Abstract Machines of Systems Biology
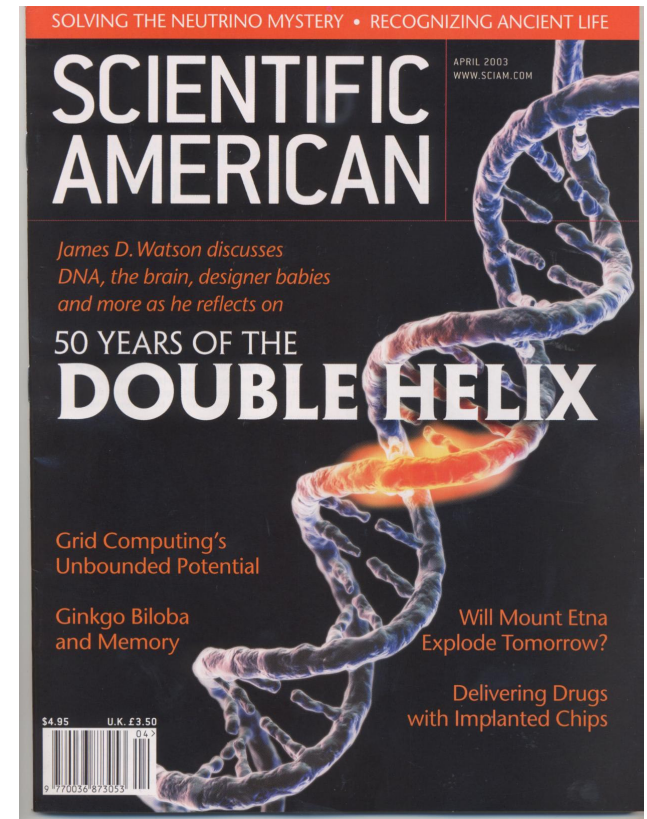
Luca Cardelli

Microsoft Research
Cambridge UK

2006-03-23 Max Planck ISS

www.luca.demon.co.uk

50 Years of <u>Molecular Cell Biology</u>

- **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

Scientific American — April 2003 — www.sciam.com
SOLVING THE NEUTRINO MYSTERY • RECOGNIZING ANCIENT LIFE
James D. Watson discusses DNA, the brain, designer babies and more as he reflects on
50 YEARS OF THE DOUBLE HELIX
Grid Computing's Unbounded Potential
Ginkgo Biloba and Memory
Will Mount Etna Explode Tomorrow?
Delivering Drugs with Implanted Chips
$4.95   U.K. £3.50

Towards <u>Systems Biology</u>

- 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?

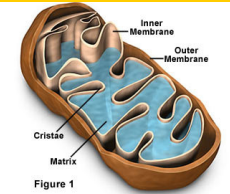# Structural Architecture

Eukaryotic Cell

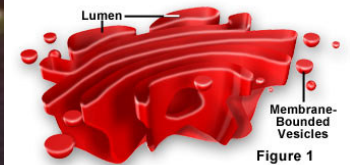(10~100 trillion in human body)

