Abstract Machines of Systems Biology

Luca Cardelli
Microsoft Research
Cambridge UK

2005-03-16 York

www.luca.demon.co.uk
50 Years of Molecular Cell Biology

- **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

- **Bootstrapping 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
  - When each component and each reaction is understood, the system is understood (?)

- But this has not led to understand how “the system” works
  - Behavior comes from complex chains 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

- Can we fix it when it breaks?
  - The question really becomes: How is information structured and processed?
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 Molecular Biology

Biochemical Networks - The Protein Machine
Gene Regulatory Networks - The Gene Machine
Transport Networks - The Membrane Machine

Systems Biology
1. “How do components interact?”
2. “Gather high-throughput data.”

Glycan Machine
Sugars
Surface and Extracellular Features

Protein Machine
Amino acids
Metabolism, Propulsion
Signal Processing
Molecular Transport

Membrane Machine
Phospholipids
Confinement
Storage
Bulk Transport

Gene Machine
Nucleotides
Regulation
Diverse:
- chemical toolkits
- instruction sets
- programming models
- notations

Model Integration
Different time and space scales
Holds genome(s), confines regulators
Makes proteins: where, when, how much
Directs membrane construction and protein embedding
Signals conditions and events
Holds receptors, actuators, hosts reactions
Implements fusion, fission
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
Some Allosteric Switches

Allosteric ("other shape") reactions modify accessibility.

**Kinase**
- donates phosphate \( P \)
- phosphorilates other proteins

**Phosphatase**
- accepts phosphate \( P \)
- dephosphorilates other proteins

**Logical AND**

at equal concentrations of the individual input stimuli, activation is much higher if both stimuli are present

"Phosphatase Kinase Kinase" = a kinase that activates a kinase that activates a phosphatase that deactivates a protein.

Humans have the same number of modular protein domains (building blocks) as worms, but twice the number of multi-domain proteins.
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 extensible 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.
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.

The Protein Machine “Instruction Set”

**On/Off switches**

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.

cf. BioCalculus [Kitano&Nagasaki], κ-calculus [Danos&Laneve]
Notations for the Protein Machine

- **Stochastic π-Calculus**
  - Priami (following Hillston's PEPA) formalizes a stochastic version of p-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 $\mu M$</th>
<th>Range of effective Hill coefficients (nM) predicted for</th>
<th>MAPKK</th>
<th>MAPKK$^*$</th>
<th>MAPKK-P</th>
<th>MAPKK-PP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. MAPKKK $\rightarrow$ MAPKK</td>
<td>0-1500</td>
<td>1.0</td>
<td>1.7</td>
<td>4.9</td>
<td>9.0</td>
<td>4.9</td>
</tr>
<tr>
<td>2. MAPKK $\rightarrow$ MAPKK $^*$</td>
<td>0-4000</td>
<td>1.0</td>
<td>1.7</td>
<td>4.9</td>
<td>9.0</td>
<td>4.9</td>
</tr>
<tr>
<td>3. MAPKK $\rightarrow$ MAPKK-P</td>
<td>0-1500</td>
<td>1.0</td>
<td>1.3-2.3</td>
<td>4.0-5.1</td>
<td>9.0</td>
<td>4.9</td>
</tr>
<tr>
<td>4. MAPKK-P $\rightarrow$ MAPKK</td>
<td>0-1500</td>
<td>1.0</td>
<td>1.5-1.9</td>
<td>3.6-6.7</td>
<td>9.0</td>
<td>4.9</td>
</tr>
<tr>
<td>5. MAPKK-P $\rightarrow$ MAPKK-PP</td>
<td>0-1500</td>
<td>1.0</td>
<td>1.3-2.4</td>
<td>3.8-5.2</td>
<td>9.0</td>
<td>4.9</td>
</tr>
<tr>
<td>6. MAPKK-PP $\rightarrow$ MAPKK-P</td>
<td>0-1500</td>
<td>1.0</td>
<td>1.7-1.8</td>
<td>4.1-6.4</td>
<td>9.0</td>
<td>4.9</td>
</tr>
<tr>
<td>7. MAPKK-P $\rightarrow$ MAPKK</td>
<td>0-1500 [300 $\mu M$]</td>
<td>1.0</td>
<td>1.7</td>
<td>3.7-6.2</td>
<td>9.0</td>
<td>4.9</td>
</tr>
<tr>
<td>8. MAPKK $\rightarrow$ MAPKK</td>
<td>0-1500</td>
<td>1.0</td>
<td>1.7</td>
<td>4.3-5.2</td>
<td>9.0</td>
<td>4.9</td>
</tr>
<tr>
<td>9. MAPKK-P $\rightarrow$ MAPKK-PP</td>
<td>0-1500</td>
<td>1.0</td>
<td>1.7</td>
<td>3.4-6.1</td>
<td>9.0</td>
<td>4.9</td>
</tr>
<tr>
<td>10. MAPKK-PP $\rightarrow$ MAPKK-P</td>
<td>0-1500</td>
<td>1.0</td>
<td>1.7</td>
<td>4.7-5.1</td>
<td>9.0</td>
<td>4.9</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.

