Biological Systems as Reactive Systems

Luca Cardelli
Microsoft Research
Cambridge UK

2006-01-26 LMU Munich

www.luca.demon.co.uk
• **Genes are made of DNA**
  - Store digital information as sequences of 4 different nucleotides
  - Direct protein assembly through RNA and the Genetic Code

• **Proteins (>10000) are made of amino acids**
  - Process signals
  - Activate genes
  - Move materials
  - Catalyze reactions to produce substances
  - Control energy production and consumption

• **Bootstrap still a mystery**
  - DNA, RNA, proteins, membranes are today interdependent. Not clear who came first
  - Separation of tasks happened a long time ago
  - Not understood, not essential
Towards Systems Biology

- Biologists now understand many of the cellular components
  - A whole team of biologists will typically study a single protein for years
  - Reductionism: understand the components in order to understand the system

- But this has not led to understand how “the system” works
  - Behavior comes from complex patterns of interactions between components
  - Predictive biology and pharmacology still rare
  - Synthetic biology still unreliable

- New approach: try to understand “the system”
  - Experimentally: massive data gathering and data mining (e.g. Genome projects)
  - Conceptually: modeling and analyzing networks (i.e. interactions) of components

- What kind of a system?
  - Just beyond the basic chemistry of energy and materials processing...
  - Built right out of digital information (DNA)
  - Based on information processing for both survival and evolution
    - Highly concurrent

- Can we fix it when it breaks?
  - Really becomes: How is information structured and processed?
Storing Processes

- Today we represent, store, search, and analyze:
  - Gene sequence data
  - Protein structure data
  - Metabolic network data
  - Signaling pathway data
  - ...

- How can we represent, store, and analyze biological processes?
  - Scalable, precise, dynamic, highly structured, maintainable representations for systems biology.
  - Not just huge lists of chemical reactions or differential equations.

- In computing...
  - There are well-established scalable representations of dynamic reactive processes.
  - They look more or less like little, mathematically based, programming languages.

Cellular Abstractions: Cells as Computation
Regev&Shapiro NATURE vol 419, 2002-09-26, 343
Structural Architecture

Eukaryotic Cell

(10~100 trillion in human body)

Membranes everywhere

Nuclear membrane

Mitochondria

Golgi

Vesicles

E.R.

Plasma membrane (<10% of all membranes)

H. Lodish et al. Molecular Cell Biology fourth edition p.1
Abstract Machines of Systems Biology

Gene Machine
- Makes proteins
- Directs membrane construction and protein embedding
- Holds genome(s)
- Confinement
- Implements fusion, fission
- Holds receptors, actuators
- Hosts reactions

Protein Machine
- Aminoacids
- Metabolism, Propulsion
- Signal Processing
- Molecular Transport

Membrane Machine
- Phospholipids
- Confinement
- Storage
- Bulk Transport

Biochemical Networks
- Nucleotides
- Models integration, different time and space scales
- Holds genome(s), confines regulators
- Implements fusion, fission
- Holds receptors, actuators
- Hosts reactions

Transport Networks
- Aminoacids
- Proteins
- Membrane

Model Integration
- Different time and space scales
- Signaling
- Storage
- Membrane

Regulation
- Gene
- Regulatory networks
- Signaling
- Storage
- Membrane

Surfaces and Extracellular Features
- Glycans
- Trees
- Hierarchical multisets

Signaling
- Strings
- Records
- Metabolism, Propulsion
- Signal Processing
- Molecular Transport

Different time and space scales
- Nucleotides
- Protein	
- Membrane

Abstract Machines of Systems Biology

Gene Machine
- Makes proteins, where/when/how much
- Directs membrane construction and protein embedding
- Holds genome(s), confines regulators
- Implements fusion, fission
- Holds receptors, actuators
- Hosts reactions

Protein Machine
- Aminoacids
- Metabolism, Propulsion
- Signal Processing
- Molecular Transport

Membrane Machine
- Phospholipids
- Confinement
- Storage
- Bulk Transport

Biochemical Networks
- Nucleotides
- Models integration, different time and space scales
- Holds genome(s), confines regulators
- Implements fusion, fission
- Holds receptors, actuators
- Hosts reactions

Transport Networks
- Aminoacids
- Proteins
- Membrane

Model Integration
- Different time and space scales
- Signaling
- Storage
- Membrane

