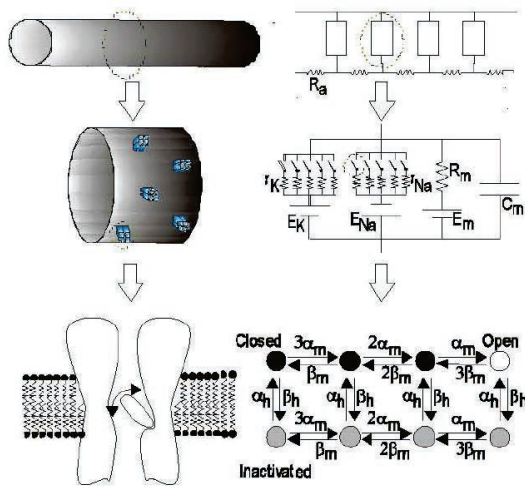
A Systems Perspective on Learning. Illustration from the talk by Randy Beer, "Dynamical Analysis."

- Building hybrid systems.** How can information in the brain be used by machine learning classifiers to create hybrid brain/computer systems? Do these require a full understanding of the hypothetical neural code, or can they be made to interpret correlations emanating from neural activity? Are such correlations a possible window

into the actual nature of neural embodiment itself?

- Towards strengthening the theoretical framework of biology.** The paucity of theories that have made robust contributions to biology begs the question of whether the field has any fundamental principles, or is it just a collection of bits and pieces? That is, does biology have enough underlying generality to permit the kind of abstraction and theoretical foundation that computer science relies upon? Is the biologist's preference for detailed mechanism a necessity or simply a product of the field's past history? Despite much discussion, the answers to these questions remain important challenges for future research.
- Flexible pattern recognition.** Flexible pattern recognition is a hallmark of biological systems. The most familiar instance is, of course, pattern recognition in the nervous system via the sensory modalities (vision, audition, etc.) but other biological systems exhibit sophisticated pattern recognition as well. For example, the immune system is able to recognize—in a decentralized and distributed way—patterns of anomalous activities that signal the presence of pathogens or other disorders that need to be dealt with. Ant colonies similarly can collectively sense changes in the environment that signal the need for reorganization of the colony's labor structure. Biological cells, genetic regulatory networks, networks of stomata in plants, and myriad other systems can likewise “perceive”, in a collective manner, spatio-temporal patterns that reflect the changing states of their environments and that allow them to respond. These phenomena

provide two major questions that beg for collaboration among biologists and computational scientists: first, are there shared principles or mechanisms for pattern recognition among all these different systems, and second, can such mechanisms be adapted to allow computers and computer networks to perform analogous kinds of adaptive pattern recognition?



The Stochastic Axon Simulator (from Faisal et al., Curr. Biol, 15, 1143-1149, 2005). . Illustration from the talk by Simon Laughlin, "The Design of Brains."

- **Improved artificial neural networks.**

Artificial neural networks (ANNs) are a pre-eminent biologically inspired method for automatic pattern recognition. However, to date, such networks have not succeeded in accomplishing anything close to human-level abilities for pattern recognition. Above, we listed a number of properties of biological systems that might be important in building more intelligent and life-like computer systems, but it is not clear which ones are the most important for improving ANNs and other biologically inspired algorithms. Can biologists give guidance on what is needed, without requiring ANN researchers to build a detailed brain?

- **Integrated and continual learning.**

As mentioned above, biological systems not only perform continual pattern recognition, they also automatically and continually learn. Unlike in current-day computing systems, learning and adaptation are integrated in living systems at all levels. How can we best develop computing systems and networks in which learning and adaptation are integral aspects of processing? Can we adapt biological principles to accomplish this? Stepping stones towards continual learning might include effective use of partial results and the building up of multifunctional representations. In the short term, this might require "satisficing," with details being integrated only over longer time scales.