Calculations: Eqs. 1-10 represent the reactions of the MAPK cascade, which are shown schematically in Fig. 1. We have used Goldbeter and Koshland's nomenclature for the rate constants— the letter $a$ denotes association, $d$ denotes dissociation without catalysis, and $k$ denotes product formation (11). KKK denotes MAPKKK, KK denotes MAPKK, and K denotes MAPK.

\[ d_1 \rightarrow KKK + E_1 \rightarrow KKK^* + E_1 \]
\[ k_1 \rightarrow KKK + E_2 \rightarrow KKK^* + E_2 \]
\[ KKK^* + E_2 \rightarrow KKK^* + E_2 \]
\[ d_2 \rightarrow K + KKK \]
\[ d_3 \rightarrow K + KKK^* \]
\[ KKK + E_1 \rightarrow KKK + E_1 \]
\[ KKK + E_2 \rightarrow KKK + E_2 \]
\[ K + KKK \rightarrow K + KKK^* \]
\[ K + KKK^* \rightarrow K + KKK^* \]
\[ KKK + E_1 \rightarrow KKK + E_1 \]
\[ KKK + E_2 \rightarrow KKK^* + E_1 \]
\[ KKK + E_2 \rightarrow KKK + E_2 \]
\[ K + KKK \rightarrow K + KKK^* \]
\[ K + KKK^* \rightarrow K + KKK^* \]
\[ KKK + E_1 \rightarrow KKK + E_1 \]
\[ KKK + E_2 \rightarrow KKK^* + E_1 \]
\[ KKK + E_2 \rightarrow KKK^* + E_2 \]
\[ K + KKK \rightarrow K + KKK^* \]
\[ K + KKK^* \rightarrow K + KKK^* \]
\[ KKK + E_1 \rightarrow KKK + E_1 \]
\[ KKK + E_2 \rightarrow KKK^* + E_1 \]
\[ KKK + E_2 \rightarrow KKK^* + E_2 \]
\[ K + KKK \rightarrow K + KKK^* \]
\[ K + KKK^* \rightarrow K + KKK^* \]

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 MAPKKs (e.g., Raf-1, B-Raf, Mos) are not yet established; here we assume that MAPKKs 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 MAPKK, respectively. MAPKK-P and MAPKK-PP denote singly and doubly phosphorylated MAPK. P'ase denotes phosphatase.
As 18 Ordinary Differential Equations
Plus 7 conservation equations

The 10 reactions described above give rise to 18 rate
equations.

One equation for each
species (8) and complex
(10), but not for constant
concentration enzymes (4)

In addition, there are seven conservation equations (Eqs.
29-35).

\[ [\text{KKK}_{\text{tot}}] = [\text{KKK}] + [\text{KKK}^*] + [\text{KKK}\cdot E1] + [\text{KKK}\cdot E2] + [\text{KKK}^*\cdot K] + [\text{KKK}^*\cdot K-P] \]

Each molecule in exactly one state

2005-03-16 15
As 12 processes (in SPiM)

let KKK() =
  (new u1@d1:Release !a1(u1); (do !u1;KKK() or !k1;KKKst()))

and KKKst() =
  (new u2@d2:Release do !a2(u2); (do !u2;KKKst() or !k2;KKK())
  or ?a3(u3); (do ?u3;KKKst() or ?k3;KKKst())
  or ?a5(u5); (do ?u5;KKKst() or ?k5;KKKst()))

let E1() =
  ?a1(u1); (do ?u1;E1() or ?k1;E1())

let E2() =
  ?a2(u2); (do ?u2;E2() or ?k2;E2())