Membranes everywhere
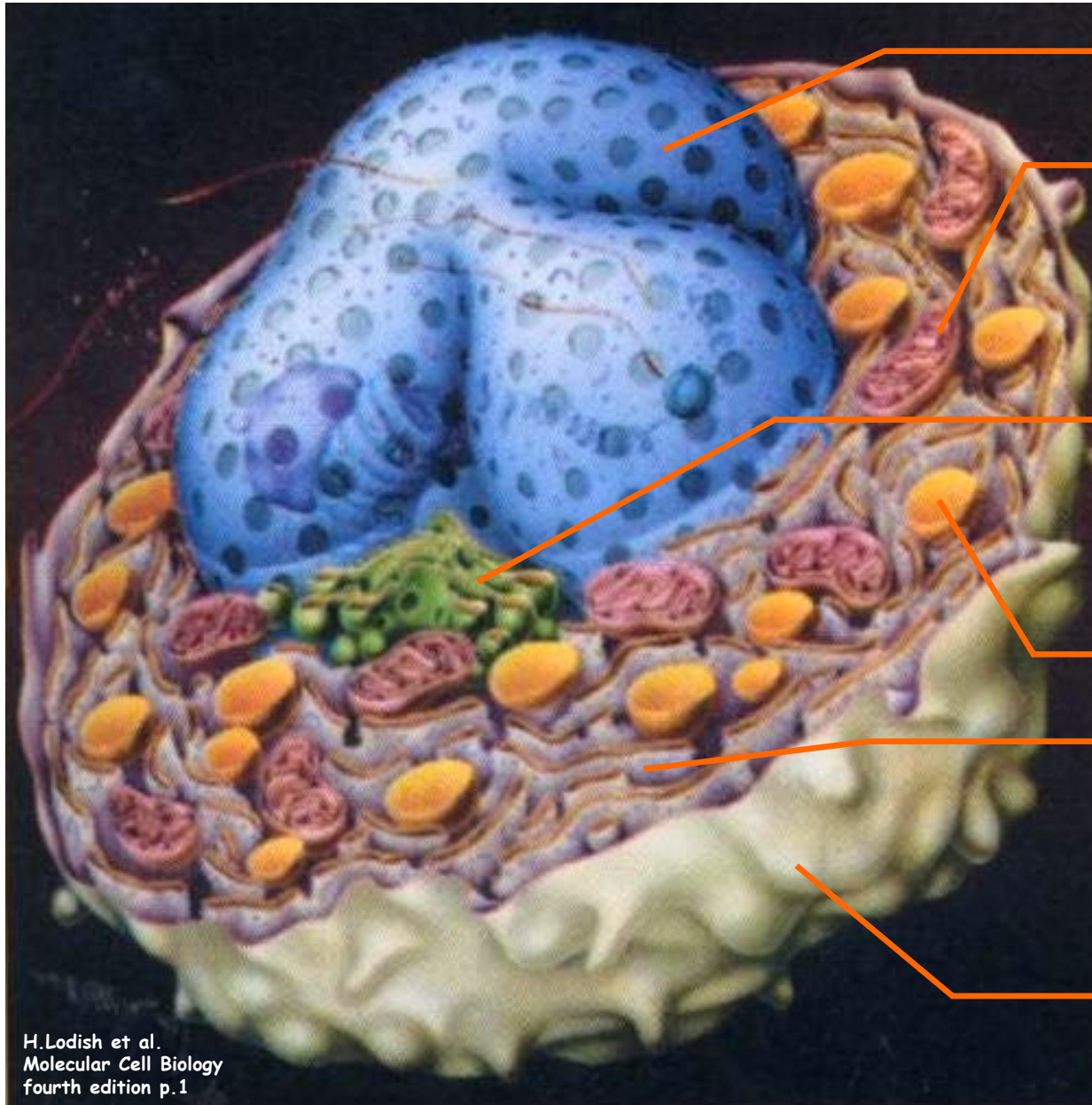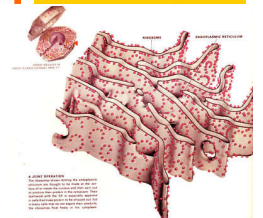
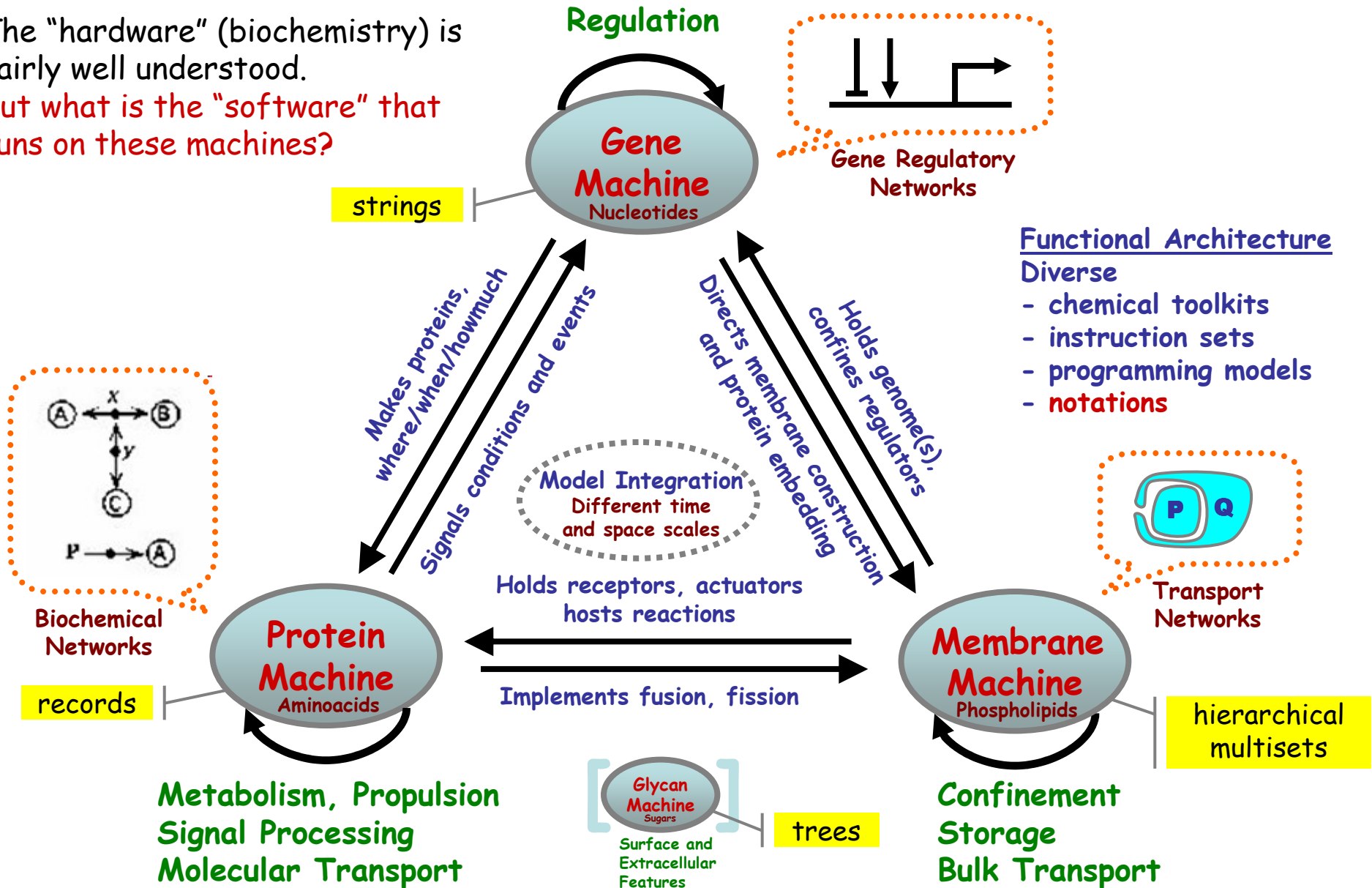Nuclear membrane

