**Department of Energy/Basic Energy Sciences
Grant No. DE-FG02-01ER25498**

Final Report
September 2001 - August 2005

**Automated Physico-Chemical
Cell Model Development through
Information Theory**

Principal Investigator
Peter J. Ortoleva, Distinguished Professor

Center for Cell and Virus Theory
Department of Chemistry
Indiana University
Bloomington, Indiana 47405
Telephone (812) 855-2717
Fax (812) 855-8300
email ortoleva@indiana.edu
<http://biodynamics.indiana.edu>

OVERVIEW

The objective of this project was to develop predictive models of the chemical responses of microbial cells to variations in their surroundings. The application of these models is optimization of environmental remediation and energy-producing biotechnical processes.

The principles on which our project is based are as follows:

- chemical thermodynamics and kinetics;
- automation of calibration through information theory;
- integration of multiplex data (e.g. cDNA microarrays, NMR, proteomics), cell modeling, and bifurcation theory to overcome cellular complexity; and
- the use of multiplex data and information theory to calibrate and run an incomplete model.

In this report we review four papers summarizing key findings and a web-enabled, multiple module workflow we have implemented that consists of a set of interoperable systems biology computational modules.

THE KARYOTE WEB-BASED CELL MODELING SYSTEM

We have developed a web-enabled cell modeling system of interoperable modules (<http://biodynamics.indiana.edu>). Special features of this Karyote system include the following:

- Cell Assembler that allows a user to define cell geometry, a network of biochemical reactions, transmembrane molecular flux reactions, and passive or active transport, as well as specifying the associated rate laws; this module also allows a user to define arrays or cell suspension systems;
- Cell Model Unification that allows a user to integrate various cell models (e.g. to create a tissue or cell suspension or a eukaryotic cell made of existing models of a given organelle;
- edit or modify a pre-existing cell model;
- a tutorial with many cell model files; and
- multiple timescale numerical methods to simulate the biochemical kinetic and transport ordinary differential equations of compartmented cell dynamics.

THE TRANSCRIPTIONAL REGULATORY NETWORK DISCOVERY SYSTEM WEBSITE

In this Karyote website microarray data is analyzed to correct and expand a transcriptional regulatory network (TRN). In the workflow of this website the user enters cDNA microarray data and a preliminary TRN is provided or extracted from our database. In one module (FTF) a new correlation method is used that uses transcription factors to guide the evaluation of the TRN; the TRN is then provided as a list of explicit up/down regulation character for each gene and its regulating transcription factors. In a second module the output of FTF is used with time series or steady state microarray data to quantify transcriptional rate constants, transcription factor/gene binding constants, and RNA degradation rate coefficients. In follow-on work we are extending this system to allow for the use of various bioinformatics techniques (e.g. gene ontology, phylogenetic

similarity, and promoter analyses). This will provide the additional data needed to reliably construct the genome-wide TRN and simulate its response to gene insertion/deletion, or to perturbations in the extracellular medium.

Simulating Cellular Dynamics through a Coupled Transcription, Translation, Metabolic Model, E.L. Weitzke and P. Ortoleva, *Computational Biology and Chemistry* 27, 469–480, 2003.

In order to predict cell behavior in response to changes in its surroundings or to modifications of its genetic code, the dynamics of a cell are modeled using equations of metabolism, transport, transcription and translation implemented in the Karyote software. Our methodology accounts for the organelles of eukaryotes and the specialized zones in prokaryotes by dividing the volume of the cell into discrete compartments. Each compartment exchanges mass with others either through membrane transport or with a time delay effect associated with molecular migration. Metabolic and macromolecular reactions take place in user-specified compartments. Coupling among processes are accounted for and multiple scale techniques allow for the computation of processes that occur on a wide range of time scales. Our model is implemented to simulate the evolution of concentrations for a user-specifiable set of molecules and reactions that participate in cellular activity. The underlying equations integrate metabolic, transcription and translation reaction networks and provide a framework for simulating whole cells given a user-specified set of reactions. A rate equation formulation is used to simulate transcription from an input DNA sequence while the resulting mRNA is used via ribosome-mediated polymerization kinetics to accomplish translation. Feedback associated with the creation of species necessary for metabolism by the mRNA and protein synthesis modifies the rates of production of factors (e.g. nucleotides and amino acids) that affect the dynamics of transcription and translation. The concentrations of predicted proteins are compared with time series or steady state experiments. The expression and sequence of the predicted proteins are compared with experimental data via the construction of synthetic tryptic digests and associated mass spectra. We present the mathematical model showing the coupling of transcription, translation and metabolism in Karyote and illustrate some of its unique characteristics.