let KK() =
  (new u3@d3:Release !a3(u3); (do !u3;KK() or !k3;KK_P()))

and KK_P() =
  (new u6@d6:Release do !a6(u6); (do !u6;KK_P() or !k6;KK_P())
  or ?a7(u7); (do ?u7;KK_P() or ?k7;KK_P())
  or ?a9(u9); (do ?u9;KK_P() or ?k9;KK_P())
  or ?a4(u4); (do ?u4;KK_P() or ?k4;KK_P())
  or ?a6(u6); (do ?u6;KKPse() or ?k6;KKPse())

and KKPse() =
  do ?a4(u4); (do ?u4;KKPse() or ?k4;KKPse())
  or ?a6(u6); (do ?u6;KKPse() or ?k6;KKPse())

let K() =
  (new u7@d7:Release !a7(u7); (do !u7;K() or !k7;K_P()))

and K_P() =
  (new u8@d8:Release new u9@d9:Release do !a8(u8); (do !u8;K_P() or !k8;K())
  or !a9(u9); (do !u9;K_P() or !k9;K_PP()))

and K_PP() =
  (new u10@d10:Release !a10(u10); (do !u10;K_PP() or !k10;K_P()))

and KPse() =
  do ?a8(u8); (do ?u8;KPse() or ?k8;KPse())
  or ?a10(u10); (do ?u10;KPse() or ?k10;KPse())

[1] substrate
[2] substrate
[3] kinase
[4] phtase
[5] substrate
[6] substrate
[7] kinase
[8] substrate
[9] kinase
[10] substrate
and 20 Interaction Channels

type Release = chan()
type Bond = chan(Release)
type React = chan()

new a1@1.0: Bond val d1=1.0 new k1@1.0: React
new a2@1.0: Bond val d2=1.0 new k2@1.0: React
new a3@1.0: Bond val d3=1.0 new k3@1.0: React
new a4@1.0: Bond val d4=1.0 new k4@1.0: React
new a5@1.0: Bond val d5=1.0 new k5@1.0: React
new a6@1.0: Bond val d6=1.0 new k6@1.0: React
new a7@1.0: Bond val d7=1.0 new k7@1.0: React
new a8@1.0: Bond val d8=1.0 new k8@1.0: React
new a9@1.0: Bond val d9=1.0 new k9@1.0: React
new a10@1.0: Bond val d10=1.0 new k10@1.0: React

... run 100 of KKK() run 100 of KK() run 100 of K()
run 1 of E2() run 1 of KKPse() run 1 of KPse()
run 1 of E1()

a_i,k_i: Two channels for each reversible chemical reaction of 2 molecules. (No behavior attached to channels except interaction rate.)
MAPK Cascade Simulation in SPiM

KKK
KK
K

KK-P
K-P

1xE1 injected

KKK* reacts
KK-PP rises quicker
K-PP flips up to 100!

All coefficients 1.0 !!!
100xKKK, 100xKK, 100xK,
1xE2, 1xKKPse, 1xKPse.

Input is 1xE1.
Output is 100xK-PP (ultrasensitivity).

KKK*  KK-PP  K-PP down slowly

1xE1 removed
1xE1 injected
MAPK Cascade Simulation in SPiM

All coefficients 1.0 !!!
100xKKK, 100xKK, 100xK,
10xE2, 10xKPpse, 10xKPse.
(so 1xE1 is no longer sufficient to produce an output)
2. The Gene Machine

The “Central Dogma” of Molecular Biology

DNA        messenger RNA        PROTEIN        SYSTEMS

regulation

transcription translation interaction

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

Lactose Operon

Metabolic space

Protein space

Gene space

5

8

A

9

0

1

'
The Gene Machine “Instruction Set”

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

Is a shorthand for:

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.

Ex: Bistable Switch

Ex: Oscillator

Expressed
Repressed
Expressing
Indirect Gene Effects

No combination of standard high-throughput experiments can reconstruct an a-priori known gene/protein network [Wagner].

One of many bistable switches that cannot be described by pure gene regulatory networks [Francois & Hakim].