Mitochondria

Golgi

Vesicles

E.R.

Plasma membrane (<10% of all membranes)

H.Lodish et al.
Molecular Cell Biology
fourth edition p.1

4

Abstract Machines of Systems Biology

The "hardware" (biochemistry) is fairly well understood.
But what is the "software" that runs on these machines?

Regulation

Gene Machine
Nucleotides

strings

Gene Regulatory Networks

Functional Architecture
Diverse
- chemical toolkits
- instruction sets
- programming models
- notations

Makes proteins, where/when/howmuch

Signals conditions and events

Directs membrane construction and protein embedding

Holds genome(s), confines regulators

Model Integration
Different time and space scales

Holds receptors, actuators hosts reactions

Biochemical Networks

Transport Networks

Protein Machine
Aminoacids

records

Implements fusion, fission

Membrane Machine
Phospholipids

hierarchical multisets

Metabolism, Propulsion
Signal Processing
Molecular Transport

Glycan Machine
Sugars

trees

Surface and Extracellular Features

Confinement
Storage
Bulk Transport

P Q

Storing *Processes*

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

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

- 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.

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

- Still, needs quantitative semantics
  - Stochastic, hybrid, etc. to talk about *rates* (and geometry).

Methods

- Model Construction (writing things down precisely)
  - Formalizing the notations used in systems biology.
  - Formulating description languages.
  - Studying their kinetics (semantics).

- Model Validation (using models for postdiction and prediction)
  - Simulation from compositional descriptions
    - Stochastic: quantitative concurrent semantics.
    - Hybrid: discrete transitions between continuously evolving states.
  - "Program" Analysis
    - Control flow analysis
    - Causality analysis
  - Modelchecking
    - Standard, Quantitative, Probabilistic

Chemistry vs. π-calculus

A process calculus (chemistry)

A different process calculus (π)

$$r: A + B \longrightarrow_{k1} C + D$$
$$s: C + D \longrightarrow_{k2} A + B$$
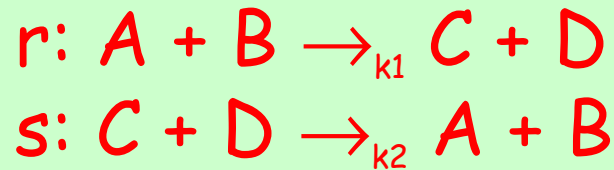
Does A become C or D?
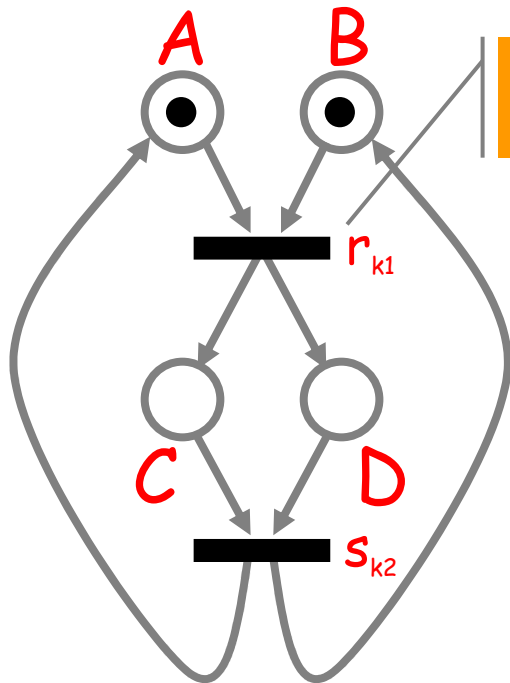
Reaction oriented
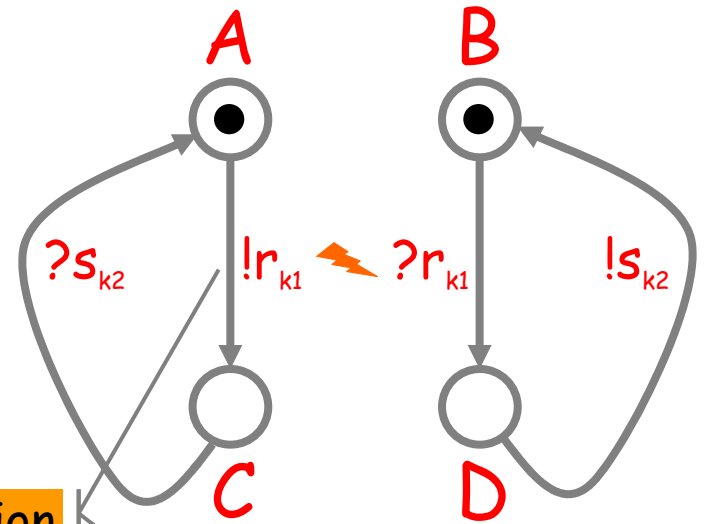
1 line per reaction