The Karyote Physico-Chemical Genomic, Proteomic, Metabolic Cell Modeling System, P. Ortoleva, E. Berry, Y. Brun, J. Fan, M. Fontus, K. Hubbard, K. Jaqaman, L. Jarymowycz, A. Navid, A. Sayyed-Ahamd, Z. Shreif, F. Stanley, K. Tuncay, E. Weitzke and L-S. Wu, *OMICS A Journal of Integrative Biology* Vol. 7, No. 3, 2003.

Modeling approaches to the dynamics of a living cell are presented that are strongly based on its underlying physical and chemical processes and its hierarchical spatio-temporal organization. Through the inclusion of a broad spectrum of processes and a rigorous analysis of the multiple scale nature of cellular dynamics, we are attempting to advance cell modeling and its applications. The presentation focuses on our cell modeling system, which integrates data archiving and quantitative physico-chemical modeling and information theory to provide a seamless approach to the modeling/data

analysis endeavor. Thereby the rapidly growing mess of genomic, proteomic, metabolic, and cell physiological data can be automatically used to develop and calibrate a predictive cell model. The discussion focuses on the Karyote cell modeling system and an introduction to the CellX and VirusX models. The Karyote software system integrates three elements: (1) a model-building and data archiving module that allows one to define a cell type to be modeled through its reaction network, structure, and transport processes as well as to choose the surrounding medium and other parameters of the phenomenon to be modeled; (2) a genomic, proteomic, metabolic cell simulator that solves the equations of metabolic reaction, transcription/translation polymerization and the exchange of molecules between parts of the cell and with the surrounding medium; and (3) an information theory module (ITM) that automates model calibration and development, and integrates a variety of data types with the cell dynamic computations. In Karyote, reactions may be fast (equilibrated) or slow (finite rate), and the special effects of enzymes and other minority species yielding steady-state cycles of arbitrary complexities are accounted for. These features of the dynamics are handled via rigorous multiple scale analysis. A user interface allows for an automated generation and solution of the equations of multiple timescale, compartmented dynamics. Karyote is based on a fixed intracellular structure. However, cell response to changes in the host medium, damage, development or transformation to abnormality can involve dramatic changes in intracellular structure. As this changes the nature of the cellular dynamics, a new model, CellX, is being developed based on the spatial distribution of concentration and other variables. This allows CellX to capture the self-organizing character of cellular behavior. The self-assembly of organelles, viruses, and other subcellular bodies is being addressed in a second new model, VirusX, that integrates molecular mechanics and continuum theory. VirusX is designed to study the in-and transmission of disease.

Toward Automated Cell Model Development through Information Theory, A.

Sayyed-Ahmad, K. Tuncay and P. Ortoleva, *Journal of Physical Chemistry A*, 107 (49); 10554-10565, 2003.

The objective of this paper is to present a methodology for developing and calibrating models of complex reaction/transport systems. In particular, the complex network of biochemical reaction/transport processes and their spatial organization make the development of a predictive model of a living cell a grand challenge for the 21st century. However, advances in reaction/transport modeling and the exponentially growing databases of genomic, proteomic, metabolic, and bioelectric data make cell modeling feasible, if these two elements can be automatically integrated in an unbiased fashion. In this paper, we present a procedure to integrate data with a new cell model, Karyote, that accounts for many of the physical processes needed to attain the goal of predictive modeling. Our integration methodology is based on the use of information theory. The model is integrated with a variety of types and qualities of experimental data using an objective error assessment approach. Data that can be used in this approach include NMR, spectroscopy, microscopy, and electric potentiometry. The approach is demonstrated on the well-studied *Trypanosoma brucei* system. A major obstacle for the development of a predictive cell model is that the complexity of these systems makes it unlikely that any model presently available will soon be complete in terms of the set of

processes accounted for. Thus, one is faced with the challenge of calibrating and running an incomplete model. We present a probability functional method that allows the integration of experimental data and soft information such as choice of error measure, *a priori* information, and physically motivated regularization to address the incompleteness challenge.

Simulated Complex Dynamics of Glycolysis in the Protozoan Parasite *Trypanosoma brucei*, A. Navid and P. Ortoleva, *Journal of Theoretical Biology* 228, 449–458, 2004.

Glycolysis in *Trypanosoma brucei* was modeled using a reaction transport simulator and tested for possible complex dynamics. The glycolytic model is multi-compartmentalized and accounts for the exchange of metabolites between the glycosomes, cytosol, mitochondrion and the host medium. The model is used to examine the effects of a range of culture medium concentrations of oxygen on the glycolysis of *T. brucei*. Our results are in good agreement with steady-state experiments. We also find that under aerobic conditions, increasing the activity of glycerol-3-phosphate dehydrogenase induces complex dynamics in the system. We report the presence of three distinct types of these dynamics. Varying the oxygen concentration in the medium can induce the transition between these dynamics.