Fig. 1. The importance of specifying gene activity when reconstructing genetic networks. (a) A hypothetical biochemical pathway involving two transcription factors, a protein kinase, and a protein phosphatase, as well as the genes encoding them. See text for details. (b) Shown is a list of perturbation effects for each of the five genes in (a), when perturbing individual genes by deleting them, and when using mRNA expression level as an indicator of gene activity. The left-most symbol in each line stands for the perturbed gene. To the right of each column is a list of genes whose activity is affected by the perturbation. (c) Analogous to (b) but for a different notion of gene activity (phosphorylation state).
The Central Dogma

DNA \[\rightarrow\] transcription \[\rightarrow\] mRNA \[\rightarrow\] translation \[\rightarrow\] Protein

RNA is not just an intermediary; it can:
- Fold-up like a protein
- Act like an enzyme
- Regulate other transcribed RNA
- Direct protein editing
- ...

Challenging the Dogma (in higher organisms)

97-98% of the transcriptional output of the human genome is non-protein-coding RNA.
30-40,000 "protein genes" (1.5% of genome)
60-100,000 "transcription units" (>30% of genome is transcribed)
**Structure of a Regulatory Region**

**Module A functions:**

- Vegetal plate expression in early development:

- Synergism with modules B and G enhancing endoderm expression in later development:

- Repression in ectoderm (modules E and F) and skeletogenic mesenchyme (module DC):

Modules E, F and DC with LiCI treatment:

Fig. 1. *Endo16* cis-regulatory system and interactive roles of module A. (A) Diversity of protein binding sites and organization into modular subregions [modified from (7)]. Specific DNA binding sites are indicated as red blocks; modular subregions are denoted by letters G to A (Bp, basal promoter). Proteins binding at the target sites considered in this work are indicated: Otx, SpOtx-1 (12); SpGCF1 (14); the proteins CG, Z, and P, which are not yet cloned; and protein C [a CREB family protein (18)] in subregion F. Proteins for which sites occur in multiple regions of the DNA sequence (indicated by the black line) are shown beneath. (B) Sequence of module A and location of protein binding sites. Sites are indicated in the same colors as in (A). A fragment containing CG3 and CG4 sites as well as Bp has no endoderm-specific activity and serves other upstream cis-regulatory systems promiscuously; similarly, the *Endo16* cis-regulatory system functions specifically with heterologous promoters substituted for Bp (5, 8, 19). Boxed sequences indicate conserved core elements of the target sites (7, 12, 14), not the complete target site sequences. (C) Integrative and interactive functions of module A (5, 8). Module A communicates the output of all upstream modules to the basal transcription apparatus. It also initiates endoderm expression, increases the output of modules B and G, and is required for functions of the upstream modules E, F, and DC. These functions are repression in nonendodermal domains and enhancement of expression in response to LiCl.
### Function of a Regulatory Region

#### Formulation

**Repression functions of modules F, E, and DC mediated by Z site**

- If \( F = 1 \) or \( E = 1 \) or \( CD = 1 \) and \( Z = 1 \):
  - \( \alpha = 1 \)
- Else:
  - \( \alpha = 0 \)

**Both P and CG\(_1\) needed for synergistic link with module B**

- If \( P = 1 \) and \( CG\(_1\) = 1 \):
  - \( \beta = 2 \)
- Else:
  - \( \beta = 0 \)

**Final step up of system output**

- If \( (CG\(_1\) = 1 \) and \( CG\(_2\) = 1 \) and \( CG\(_3\) = 1 \)):
  - \( \gamma = 2 \)
- Else:
  - \( \gamma = 1 \)

**Positive input from modules B and G**

- \( \delta(t) = B(t) + G(t) \)

**Synergistic amplification of module B output by CG\(_1\)-P subsystem**

- \( \varepsilon(t) = \beta \delta(t) \)

**Switch determining whether Otx site in module A, or upstream modules (i.e., mainly module B), will control level of activity**

- If \( \zeta(t) = 0 \):
  - \( \xi(t) = Otx(t) \)
- Else:
  - \( \xi(t) = \eta(t) \)

**Repression function inoperative in endoderm but blocks activity elsewhere**

- \( \eta(t) = 0 \)
- Else:
  - \( \eta(t) = \xi(t) \)

**Final output communicated to BTA**

- \( \theta(t) = \gamma \eta(t) \)

---

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.