Interaction oriented

1 line per component

The same "model"

Maps to a CTMC ⋮ Maps to a CTMC

A becomes C not D!

$$A = !r_{k1}; C$$
$$C = ?s_{k2}; A$$
$$B = ?r_{k1}; D$$
$$D = !s_{k2}; B$$

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.

Why π-calculus, in particular

- **Well studied, compact, precise, and general**
  - A "programming language" first, a mathematical model second
  - Syntax (configurations):   P ::= O | P+P | P|P | ?n(n).P | !n(n).P | (νn)P | *P
  - Semantics (reactions):   (P' + !n(m).P) | (Q' + ?n(m').Q) → P | Q{m'←m}
- **Binary interactions**
  - I.e., "collisions"
- **Reactive and compositional**
  - Each subsystem is a separate (composition of) nondeterministic automata interacting with the environment (more automata)
- **Dynamic network evolution and species evolution**
  - Each subsystem can create fresh connections or spawn new subsystems
- **Compact description of combinatorics (like any programming language)**
  - $(Bit_1 | Bit_2 | \ldots | Bit_n)$ size-n description, where Bit is a 2-state subsystem
    - E.g. a protein with with 2^n phosphorilation configurations (i.e. "different chemical species")
- **Complexation/polymerization**
  - The most characteristic feature of π-calculus (fresh names) models "sticking"

Actin-like
Poly/depolymerization

Monomer Automata
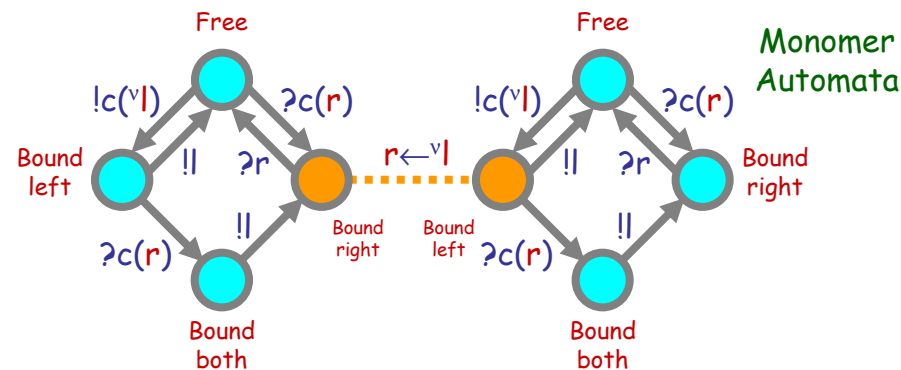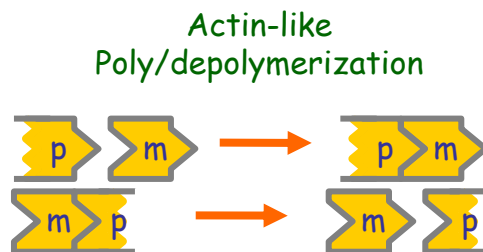
Stochastic π-calculus