Regulation
- Gene
- Regulatory networks
- Signaling
- Storage
- Membrane

Surfaces and Extracellular Features
- Glycans
- Trees
- Hierarchical multisets

Signaling
- Strings
- Records
- Metabolism, Propulsion
- Signal Processing
- Molecular Transport

Different time and space scales
- Nucleotides
- Protein	
- Membrane
Reactive Systems

- **Modeling biological systems**
  - Not as continuous systems (often highly nonlinear)
  - But as discrete *reactive systems*; abstract machines where:
    - **States** represent situations
    - Event-driven **transitions** between states represent dynamics
  - The adequacy of describing (discrete) complex systems as reactive systems has been argued convincingly [Harel]

- **Many biological systems exhibit features of reactive systems:**
  - Discrete transitions between states
  - Deep layering of abstractions ("steps" at multiple levels)
  - Complexity from combinatorial interaction of simple components
  - High degree of concurrency and nondeterminism
  - "Emergent behavior" not obvious from part list
**Chemistry vs. **π**-calculus**

**A process calculus (chemistry)**

\[ r: A + B \xrightarrow{k_1} C + D \]
\[ s: C + D \xrightarrow{k_2} A + B \]

- Reaction oriented
- 1 line per reaction

**A different process calculus (π)**

\[ A = !r_{k_1} \cdot C \]
\[ C = ?s_{k_1} \cdot A \]
\[ B = ?r_{k_1} \cdot D \]
\[ D = !s_{k_2} \cdot B \]

- Interaction oriented
- 1 line per component

A Petri-Net-like representation. Precise and dynamic, but not modular, scalable, or maintainable.

A compositional graphical representation (precise, dynamic and modular) and the corresponding calculus.
1. The Protein Machine

- **Complex folded-up shapes that:**
  - Fit together, dock, undock.
  - Excite/unexcite, warp each other.
  - Bring together, catalyze, transform materials.
  - Form complex aggregates and networks.

- **Mapping out such networks:**
  - In principle, it’s “just” a very large set of chemical equations.
  - Notations have been developed to summarize and abstract.

An actual molecular interaction network.
(Nodes are distinct protein kinds, arcs mean that two kinds of proteins interact.)
Protein Structure

Primary

The 20 Aminoacids

Tryptophan

Secondary

Alpha Helix, Beta Sheet

Tertiary

Green Fluorescent Protein

Quaternary

Triose Phosphate Isomerase

http://www.cmbi.kun.nl/gvteach/bioinformatica1/
Protein Function

- Regulation
- Degradation
- Metabolism
- Movement
- Assembly
- Transport
- Structure
- Signalling

Taken from the web?
**MIM: Molecular Interaction Maps (Kohn)**

- The double-arrowed line indicates that proteins A and B can bind to each other. The "node" placed on the line represents the A:B complex.
- Asymmetric binding where protein A donates a peptide that binds to a receptor site or pocket on protein B.
- Representation of multimolecular complexes: x is A:B; y is (A:B):C. This notation is extendible to any number of components in a complex.
- Covalent modification of protein A. The single-arrowed line indicates that A can exist in a phosphorylated state. The node represents the phosphorylated species.
- Cleavage of a covalent bond: dephosphorylation of A by a phosphatase.
- Proteolytic cleavage at a specific site within a protein.

Stoichiometric conversion of A into B.

Transport of A from cytosol to nucleus. The node represents A after it has been transported into the nucleus.

Formation of a homodimer. Filled circle on the right represents another copy of A. The node on the line represents the homodimer A:A.

z is the combination of states defined by x and y.

Enzymatic stimulation of a reaction.

General symbol for stimulation. A bar behind the arrowhead signifies necessity.

General symbol for inhibition.

Shorthand symbol for transcriptional activation.

Shorthand symbol for transcriptional inhibition.

Degradation products.
Molecular Interaction Maps

The p53-Mdm2 and DNA Repair Regulatory Network


Figure 6B: The p53-Mdm2 and DNA repair regulatory network (version 2) - May 19, 1999.
Each protein has a structure of binary switches and binding sites. But not all may be always accessible.

Switching of accessible switches.
- May cause other switches and binding sites to become (in)accessible.
- May be triggered or inhibited by nearby specific proteins in specific states.