- **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
A gene gate
\[ \text{neg}(a, b) \triangleq ?a_r; \tau_\eta; \text{neg}(a, b) + \tau_v; (\text{tr}(b) \mid \text{neg}(a, b)) \]
\[ \text{tr}(p) \triangleq (\lnot p_r; \text{tr}(p)) + \tau_\delta \]

The stochastic-\(\pi\) program
\[ \text{val} \ dk = 0.001 \quad (\text{\text{* Decay rate \text{*}}}) \]
\[ \text{val} \ inh = 0.001 \quad (\text{\text{* Inhibition rate \text{*}}}) \]
\[ \text{val} \ cst = 0.1 \quad (\text{\text{* Constitutive rate \text{*}}}) \]

let \( \text{tr}(p:\text{chan}) = \)
\[ \text{do} \ l p; \text{tr}(p) \text{ or delay}@dk \]

let \( \text{neg}(a:\text{chan}, b:\text{chan}) = \)
\[ \text{do} \ l a; \text{delay}@inh; \text{neg}(a, b) \]
\[ \text{or} \ text{delay}@cst; \text{tr}(b) \mid \text{neg}(a, b) \]

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

A genetic circuit (engineered in E.Coli)

A stochastic simulation (in SPiM)
3. The Membrane Machine

Molecular transport and transformation through dynamic compartment fusion and fission.

The Instruction Set

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].

Voet, Voet & Pratt
Fundamentals of Biochemistry
Wiley 1999, Ch10 Fig 10-22.
Membrane Fusion

Proposed sequence of events in pH sensitive hemagglutinin membrane fusion

1. Cell membrane
2. Virus membrane
3. Aggressive fusion (virus)
4. By unknown mechanisms, the exoplasmic leaflets of the two membranes fuse” [MCB p745]
5. Cooperative fusion (vesicle)
6. “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”

- **Mate**
- **Mito**
- **Arbitrary subsystem**

- **Zero case**
- **Drip**
- **One case**
- **Bud**

- **Fusion**
- **Fission**

- **Exo**
- **Endo**
- **Arbitrary subsystem**

- **Zero case**
- **Pino**
- **One case**
- **Phago**
Locally Implementable!

Global Views

Mito \rightarrow (Fission)

Mate \rightarrow (Fusion)

Endo \rightarrow (Fission)

Exo \rightarrow (Fusion)

Same Local View!
Mito/Mate by 3 Endo/Exo
Notations for the Membrane Machine

- "Snapshot" diagrams
  - In biology literature.

- P-Systems
  - G. Paun uses ideas from the theory of grammars and formal languages to model "Membrane Computing" (book 2002).
    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

Viral Replication

LDL-Cholesterol Degradation


**Brane Calculi**

**systems**  
\[ P, Q ::= \circ \mid P \circ Q \mid !P \mid \sigma(P) \]

*ests of membranes

**branes**  
\[ \sigma, \tau ::= 0 \mid \sigma|\tau \mid !\sigma \mid \alpha.\sigma \]

**actions**  
\[ \alpha ::= 1 \mid ... \]

**1D fluids (\(\sigma\)) inside a 2D fluid (P)**

TWO commutative monoids instead of ONE of normal process calculi

\[ \sigma(P) \]

\[ \sigma|\tau(P) \]

\[ a.\sigma|\tau = (a.\sigma)|\tau \]

N.B. Restriction (\(\nu\)) could be added to both systems and branes. It usually would originate in branes, but would extrude to whole systems.
Brane Reactions (Cartoons)

A Turing-Complete language
[Busi Gorrieri]
Brane-Molecule Reactions (Cartoons)

With molecule multisets $p, q$:

$$
\begin{align*}
\text{p}_1 &\quad \text{p}_2 \\
\text{q}_1 &\quad \text{q}_2
\end{align*}
$$
Phago \( \cap_n \sigma | \sigma' (P) \circ \cap_n (\rho) . \tau | \tau' (Q) \rightarrow \tau | \tau' (\rho(\sigma | \sigma' (P)) \circ Q) \)

Exo \( \cap_n . \tau | \tau' (\cap_n . \sigma | \sigma' (P) \circ Q) \rightarrow \rho \circ \sigma | \sigma' | \tau | \tau' (Q) \)

Pino \( \cap (\rho). \sigma | \sigma' (P) \rightarrow \sigma | \sigma' (\rho(\circ) \circ P) \)

N.B.: the parity of nesting of P and Q is preserved; this makes the reactions preserve bitonality.