- **Automated maintenance and repair.**

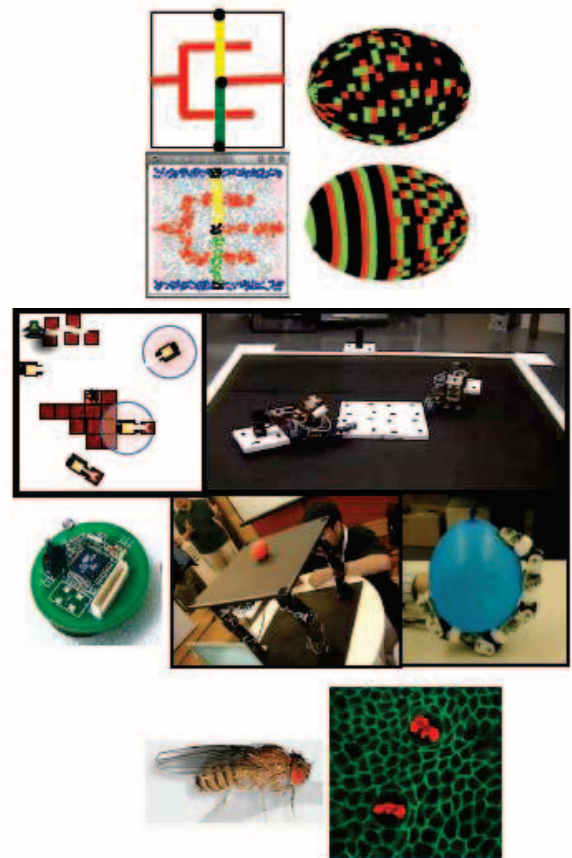
Related to the ability of continual adaptation and learning is the ability of biological systems to continually repair themselves to maintain health and integrity. This occurs at all scales, ranging from molecular DNA repair mechanisms to population effects (e.g., the collective maintenance and repair in social insect colonies). Living systems, moreover, do not "decommission" their components or "go offline" for repair, but rather rebuild parts while staying in operation. The notion of automated maintenance and repair in computing systems is an active topic of research, sometimes termed "autonomic computing". So far progress in this area has been limited, but there is likely great potential in mining biological principles for designing mechanisms for such processes.

- **Reliable performance from unreliable components and constant environmental variability.** Another

related topic is the remarkable ability of living systems to operate in a stable, reliable, and predictable way despite the unreliability and unpredictability of their underlying components and their environment. This theme again runs through biological systems at every scale, and is becoming increasingly important in computer science. As primitive computing components shrink to molecular scales, intrinsic unreliability, noise, and environmental fluctuations will require fundamentally new strategies for ensuring the reliability and predictability of systems. Collaborations between biological and computational scientists are likely to be a fertile source of new ideas for these issues.

- **Co-evolution and dealing with adaptive adversaries.** The problem of dealing with adversaries is at the forefront of computing research, as the perpetrators of spam, malware, network attacks, and other undesired and criminal activities become increasingly dynamic and able to adapt to countermeasures. This is an area in which biological inspiration might be particularly helpful, since dealing with adaptive adversaries is a constant way of life for living organisms. Several related biological ideas are already making their way into computer science, including host-parasite co-evolution and immune-system-inspired computer security. We see this as a very fruitful area for future cross-disciplinary collaborations.
- **Computing with biomolecules.** The field of synthetic biology is beginning to find principled ways to design computing systems with biomolecular elements and reactions. Can we develop more fully the analogs of high-

level data structures and programming languages, verification and model checking methods, and CAD tools for implementing computations with biomolecules? The vast differences between traditional computer science and biological systems—spelled out above in the section titled “Biology Versus Traditional Computer Science”—present enormous challenges to accomplishing what has been called “Executable Biology”, but there is already a small, dedicated community working on these questions, and the potential for much additional collaboration among computer scientists and biologists.



Self-organizing Systems Research. Illustration from the talk by Radhika Nagpal. "Collective Intelligence and Self-organizing Systems."

- **Towards a theoretical framework for biological computation.** The theories of information and computation have

provided mathematical frameworks for formalizing, designing, and reasoning about communication and information processing in human-created systems. These theories use concepts such as channel capacity and coding, formal languages, abstract machines, space and time complexity, and other mathematical notions. To what extent can these frameworks be adapted and extended to provide a theoretical underpinning for studying processes in biological systems? Such a theory might be one way in which the disparate details of different biological systems could be made sense of via common building blocks and mechanisms for information processing.

computer scientists can offer, and of what the conceptual foundations of computer science are. This is a consequence of the lack of background knowledge among biologists about areas outside of the traditional training in chemistry and elementary physics.

- There is a lack of interest on the part of computer scientists in understanding biological systems. Computer scientists want to build systems that do things; they are not typically very interested in actual biology, and are often not aware of the potential ideas it can offer to computer science, or the potential contributions computer scientists might make to the field of biology, beyond algorithms for bioinformatics.
- Biologists tend to want their models to be realistic, detailed, and complex. The culture in biology says "building a model means getting every detail right." A frequent quote from biologists to computer scientists: "We can't model this because we don't have enough data yet." This situation prevails despite the fact that most biological models serve mainly to confirm existing hypotheses, rather than to point to new directions or predictions.

IV. Potential Barriers for Collaboration and for Investigation of Shared Principles

The workshop participants identified the following issues as some of the potential barriers to effective collaboration between computer scientists and biologists on the questions listed above.