- **A quantitative variant of π-calculus:**
  - Channels have stochastic "firing" rates with exponential distribution.
  - Nondeterministic choice becomes *stochastic race*.
  - Cuts down to CTMCs (Continuous Time Markov Chains) in the finite case (not always). Then, standard analytical tools are applicable.
  - Can be given friendly automata-like scalable graphical syntax (Andrew Phillips et al.).
  - Is directly executable (e.g. via the Gillespie algorithm from physical chemistry).
  - Is analyzable (large body of literature, at least in the non-stochastic case).
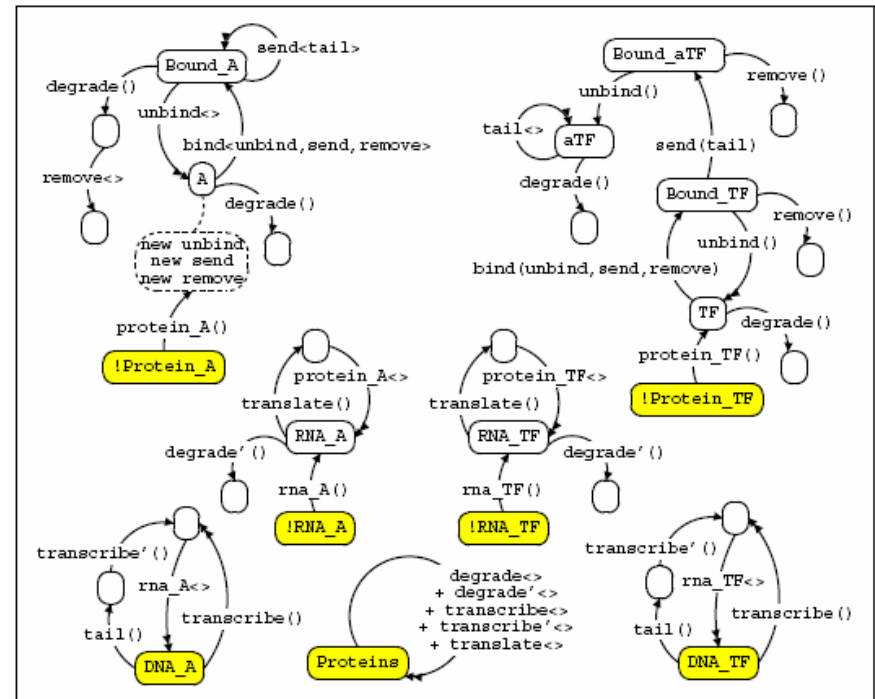
Figure 2. Regulating Gene Expression by Positive Feedback [9]
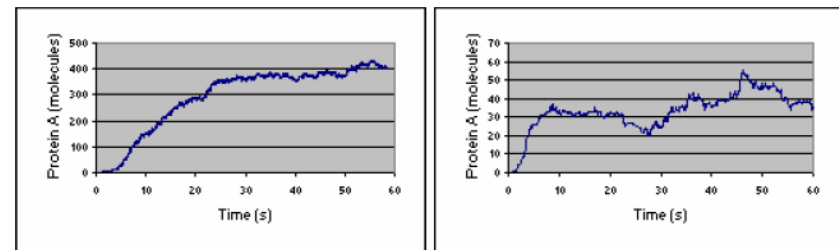
Figure 3. Protein A molecules v.s. time in presence (left) and absence (right) of TF

A.Phillips, L.Cardelli. BioConcur'04.

The Rate of What?

- **In chemistry:**
  - Each reaction involves **2** molecules, and each reaction has a **rate**. Rates belongs to **reactions**. Molecules do *not* have rates.

- **In process algebras:**
  - Should rates belong to:
    - each individual action? only outputs? delays only?
  - The rate of a synchronization of two actions should be the:
    - max? product? undefined if different? infinite (except for delays)?
  - All that has been tried.

- **We go back to chemistry**
  - Rates belong to **channels**. (This is called the "biochemical" stochastic π-calculus by Priami-Regev-Shapiro-Silverman)

- **Issues:**
  - Multiple activities on the same channel (concentrations of molecules involved in a reaction: mass action law of chemistry).
  - Choices between different channels (molecules involved in multiple reactions: still standard chemistry).
  - In biochemistry, rates of homodimerization (a molecule can interact with a copy of itself, but not with itself).
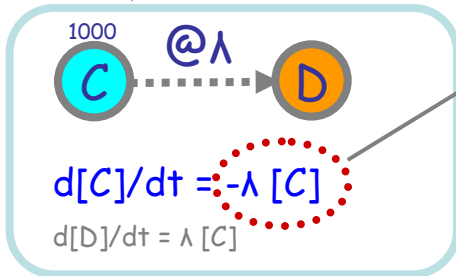
Rates belong to **channels** not to actions!
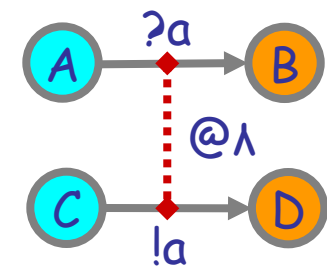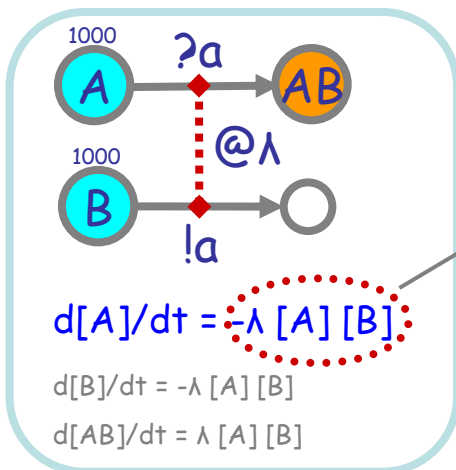
Law of Mass Interaction

The speed of interaction[†] is proportional to the number of possible interactions[‡].

Decay

1000
C @λ ⟶ D

$d[C]/dt = -λ [C]$

$d[D]/dt = λ [C]$

Exponential Decay law
Rate of change proportional to number of possible decays.