B&R \( p_1 \circ p_1 (p_2) \Rightarrow q_1 (q_2). \alpha | \sigma (p_2 \circ P) \rightarrow q_1 \circ \alpha | \sigma (q_2 \circ P) \)

(multiset rewriting, inside and outside membranes)
Derivable Reactions (Cartoons)

A Decidable-Termination language
[Busi Gorrieri]

Diagram:
- Mate: P Q → P Q
- Bud: P Q → P Q
- Drip: P → P
Viral Reproduction

- **Virus**
- **Endosome**
- **Phago**
- **Mate**
- **Exo**
- **Nucleus**
- **Endoplasmic Reticulum**
- **RNase**
- **Capsid**
- **Membrane**
- **Envelope protein**
- **Nucleocapsid**
- **Translation**
- **Assembly**
- **Budding**
- **Exo**
- **Drip**
- **Vesicle**

**Steps:**
- Infection
- Replication
- Progeny

[MBC p.279] annotated
Ex: Viral Progeny

Assume:
\[ \text{nucap} \circ \text{cytosol} \rightarrow \rightarrow \text{nucap}^n \circ \text{envelope-vesicle}^m \circ \text{cytosol}' \]
by available cellular machinery

Then:

\[ \text{cell} \]
\[ \text{envelope-vesicle} \quad \text{nucap} \]
\[ \text{bud}^{\downarrow} (\text{vRNA}) \circ \text{cytosol}' \]
\[ \text{Exo} \]

\[ \text{envelope} \quad \text{nucap} \]
\[ \text{bud}^{\downarrow} (\text{vRNA}) \circ \text{cytosol}' \]
\[ \text{Bud} \]

\[ \text{cell} \quad \text{virus} \]
\[ \text{nucap} \circ \text{cytosol}' \]
Lysosome and target don’t just merge.

Biologically, Mito/Mate clearly happens. However, weird sequences of Endo/Exo are also common.
Surprisingly enough, we have found that parameter values that give rise to a stable steady state in the deterministic limit continue to produce reliable oscillations in the stochastic case, as shown in Fig. 5. Therefore, the presence of noise not only changes the behavior of the system by adding more disorder but can also lead to marked qualitative differences.

Fig. 5. Time evolution of $R$ for the deterministic Eq. [1] (a) and stochastic (b) versions of the model. The values of the parameters are as in the caption of Fig. 1, except that now we set $\delta = 0.05 \text{ h}^{-1}$. For these parameter values, $\tau < 0$, so that the fixed point is stable.

**Mechanisms of noise-resistance in genetic oscillators**

José M. G. Vilar, Hao Yuan Kueh, Naama Barkai, and Stanislas Leibler

*PNAS* April 30, 2002 vol. 99 no. 9 5991
Scaling up to Big Systems: ODE's vs Processes
- A process calculus:
  - The modular representation of concurrent (and stochastic) processes of all kinds.
  - Cuts down to CTMCs (Continuous Time Markov Chains) in the finite case (not always). Then, standard tools are applicable.
  - Can be given friendly automata-like scalable graphical syntax (work in progress: Andrew Phillips).
  - Is directly executable (e.g. via Gillespie).
  - Is analyzable (large body of literature, at least in the non-stochastic case).
**Chemistry vs. \(\pi\)-calculus**

**A process calculus (chemistry, or SBML)**

\[
\text{Na} + \text{Cl} \rightarrow_{k_1} \text{Na}^+ + \text{Cl}^- \\
\text{Na}^+ + \text{Cl}^- \rightarrow_{k_2} \text{Na} + \text{Cl}
\]

This Petri-Net-like graphical representation degenerates into spaghetti diagrams: precise and dynamic, but not scalable, structured, or maintainable.

**A compositional graphical representation, and the corresponding calculus.**

\[
\begin{align*}
\text{Na} &= !r_{k_1}; ?s_{k_2}; \text{Na} \\
\text{Cl} &= ?r_{k_1}; !s_{k_2}; \text{Cl}
\end{align*}
\]

**The same “model”**

Maps to a CTMC

Maps to a CTMC


**A different process calculus (\(\pi\))**
From Reactions to ODE's

r₁: A + B → $k_1$ C + C
r₂: A + C → $k_2$ D
r₃: C → $k_3$ E + F
r₄: F → $k_4$ B

Write the coefficients by columns

Concentration changes
Stoichiometric matrix

Read the rate laws from the columns
$v_i(x, e_i, k_i)$

$x$: chemical species
$[-]$: concentrations
$v$: rate laws
$k$: kinetic parameters
$N$: stoichiometric matrix
$e$: catalysts (if any)