Binding on accessible sites.
- May cause other switches and binding sites to become (in)accessible.
- May be triggered or inhibited by nearby specific proteins in specific states.
Notations for the Protein Machine

- **Stochastic π-Calculus**
  - Priami (following Hillston’s PEPA) formalizes a stochastic version of π-calculus where channels have communication rates.

- **BioSPI**
  - Regev-Shapiro-Silverman propose modeling chemical interactions (exchange of electrons and small molecules) as “communication”.
  - Standard stochastic simulation algorithms (Gillespie) can be used to run in-silico experiments.
  - Complex formation is encoded via p-restriction.

- **PEPA**
  - Calder Gilmore and Hillston model the ERK pathway.

- **k-calculus**
  - Danos and Laneve (following Kitano’s BioCalculus) define a calculus where complex formation is primitive.

- **(Stochastic) Petri Nets**
  - S.Reddy’94 modeling pathways.
  - Srivastava Perterson and Bentley analyze and simulate E.coli stress response circuit.

- **Bio State Charts**
  - Harel uses State Charts to model biological interactions via a semi-graphical FSM notation.

- **Pathway Logic**
  - Talcott-Eker-Knapp-Lincoln use term-rewriting.

- **BioCham**
  - ChabrierRivier-Fages-Soliman use term-rewriting and CLT modelchecking.

- **Kohn Diagrams, Kitano Diagrams**

- **SBML (Systems Biology Markup Language)**
  - XML dialect for MIM’s:
    - Compartments (statically nested)
    - Reagents with concentrations
    - Reactions with various rate laws
  - Read and written by many tools via the Systems Biology Workbench protocol

### Table 2. Predicted Hill coefficients for MAP kinase cascade components: Varying the assumed $K_m$ values

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Range of assumed $K_m$ values</th>
<th>Range of effective Hill coefficients (mM) predicted for</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. MAPK -&gt; MAPKK</td>
<td>60-1500 nM</td>
<td>MAPKK: 1.0, MAPKK*: 1.7, MAPK: 4.9</td>
</tr>
<tr>
<td>2. MAPKK* -&gt; MAPKK</td>
<td>60-1500 nM</td>
<td>MAPKK: 1.0, MAPKK*: 1.7, MAPK: 4.9</td>
</tr>
<tr>
<td>3. MAPK -&gt; MAPKK-P</td>
<td>60-1500 nM</td>
<td>MAPKK: 1.0, MAPKK-P: 1.3-2.3, MAPK: 4.0-5.1</td>
</tr>
<tr>
<td>4. MAPKK-P -&gt; MAPKK</td>
<td>60-1500 nM</td>
<td>MAPKK: 1.0, MAPKK-P: 1.5-1.9, MAPK: 3.6-6.7</td>
</tr>
<tr>
<td>5. MAPKK-P -&gt; MAPKK-PP</td>
<td>60-1500 nM</td>
<td>MAPKK: 1.0, MAPKK-P: 1.2-2.4, MAPK: 3.2-5.2</td>
</tr>
<tr>
<td>6. MAPKK-PP -&gt; MAPKK-P</td>
<td>60-1500 nM</td>
<td>MAPKK: 1.0, MAPKK-P: 1.7-1.8, MAPK: 4.1-6.4</td>
</tr>
<tr>
<td>7. MAPK -&gt; MAPKK-P</td>
<td>60-1500 nM</td>
<td>MAPKK: 1.0, MAPKK-P: 1.7, MAPK: 3.7-6.2</td>
</tr>
<tr>
<td>8. MAPK -&gt; MAPKK-PP</td>
<td>60-1500 nM</td>
<td>MAPKK: 1.0, MAPKK-PP: 1.7, MAPK: 4.3-5.2</td>
</tr>
<tr>
<td>9. MAPK-P -&gt; MAPKK-P</td>
<td>60-1500 nM</td>
<td>MAPKK: 1.0, MAPKK-P: 1.7, MAPK: 3.4-6.1</td>
</tr>
<tr>
<td>10. MAPK-P -&gt; MAPKK-PP</td>
<td>60-1500 nM</td>
<td>MAPKK: 1.0, MAPKK-PP: 1.7, MAPK: 4.7-5.1</td>
</tr>
</tbody>
</table>

The assumed $K_m$ values for each reaction were individually varied over the ranges shown, with the assumed $K_m$ values for the other nine reactions held constant. The effective Hill coefficients were calculated from the steepness of the predicted stimulus/response curves, as described in the text.