- There is a profound lack of understanding by biologists of what

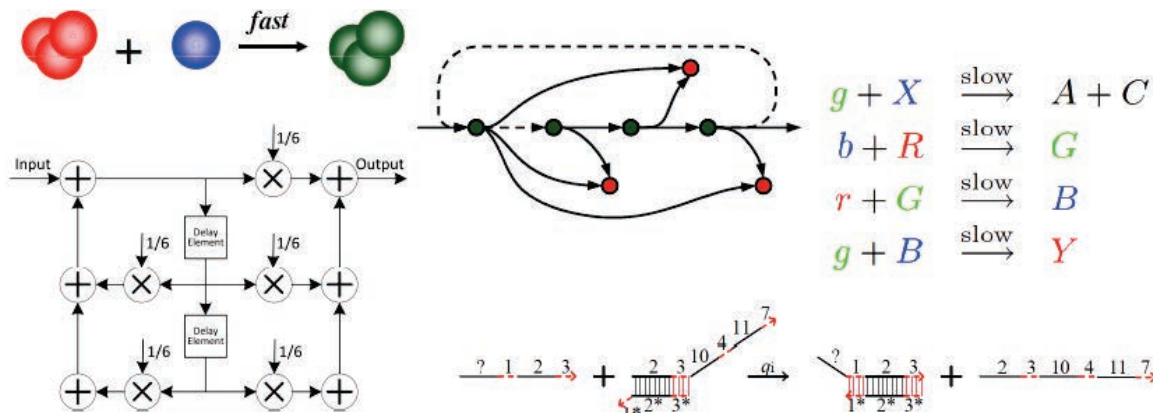


Illustration from the talk by Marc Riedel, "Robust Stochastic Computation with Biomolecular Reactions"

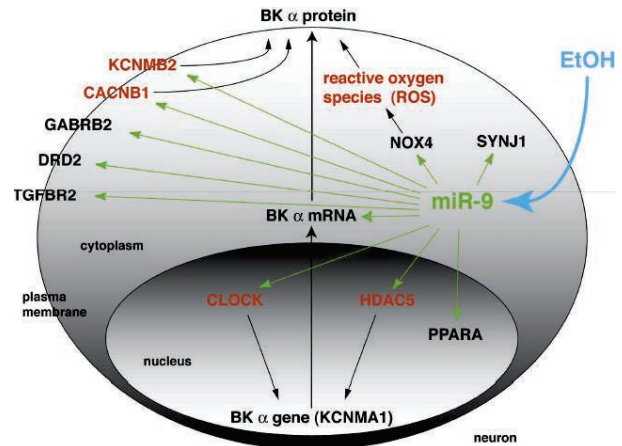
- Computer scientists are often too attached to traditional computational paradigms, such as logic gates or universal Turing machines, and shy away from notions of computation that they do not feel comfortable with.

V. Practical Considerations: Routes to Foster the Synthesis of Computing and Biological Science

In addition to the list of questions for potential collaborative research (above), the workshop participants discussed recommendations to facilitate and accelerate the development of synergy between the computing and biological sciences, with an emphasis on the training of young scientists in both fields.

- Participants unanimously and enthusiastically agreed that the development of one or more programs that specifically solicit proposals addressing shared principles between the computing and biological sciences is a high priority. Features of such programs would include a requirement (or at least strong encouragement) for explicit collaboration across these disciplines, along the lines of the questions described in this report. It is essential that the collaborations be reciprocal and intellectually equal; that is, the work should contribute to both computer science and biology.
- Biologists need to be better educated in the foundations of computer science, and vice versa. This is obviously a long-term effort and will occur over the next few generations of scientists. However, two short-term strategies to

facilitate progress are (1) the development of cross-disciplinary courses at the undergraduate and graduate levels; and (2) concrete mechanisms to promote the sharing of



miR 9 as a Master Switch. Illustration from the talk by Steve Treisman, "Molecular Plasticity."

graduate students and postdocs between different labs/research groups. Regarding the first approach, several participants pointed out that cross-disciplinary courses linking biology and computer science already exist at many institutions. Mechanisms for facilitating the sharing of course materials, syllabi, etc. would be very useful for faculty trying to develop such courses at their own institutions. Regarding the second approach, it was suggested that funding agencies could provide short (e.g., 6 month), fast track fellowships that would allow students and postdocs in biology or computation to briefly work in the lab of their mentors' collaborators in the other field. Such fellowships could be administered as supplements to existing cross-disciplinary grants.