Mass interaction

1000
A ?a ⟶ AB
@λ
1000
B !a ⟶ ○

Interaction Law generalizes Decay Law

$d[A]/dt = -λ [A] [B]$

$d[B]/dt = -λ [A] [B]$

$d[AB]/dt = λ [A] [B]$

Mass Interaction law
Rate of change proportional to number of possible interactions

[†] speed of interaction = apparent rate (formally definable) = number of interactions over time

[‡] not proportional to the number of interacting processes!

[P] is the number of processes P (this is informal; it is only meaningful for a set of processes offering a given action, but a set of such processes can be counted and plotted)

— A() — B() — AB()
— C() — D()

— A() — B() — AB()
— C() — D()
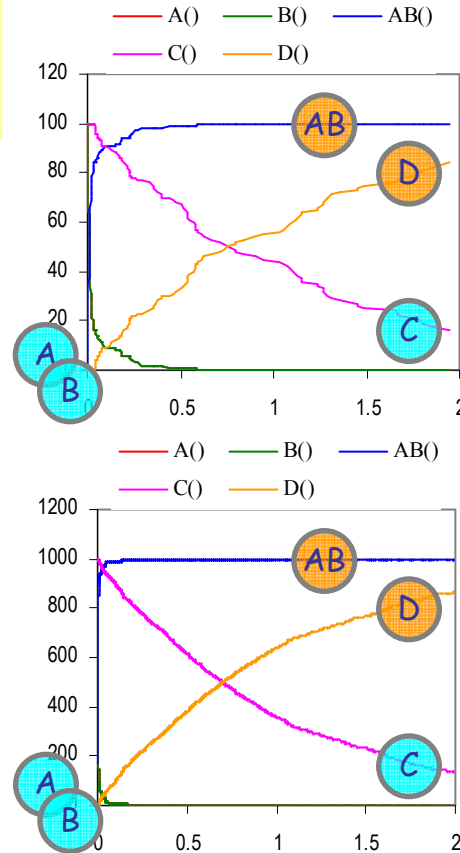
N interactions on the same channel are "faster" than N delays at the same rate (on N independent channels).

Each C has λ chances to decay per second (no matter how many other Cs there are), but each A has [B]*λ chances to interact per second: it depends on how many Bs there are.

```
directive sample 2.0 10000
directive plot A(); B(); AB(); C(); D()
new stop@1.0:chan

val del = 1.0
new a@del:chan

let A() = ?a; AB()
and B() = !a
and AB() = ?stop

let C() = delay@del; D()
and D() = ?stop
```

```
run 100 of (A() | B() | C())
```

```
run 1000 of (A() | B() | C())
```

Chemical Law of Mass Action
The speed of a chemical reaction is proportional to the activity of the reacting substances.

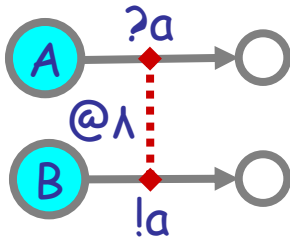(Activity = concentration, for well-stirred aqueous medium)

(Concentration = number of moles per liter of solution)

(Mole = 6.022141×10^{23} particles)

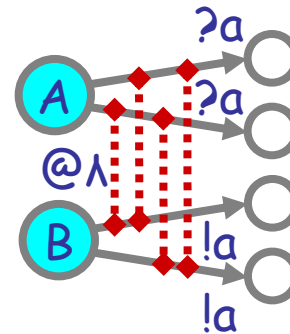# Activity and Apparent Rate
stochastic algebras disagree!

The speed of interaction is proportional to the number of possible interactions.

The mass interaction law [Buchholz] [Priami-Regev-Shapiro-Silverman] is compatible with chemistry [Gillespie] and *incompatible* with any other stochastic algebra in the literature! (including [Priami]; see [Hermanns])

Other algebras assign rates to actions, not channels, with apparent rates:
$2\lambda * 2\lambda = 4\lambda^2$
$\max(2\lambda, 2\lambda) = 2\lambda$ [Goetz]
$\min(2\lambda, 2\lambda) = 2\lambda$ [Priami]
$1/(1/(2\lambda)+1/(2\lambda)) = \lambda$ [PEPA]
$2\lambda * 1 = 2\lambda$ (passive inputs)

#A processes: 1
#B processes: 1
?a activity: 1
!a activity: 1

apparent rate: λ
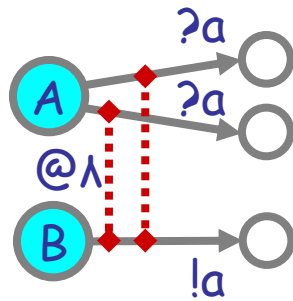
#A processes: 1
#B processes: 1
?a activity: 2
!a activity: 1

apparent rate: 2λ

#A processes: 1
#B processes: 1
?a activity: 2
!a activity: 2

apparent rate: 4λ

Power Actions [Buchholz]

$=_{def}$

$\times k$

Activity = k
(similarly for $?^k a$)

Can generalize k to a real number, and to a dynamically bound variables.