The $K_m$ value for reaction 7 has been measured to be 300 nM for the phosphorylation of a mammalian MAPK by a MAPKK (N. Ahn, personal communication). All of the other $K_m$ values were initially assumed to be 300 nM as well.

---

**Fig. 1.** Schematic view of the MAPK cascade. Activation of MAPK depends upon the phosphorylation of two conserved sites [Thr-183 and Tyr-185 in rat p42 MAPK/Erk2 (4, 5)]. Full activation of MAPK also requires phosphorylation of two sites [Ser-218 and Ser-222 in mouse Mek-1/MKK1 (6-10)]. Detailed mechanisms for the activation of various MAPKKks (e.g., Raf-1, B-Raf, Mos) are not yet established; here we assume that MAPKKks are activated and inactivated by enzymes we denote E1 and E2. MAPKK+ denotes activated MAPKK, MAPKK-P and MAPKK-PP denote singly and doubly phosphorylated MAPK, respectively. MAPKK-P and MAPKK-PP denote singly and doubly phosphorylated MAPK, respectively.

**10 chemical reactions**

**INPUT**

**OUTPUT**

**Reservoirs**

**Back Enzymes**
Enzymatic Reactions

Reaction View

\[ S \rightarrow P \]

\[ E \xrightarrow{(c,d,e)} \]

\[ E+S \xrightarrow{c} ES \xrightarrow{e} P+E \]

Interaction View

Private bindings between one S and one E molecule

\[ S() \triangleq \text{new u@d new k@e} \]

\[ !a_c(u,k); (!u_d; S() + !k_e; P()) \]

\[ E() \triangleq ?a_c(u,k); (?u_d; E() + ?k_e; E()) \]

\[ P() \triangleq ... \]
MAPK Cascade Simulation in SPiM

1st stage:
- KKK* barely rises

2nd stage:
- KK-PP rises, but is not stable

3rd stage:
- K-PP flips up to max even anticipating 2nd stage

Rates and concentrations from paper:
1xE1 (0.3 nM)
1xKKP'ase (0.3 nM)
120xKKP'ase (120 nM)
3xKKK (3 nM)
1200xKK (1.2 uM)
1200xK (1.2 uM)

dx = rx = 150, ax = 1
(Km x = (dx + rx) / ax, Km = 300 nM)

1xE1 injected
MAPK Cascade Simulation in SPiM

All coefficients 1.0 !!!
100xKKK, 100xKK, 100xK,
13xE2, 13xKKPse, 13xKPse.
nxE1 as indicated
(1xE1 is not sufficient to produce an output)
MHC Class I Antigen Presentation


- part of the cellular immune response
  - MHC class I complexes present self and foreign peptide on the cell surface
  - recognized by T lymphocytes and natural killer cells
  - critical for development of self tolerant T cell in thymus

MHC Class I Antigen Presentation

Flytrap model

- Continuous generation of HLA-A2 and peptide
- Degradation of free empty HLA-A2 and free peptide
- Tapasin pathway
  - accelerated peptide dissociation
  - reduced closure trigger rate

- Egression
  - spontaneous opening/closing of empty MHC complex

A stochastic pi-calculus model of MHC class I antigen presentation, Leonard Goldstein.

with Luca Cardelli and Andrew Phillips (Microsoft) and Tim Elliott and Joern Werner (U. Southampton)
2. The Gene Machine

The “Central Dogma” of Molecular Biology

4-letter digital code
4-letter digital code
20-letter digital code
50,000 (?) shapes

The Lactose Operon

Pretty far from the atoms.
Regulation of a gene (positive and negative) influences transcription. The regulatory region has precise DNA sequences, but not meant for coding proteins: meant for binding regulators.

Transcription produces molecules (RNA or, through RNA, proteins) that bind to regulatory region of other genes (or that are end-products).

Human (and mammalian) Genome Size
3Gbp (Giga base pairs) 750MB @ 4bp/Byte (CD)
Non-repetitive: 1Gbp 250MB
In genes: 320Mbp 80MB
Coding: 160Mbp 40MB
Protein-coding genes: 30,000-40,000

M.Genitalium (smallest true organism)
580,073bp 145KB (eBook)
E.Coli (bacteria): 4Mbp 1MB (floppy)
Yeast (eukarya): 12Mbp 3MB (MP3 song)
Wheat 17Gbp 4.25GB (DVD)
Gene Composition

Under the assumptions [Kim & Tidor]
1) The solution is well-stirred (no spatial dependence on concentrations or rates).
2) There is no regulation cross-talk.
3) Control of expression is at transcription level only (no RNA-RNA or RNA-protein effects)
4) Transcriptions and translation rates monotonically affect mRNA and protein concentrations resp.

