SANDIA REPORT

SAND2004-6416
Unlimited Release
Printed January 2005

Biomolecular Decision-Making Processes for Self-Assembly: LDRD Final report

Gordon C. Osbourn

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865)576-8401
Facsimile: (865)576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800)553-6847
Facsimile: (703)605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2004-6416
Unlimited release
Printed January 2005

Biomolecular Decision-Making Processes for Self-assembly:

LDRD Final Report

Gordon C. Osbourn
Complex Systems Science
Sandia National Laboratory
P.O. Box 5800
Albuquerque, NM 87185-1423

Abstract

The brain is often identified with decision-making processes in the biological world. In fact, single cells, single macromolecules (proteins) and populations of molecules also make simple decisions. These decision processes are essential to survival and to the biological self-assembly and self-repair processes that we seek to emulate. How do these tiny systems make effective decisions? How do they make decisions in concert with a cooperative network of other molecules or cells? How can we emulate the decision-making behaviors of small-scale biological systems to program and self-assemble microsystems? This LDRD supported research to answer these questions. Our work included modeling and simulation of protein populations to help us understand, mimic, and categorize molecular decision-making mechanisms that non-equilibrium systems can exhibit. This work is an early step towards mimicking such nanoscale and microscale biomolecular decision-making processes in inorganic systems.

Acknowledgements

This work was sponsored by the U.S. Department of Energy under Contract DE-AC04-94AL85000. Sandia is a multiprogram laboratory operated by Sandia Corp., a Lockheed Martin Co., for the U.S. Department of Energy.

Contents

Proposal Background

Biological self-assembly can be viewed as a series of individual decisions that protein complexes and cells must make to form larger protein, cellular, and organism structures. For example, cells in the developing embryo of a multi-cellular organism must individually decide what specialized cell type they will take on (e.g., neural, epidermal, etc.) and when to make this commitment. They must do so in a way that is consistent with the decisions of other cells so that all cell types are generated in the correct numbers and locations. Recent biological studies are now revealing some of the specific molecular processes (e.g. Notch-Delta signaling) that enable such decisions to be made at the cellular level so that a multi-cellular organism can self-assemble correctly.

Another example is the formation of two-dimensional (2-D) maps in the brain, such as those that receive and process retinal inputs. Axons that will ultimately carry retinal response information must find their way through the brain to the target location for subsequent processing, and then must form synapses at terminal locations that preserve the 2-D visual patterns of the retinal receptors. Again, recent biological experiments are clarifying the chemical gradient signals that the axons use to navigate to the mapping site and to find their correct 2-D destination site among the millions of other axons that are simultaneously finding their destinations. We would like to understand a broad set of these remarkable decision (and self-assembly) processes from a scientific perspective, so that we can also learn to emulate them in smart, integrated nanosystems and microsystems.

There are several key questions that motivated this work. (1) How can tiny, essentially stochastic, molecular systems make decisions at all, and how can they do so in a reliable way? (2) How can these molecular systems make decisions in concert with a cooperative network of other molecules or cells, as typically occurs in biological systems? (3) How can we reproduce these decision-making behaviors in order to program and self-assemble microsystems?

Based on our recent self-assembly research, we saw evidence that many distinct biomolecular decision-making processes at many length scales in fact have certain elements in common. These processes exhibit similarities when viewed from a statistical mechanics point of view, in that they all require similar far-from-equilibrium system properties. Thus, we expected that our observation could be made scientifically rigorous by understanding the detailed statistical mechanics of these far-from-equilibrium systems. Further, we expected that the key statistical mechanics properties could also be exhibited by properly designed inorganic (nano and micro) systems. This work offers an unusual opportunity to elucidate a rare “systems” biology phenomenon -- a truly general description of these decision-making processes that will apply to many different types of far-from-equilibrium self-assembly systems at a variety of size scales.

Our goal in this one-year project was to develop a general statistical mechanics theory for these stochastic molecular decision-making processes. We carried out stochastic simulations of several model biomolecular decision processes. With these simulation and theory results in hand, our goal was to describe

how a variety of other distinct model biomolecular decision processes can also be understood, but without the need for detailed simulations in these other cases. We sought a new understanding of the general principles that underlie the many types of decision-making processes in non-equilibrium biomolecular systems, and to understand how to translate these principles into the design of certain classes of dynamically self-assembling structures.