!^{2}a

?faster   ?slower

!a
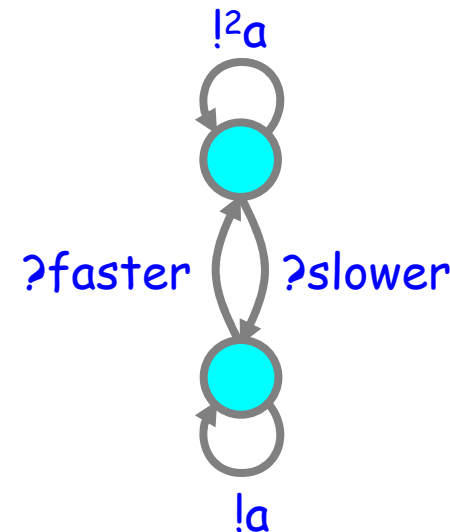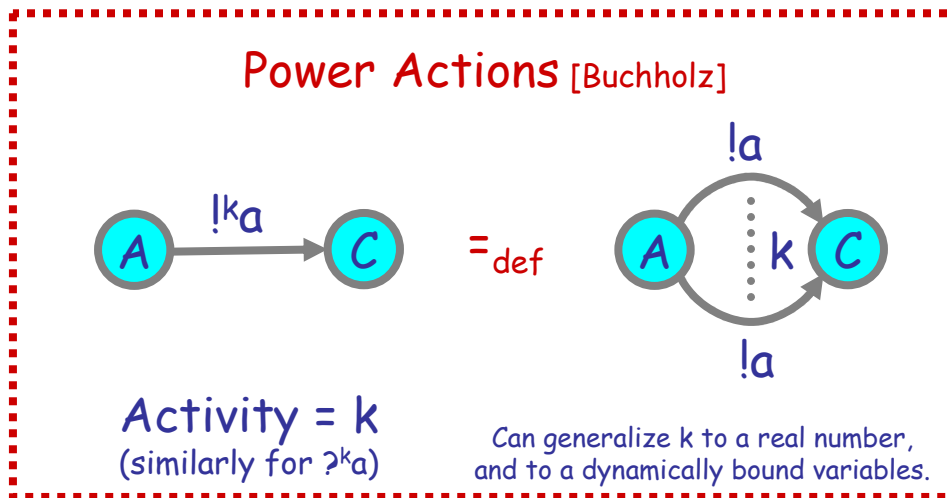
1. The Protein Machine

- Complex folded-up shapes that:
  - Fit together, dock, undock.
  - Excite/unexcite, warp each other.
  - Bring together, catalyze, transform materials.
  - Form complex aggregates and networks.
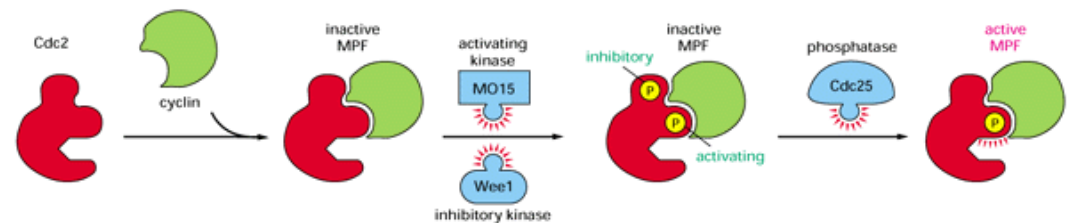
- Mapping out such networks:
  - In principle, it's "just" a very large set of chemical equations.
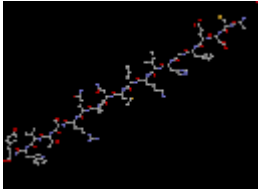 - Notations have been developed to summarize and abstract.

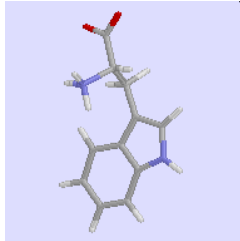An actual molecular interaction network.
(Nodes are distinct protein kinds,
arcs mean that two kinds of proteins interact.)
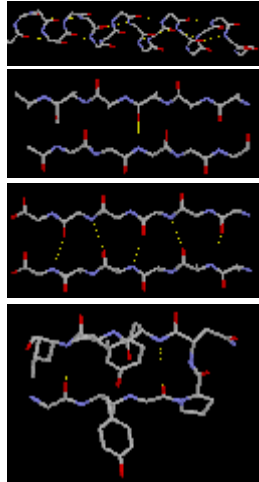
Protein Structure

Primary

Secondary

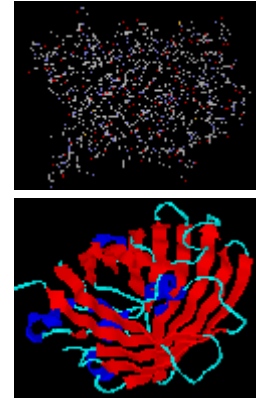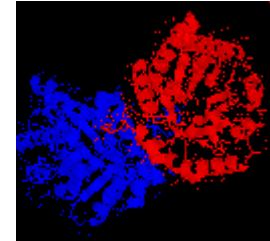Tertiary