Is a shorthand for:

Ex: Bistable Switch

Ex: Oscillator

Expressed
Repressed
Expressing
Gene Regulatory Networks

http://strc.herts.ac.uk/bio/maria/NetBuilder/


The Programming Model

• Strange facts about genetic networks:
  - Not an operator algebra. The output of each gate is fixed and pre-determined; it is never a function of the input!
  - Not term-rewriting, nor Petri nets. Inhibition is widespread.
  - Not Communicating Sequential Processes. Feedback is widespread: asynchronous communication needed to avoid immediate self-deadlocks. Even the simplest gates cannot be modeled as a single synchronous automata.
  - Not Message-Passing between genes. Messages themselves have behavior (e.g., they stochastically decay and combine), hence messages are processes as well.
  - Not Data-Flow. Any attempt to use data-flow-style modeling seems doomed because of widespread loops that lead to deadlocks or unbounded queues. Data-flow tokens do not “decay” like proteins.

• How can it possibly work?
  - Stochastic broadcasting. The apparently crude idea of broadcasting a whole bunch of asynchronous decaying messages to activate a future gate, means there are never any “pipeline full” deadlocks, even in presence of abundant feedback loops.
  - Stochastic degradation. Degradation is fundamental for system stability, and at the same time can lead to sudden instability and detection of concentration levels.
Notations for the Gene Machine

- Many of the same techniques as for the Protein Machine apply.
  - Process Calculi, Petri Nets, Term-Rewriting Systems...

- But the “programming model” is different.
  - Asynchronous stochastic control.
  - Biologically poorly understood.
  - Network “motifs” are being analyzed.

- Specific techniques:
  - Hybrid Petri Nets
    - [Matsuno, Doi, Nagasaki, Miyano] Gene Regulation
    - Genomic Object Net www.genomicobject.net

- Gene Regulation Diagrams

- Mixed Gene-Protein Diagrams
Gene Gates and Circuits

A gene gate

\[ \neg(a, b) \triangleq \begin{cases} \neg r; \tau_\eta; \neg(a, b) + \\ \tau_\epsilon; (tr(b) \mid \neg(a, b)) \end{cases} \]

\[ tr(p) \triangleq (l_p; tr(p)) + \tau_\delta \]

A genetic circuit (engineered in E.Coli)

\[ \neg(a, b) \mid \neg(b, c) \mid \neg(c, a) \]

The stochastic-π program

\[
\text{val } dk = 0.001 (* \text{ Decay rate } *) \\
\text{val } inh = 0.001 (* \text{ Inhibition rate } *) \\
\text{val } cst = 0.1 (* \text{ Constitutive rate } *) \\
\]

let \( tr(p:chan()) = \)

\[
\text{do } l_p; tr(p) \text{ or delay@dk } \\
\]

let \( \neg(a:chan(), b:chan()) = \)

\[
\text{do } ?a; \text{ delay@inh}; \neg(a, b) \text{ or delay@cst}; (tr(b) \mid \neg(a, b)) \\
\]

(* The circuit *)

\[
\text{val } bnd = 1.0 (* \text{ Protein binding rate } * ) \\
\text{new } a@bnd:chan() \text{ new } b@bnd:chan() \text{ new } c@bnd:chan() \\
\text{run } (\neg(c, a) \mid \neg(a, b) \mid \neg(b, c)) \\
\]

A stochastic simulation (in SPiM)

\[ r=1.0, \varepsilon=0.1, h=0.001, \delta=0.001 \]
Guet et al.: D038/lac-


Experiment:
- **aTc**: 0101
- **IPTG**: 0011
- **GFP**: 0100

The output of some circuits did not seem to make any sense...

3. The Membrane Machine

Molecular transport and transformation through dynamic compartment fusion and fission.