Project Results

The decision-making properties of protein networks are more readily understood by viewing them as part of the larger information processing capabilities of such networks. Biochemical reactions taking place in living systems that map different inputs to specific outputs are intuitively recognized as performing information processing. Conventional wisdom distinguishes such proteins, whose primary function is to transfer and process information, from proteins that perform the vast majority of the construction, maintenance, and actuation tasks of the cell (assembling and disassembling macromolecular structures, producing movement, and synthesizing and degrading molecules). In this LDRD, we examined the computing and decision-making capabilities of biological processes in the context of the formal model of computing known as the random access machine (RAM), which is equivalent to a Turing machine (Minsky, 1967). When viewed from the novel RAM perspective, we were able to show that many of these dynamic self-assembly processes-synthesis, degradation, assembly, movement-do carry out computational operations and can implement decision-making processes (i.e., branching in the RAM model).

Branching/deciding between two alternative behaviors at the molecular scale often results from a competition between two populations, e.g., kinases and phosphatases that act on the same type of protein. Assume that when the target protein is phosphorylated it is activated as an enzyme for some reaction. In a deterministic sense, if there are more kinases than phosphatases for the target proteins, then the kinases will "win," and some target proteins will remain phosphorylated, so that they can catalyze their reaction. If there are more phosphatases, then they "win," and no target proteins stay activated long enough to perform their catalytic function. This represents an algorithm like "if $a > b$, then do x", where a represents the population of kinases, b the population of phosphatases, and "do x" is the reaction catalyzed by the target protein. However, because the kinase and phosphatase molecules diffuse around in the "soup" of the cytosol, reacting at random with the target proteins, there is a "race" between the two species to react with the target protein. Sometimes when there are fewer kinases than phosphatases, a phosphorylated target protein will have the opportunity to perform its catalytic function before it becomes dephosphorylated. Similarly, occasionally when there are more kinases than phosphatases, the target protein will not get a chance to catalyze a reaction before being dephosphorylated. Statistically, when $a >> b$, "do x" will almost always occur, and when $a << b$, "do x" will almost never occur, but when $a \sim b$, there is a smooth

sigmoidal transition where the probability of "doing *x*" goes from 0 to 1. Thus, the stochastic nature of such races leads to "errors" in computation, particularly when the race is close.

Decision-making or branching can be accomplished by a wide variety of molecular "hardware" configurations, including a set of only one or two molecules. The only requirement is two possible outcomes. For example, a new tubulin dimer could be added to the end of a microtubule (increment), or the last tubulin dimer might dissociate (decrement). A motor protein may take another step along the cytoskeletal fiber to which it is currently bound, or it might fall off. In both of these cases, there is a "race" between two stochastic events. There are two possible outcomes; the decision is made by whichever one happens first. In general, the decision-making/branching we observe in Nature involves some sort of race or competition. The decision is made to follow the "winner"—the faster, stronger, or more numerous.

To understand how living systems can function effectively using such non-deterministic decision processes, we also developed models of error-correction strategies (algorithms) employed by living systems. Error correction simulation code was developed, but we did not complete these simulations during the one year funding period. We also showed that the same RAM computing model is applicable at other hierarchical levels of biological systems (e.g., cellular or organism networks as well as molecular networks). We carried out stochastic simulations of idealized protein networks designed explicitly to carry out a numeric calculation. We also explored the reliability of such computations. Finally, we identified a representive set of real examples of dynamic self-assembly processes that occur in living systems, and describe the RAM computer programs they implement. The details of all of the simulations described above will be presented in a separate publication.

Conclusions

Our two major accomplishments in this work were: Discovering that, by viewing the non-equilibrium processes of living systems from the RAM model perspective, a far greater fraction of these processes can be understood as computing and decision-making processes than has been previously recognized by the scientific community; Achieving a deeper understanding of the common elements of diverse stochastic molecular decision-making proceses in biological systems. Both of these will prove essential to our quest to develop new classes of inorganic dynamic self-assembling systems.

Distribution:

1	MS-0123	Donna Chavez, LDRD office, 1011
2	MS-1423	G. C. Osbourn, 1001
2	MS-0899	Technical Library, 9616
1	MS-9018	Central Technical Files, 8945-1
1	MS-0161	Patent and Licensing Office, 11500