VI. Workshop Participants

Organizers:

Ralph Greenspan Kavli Institute for Mind and Brain UC San Diego	Melanie Mitchell Computer Science Portland State University	Jo Ann Wise Molecular and Microbiology Case Western Reserve University
--	---	--

Participants:

Luis Amaral Chemical & Biological Engineering Northwestern University	Chris Bailey-Kellogg Computer Science Dartmouth College	John Byrne Neurobiology and Anatomy U. Texas Med. School, Houston
Ziv Bar-Joseph Computer Science Carnegie Mellon Univ.	Randy Beer School of Informatics Indiana University	Tony Bell Redwood Center for Theoretical Neuroscience UC Berkeley
Hana El-Samad Biochemistry UC San Francisco	Tom Henzinger Computer and Communication Sciences EPFL	Tom Kepler Biostat. & Bioinformatics Duke University
Jeffrey Krichmar Cognitive Sciences UC Irvine	Dmitri Krioukov CAIDA UC San Diego	Simon Laughlin Department of Zoology University of Cambridge
Eve Marder Biology Brandeis University	Olgica Milenkovic Electrical and Computer Engineering Univ. Illinois, Urbana-Champaign	Partha Mitra Cold Spring Harbor Laboratory
Radhika Nagpal Computer Science Harvard University	Ilya Nemenman Physics and Biology Emory University	Tomaso Poggio Brain and Cognitive Sciences MIT
Marc Riedel Electrical and Computer Engineering University of Minnesota	Ehud Shapiro Computer Science & Applied Math Biological Chemistry Weizman Institute of Science	Olaf Sporns Neuroscience Indiana University
Michael Simpson Oak Ridge National Laboratory		Steven Treisman Institute of Neurobiology University of Puerto Rico

Web site links for the organizers and participants as well as slides for most of the presentations can be found at the following URL:

<http://web.cecs.pdx.edu/~mm/SharedPrinciplesWorkshopMay2010/>

NSF Workshop on
Shared Organizing Principles
In the Computing and Biological Sciences

Program

Dates: Tuesday & Wednesday May 25th & 26th 2010

Location: Hilton Arlington, Arlington, VA.

DAY 1 (Tuesday May 25th)

- | | |
|-----------|--|
| 7:30–8:15 | Registration and continental breakfast |
| 8:15–8:30 | Introductions and welcome |
| | NSF CISE and BIO personnel |
| 8:30–8:40 | Opening Talk (1): Melanie Mitchell, Portland State University & Santa Fe Institute |
| 8:40–8:50 | Opening Talk (2): Ralph Greenspan, The Neurosciences Institute |

Session 1: Information Representation and Information Processing

- | | |
|-------------|---|
| 9:00–9:10 | Michael Simpson, ORNL |
| 9:10–9:20 | Ilya Nemenman, Emory University |
| 9:20–9:30 | Tom Kepler, Duke University |
| 9:30–9:40 | Ehud Shapiro, Weizmann Institute |
| 9:40–9:50 | Chris Bailey-Kellogg, Dartmouth |
| 9:50–10:00 | Olga Milenkovic, University of Illinois |
| 10:00–10:10 | Tony Bell, University of California, Berkeley |
| 10:10–10:20 | Tomaso Poggio, MIT |

10:20–10:40 Coffee Break

10:40–11:10 Group Discussion

Session 2: Networks and Communication

11:10–11:20 Luis Amaral, Northwestern University

11:20–11:30 Tom Henzinger, EPFL

11:30–11:40 Radhika Nagpal, Harvard University

11:40–11:50 Olaf Sporns, Indiana University

11:50–12:00 Dmitri Krioukov, University of California, San Diego

12:00–12:20 Group discussion

12:20–1:30 Lunch

Session 3: Synthetic Systems, Design, and Control

1:30–1:40 Ziv Bar-Joseph, Carnegie-Mellon University

1:40–1:50 Hana El-Samad, UC San Francisco

1:50–2:00 Marc Reidel, University of Minnesota

2:00–2:10 Simon Laughlin, University of Cambridge

2:10–2:20 Partha Mitra, Cold Spring Harbor Laboratory

2:20–2:40 Coffee break

Session 4: Learning and Adaptation

2:40–2:50 Steve Treisman, University of Puerto Rico

2:50–3:00 Eve Marder, Brandeis University

3:00–3:10 John Byrne, University of Texas

3:10–3:20 Jeffrey Krichmar, University of California, Irvine
3:20–3:30 Randy Beer, Indiana University
3:30–4:15 Group discussion
4:15–6:00 Break
6:00–?? Working dinner: Discussion in breakout groups

DAY 2 (Wednesday May 26th)

8:30–10:00 Breakout groups continue to meet, draft report
10:00–10:20 Coffee break
10:20–12:00 Breakout groups present reports to whole group; discussion
12:00–1:30 Lunch, adjourn
1:30–3:00 Workshop organizers finish report