Well, what is all that for?
“Given the complicated pathways that have evolved to synthesize them, it seems likely that these [modified proteins] have important functions, but for the most part these functions are not known” [MBC p.609]
Membrane Fusion

Positive curvature to Negative curvature transition in 3D

Aggressive fusion (virus)

Proposed sequence of events in pH sensitive hemagglutinin membrane fusion

Cooperative fusion (vesicle)

“Fusion of the two membranes immediately follows prefusion, but precisely how this occurs is not known” [MCB p742]
Membrane Fission

Assembly and disassembly of the clathrin coat

Vesicle Formation

"Nonetheless, the actual process whereby a segment of phospholipid bilayer is 'pinched off' to form a pit and eventually a new vesicle is still not understood" [MCB p.746]

Cytokinesis (Mitosis)
The Membrane Machine “Instruction Set”

Mito: special cases

Xo: special cases

Endo: special cases

Arbitrary subsystem

Mito

Mate

Pino

Phago

Zero case

One case

Fusion

Fission

P Q

P Q

P Q

P Q

P Q

P Q

P Q

P Q

P Q

P Q

P Q

P Q
Mito/Mate by 3 Endo/Exo
Notations for the Membrane Machine

- **“Snapshot” diagrams**
  - In biology literature.

- **P-Systems**
    http://psystems.disco.unimib.it/.

- **BioAmbients**
  - An extension of BioSPI along Ambient Calculus lines (with more bio-relevant mobility primitives) to model dynamic compartments.

- **Brane Calculi**
  - Computation on the membrane...
Membrane Algorithms

Protein Production and Secretion

LDL-Cholesterol Degradation

Viral Replication


Abstract Machines of Systems Biology

Gene Machine
Nucleotides

Regulation

Gene Regulatory Networks

Membrane Machine
Phospholipids

Confinement Storage
Bulk Transport

Protein Machine
Aminoacids

Metabolism, Propulsion
Signal Processing
Molecular Transport

Model Integration
Different time
and space scales

Holds genome(s),
confines regulators

Directs membrane construction
and protein embedding

Holds receptors, actuators
hosts reactions

Implements fusion, fission

Signal Processing Networks

Biochemical Networks

Surface and Extracellular Features

Transport Networks

Glycan Machine
Sugars

Membrane-embedded
Glycans

Aminoacids
Nucleotides
Phospholipids

A
B
C
P
A
x
y

PP PP QQ QQ
Model Construction and Validation
Reactive Systems

- **Modeling biological systems**
  - Not as continuous systems (often highly nonlinear)
  - But as discrete **reactive systems**; abstract machines where:
    - **States** represent situations
    - Event-driven **transitions** between states represent dynamics
  - The adequacy of describing (discrete) complex systems as reactive systems has been argued convincingly [Harel]

- **Many biological systems exhibit features of reactive systems:**
  - Discrete transitions between states
  - Deep layering of abstractions ("steps" at multiple levels)
  - Complexity from combinatorial interaction of simple components
  - High degree of concurrency and nondeterminism
  - "Emergent behavior" not obvious from part list
Model Validation: Simulation

- **Basic stochastic algorithm: Gillespie**
  - Exact (i.e. based on physics) stochastic simulation of chemical kinetics.
  - Can compute concentrations and reaction times for biochemical networks.

- **Stochastic Process Calculi**
  - **BioSpi** [Shapiro, Regev, Priami, et. al.]
    - Stochastic process calculus based on Gillespie.
  - **BioAmbients** [Regev, Panina, Silverma, Cardelli, Shapiro]
    - Extension of BioSpi for membranes.
  - **Case study: Lymphocytes in Inflamed Blood Vessels** [Lecca, Priami, Quaglia]
    - Original analysis of lymphocyte rolling in blood vessels of different diameters.
  - **Case study: Lambda Switch** [Celine Kuttler, IRI Lille]
    - Model of phage lambda genome (well-studied system).
  - **Case study: VICE** [U. Pisa]
    - Minimal prokaryote genome (180 genes) and metabolism of whole VIrtual CEll, in stochastic π-calculus, simulated under stable conditions for 40K transitions.

- **Hybrid approaches**
  - **Charon language** [UPenn]
    - Hybrid systems: continuous differential equations + discrete/stochastic mode switching.
  - Etc.
Model Validation: “Program” Analysis

- **Causality Analysis**
  - *Biochemical pathways,* (“concurrent traces” such as the one here), are found in biology publications, summarizing known facts.
  - This one, however, was automatically generated from a program written in BioSpi by comparing traces of all possible interactions. [Curti, Priami, Degano, Baldari]
  - One can play with the program to investigate various hypotheses about the pathways.