Read the concentration changes from the rows

E.g. $d[A]/dt = -k_1[A][B] - k_2[A][C]$

$d[A]/dt = -v_1 - v_2$
$d[B]/dt = -v_1 + v_4$
$d[C]/dt = 2v_1 - v_2 - v_3$
$d[D]/dt = v_2$
$d[E]/dt = v_3$
$d[F]/dt = v_3 - v_4$

$\frac{d[x]}{dt} = N \cdot v$

Stoichiometric Matrix

<table>
<thead>
<tr>
<th>Species</th>
<th>N₁</th>
<th>N₂</th>
<th>N₃</th>
<th>N₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-1</td>
<td>-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>-1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>D</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td>1</td>
<td>-1</td>
<td></td>
</tr>
</tbody>
</table>
# From Reactions to Processes

## Processes

- \( r_1: A+B \rightarrow k_1 \ C+C \)
- \( r_2: A+C \rightarrow k_2 \ D \)
- \( r_3: C \rightarrow k_3 \ E+F \)
- \( r_4: F \rightarrow k_4 \ B \)

## Interactions

<table>
<thead>
<tr>
<th>N</th>
<th>( r_1 )</th>
<th>( r_2 )</th>
<th>( r_3 )</th>
<th>( r_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-1</td>
<td>-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>-1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>-1</td>
<td>-1</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td>1</td>
<td>-1</td>
</tr>
</tbody>
</table>

For binary reactions, first species in the column does an input and produces result, second species does an output. For unary reactions, species does a tau action and produces result. No ternary reactions.

\[
A = \tau v_1 k_1 (C|C) + \tau v_2 k_2 . D + \tau a
\]

\[
B = \lambda v_1 k_1 + \tau b
\]

\[
C = \lambda v_2 k_2 + \tau k_3 (E|F) + \tau c
\]

\[
D = 0 + \tau d
\]

\[
E = 0 + \tau e
\]

\[
F = \tau k_3 . B + \tau f
\]

Add a barb for counting and plotting

Read the process interactions from the rows

(Rate laws are implicit in stochastic semantics)
Stoichiometric Matrices Blow Up

- Can Translate Chemistry to ODE’s or Processes
  - It is standard to go from chemical equations to ODE’s via stoichiometric matrices.
  - It is similarly possible to go from chemical equations to processes via stoichiometric matrices.

- But there is a better way:
  - Stoichiometric matrices blow-up exponentially for biochemical systems (unlike for ordinary chemical systems) because proteins have combinatorial state and complexes are common.
  - To avoid this, we should describe biochemical systems compositionally without going through stochiometric matrices (and hence without ODE’s).
Complexes: The ODE Way

- **n** domains
  - A, B, C
- **2n** domain reactions
  - \( A \approx A_p \)
  - \( B \approx B_p \)
  - \( C \approx C_p \)
- **1** complex
  - \( ABC \)

**Stoichiometric Matrix**

<table>
<thead>
<tr>
<th>N</th>
<th>( v_1 )</th>
<th>( v_2 )</th>
<th>( v_3 )</th>
<th>( v_4 )</th>
<th>( v_5 )</th>
<th>( v_6 )</th>
<th>( v_7 )</th>
<th>( v_8 )</th>
<th>( v_9 )</th>
<th>( v_{10} )</th>
<th>( v_{11} )</th>
<th>( v_{12} )</th>
<th>( v_{13} )</th>
<th>( v_{14} )</th>
<th>( v_{15} )</th>
<th>( v_{16} )</th>
<th>( v_{17} )</th>
<th>( v_{18} )</th>
<th>( v_{19} )</th>
<th>( v_{20} )</th>
<th>( v_{21} )</th>
<th>( v_{22} )</th>
<th>( v_{23} )</th>
<th>( v_{24} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( A_pBC )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( ABpC )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( ABC_p )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( ApBpC )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( ABpCp )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( ApBpCp )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td>( 2^n \times 2n(2n^{-1}) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The matrix is very sparse, so the corresponding ODE system is not dense. But it still has \( 2^n \) equations, one per species, plus conservation equations ([ABC]+[\( A_pBC \)=constant, etc.).
Where the local domain reactions are not independent, we can use lateral communication so that each component is aware of the relevant others.

System description is linear in the number of basic components.

(Its “run-time” behavior or analysis potentially blows-up just as in the previous case.)
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