- **Control Flow Analysis**
  - Flow analysis techniques applied to process calculi.
  - Overapproximation of behavior used to answer questions about what “cannot happen”.
  - Analysis of positive feedback transcription regulation in BioAmbients [Flemming Nielson].

- **Probabilistic Abstract Interpretation**
  - [DiPierro Wicklicky].
Model Validation: Modelchecking

- **Temporal**
  - Software verification of biomolecular systems (NA pump) [Ciobanu]
  - Analysis of mammalian cell cycle (after Kohn) in CTL. [Chabrier-Rivier Chiaverini Danos Fages Schachter]
    - E.g. is state $S_1$ a necessary checkpoint for reaching state $S_2$?

- **Quantitative: Simpathica/xssys** [Antioniotti Park Policriti Ugel Mishra]
  - Quantitative temporal logic queries of human Purine metabolism model.
    
    $$\text{Eventually(Always (PRPP = 1.7 \times PRPP1) \ implies steady\_state()) and Eventually(Always(IMP < 2 \times IMP1)) and Eventually(Always(hx\_pool < 10\times hx\_pool1)))}$$

- **Stochastic: Spring** [Parker Normal Kwiatkowska]
  - Designed for stochastic (computer) network analysis
    - Discrete and Continuous Markov Processes.
    - Process input language.
    - Modelchecking of probabilistic queries.
What Reactive Systems Do For Us

We can write things down precisely
- We can modularly describe high structural and combinatorial complexity ("do programming").

We can calculate and analyze
- Directly support simulation.
- Support analysis (e.g. control flow, causality, nondeterminism).
- Support state exploration (modelchecking).

We can visualize
- Automata-like presentations.
- Petri-Net-like presentations.
- State Charts, Live Sequence Charts [Harel]
  - Hierarchical automata.
  - Scenario composition.

We can reason
- Suitable equivalences on processes induce algebraic laws.
- We can relate different systems (e.g. equivalent behaviors).
- We can relate different abstraction levels.
- We can use equivalences for state minimization (symmetries).

Disclaimers
- Some of these technologies are basically ready (medium-scale stochastic simulation and analysis, medium-scale nondeterministic and stochastic modelchecking).
- Others need to scale up significantly to be really useful. This is (has been) the challenge for computer scientists.

Many approaches, same basic philosophy, tools being built:
**Conclusions**

**Q:** “The data are accumulating and the computers are humming, what we are lacking are the words, the grammar and the syntax of a new language…”

D. Bray (TIBS 22(9):325-326, 1997)

**A:** “The most advanced tools for computer process description seem to be also the best tools for the description of biomolecular systems.”

E. Shapiro (Lecture Notes)
References


Papers

BioAmbients
a stochastic calculus with compartments.

Brane Calculi
process calculi with computation “on” the membranes, not inside them.

Bitonal Systems
membrane reactions and their connections to “local” patch reactions.

Abstract Machines of Systems Biology
the abstract machines implemented by biochemical toolkits.

www.luca.demon.co.uk/BioComputing.htm
Positions

Postdoc at Imperial College London:
Centre for Integrative Systems Biology

Computational Modelling of Biological Processes

Deadline for Applications: 10th February 2006

Applications are invited for the position of a research assistant/associate for up to three years to work on the application of process-modelling techniques to the signalling of phagocytosis. This position has been awarded to Dr Philippa Gardner and Dr Luca Cardelli (Microsoft Research Cambridge), funded by a large BBSRC/EPSRC grant to support a new Centre for Systems Biology at Imperial. It complements two equivalent positions (one for a biologist, one for a mathematician) in Centre for Molecular Microbiology and Infection & Division of Cell and Molecular Biology, to investigate the spatio-temporal control of phagocytic signalling during uptake of bacteria. We expect the three researchers to work closely together.

Applicants should complete an application form, downloadable from http://www.imperial.ac.uk/employment/academicform.htm. Applications will not be accepted unless they are on the correct form and clearly marked with the Job Reference Number PG Bio 05. The application form should be accompanied by a full CV with names and addresses of 3 referee and should be sent to: Mrs Nicola Rogers Department of Computing Imperial College London South Kensington Campus London, SW7 2AZ UK Email: n.c.rogers@imperial.ac.uk.

Various positions in Trento:

http://www.msr-unitn.unitn.it