Quaternary

The 20 Aminoacids

Tryptophan

Alpha Helix, Beta Sheet

Green Fluorescent Protein

Triose Phosphate Isomerase

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

Protein Function

Regulation

Degradation

Metabolism

Movement

Assembly

Transport

Structure

Signalling
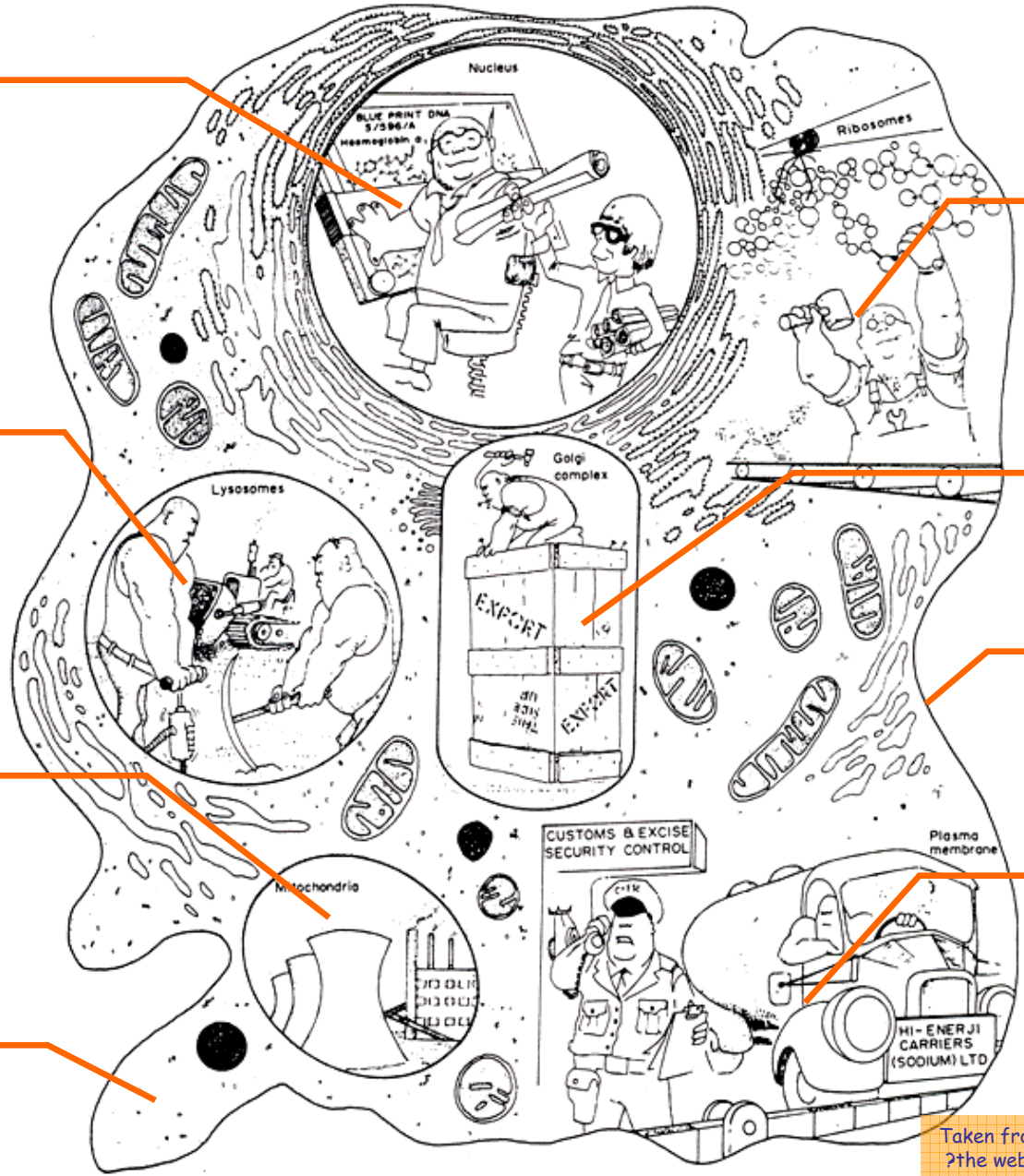
Taken from ?the web?

Some Allosteric Switches

Domain architecture and autoinhibitory interactions in modular switch proteins. (a) Src family kinases contain N-terminal SH3 and SH2 domains, and a kinase domain flanked by intramolecular SH3-binding and SH2-binding sites (when the C-terminal motif tyrosine is phosphorylated by Csk). The crystal structures of several family members show that both intramolecular domain interactions function in concert to lock the kinase in an inactive conformation. Activating stimuli (red) include external SH2 or SH3 ligands. After initial activation, the kinase is maintained in an active state by autophosphorylation of its activation loop. (b) SHP-2 phosphatase contains two SH2 domains and a phosphatase domain. The crystal structure of the phosphatase shows that the N-terminal SH2 domain participates in an autoinhibitory interaction that directly blocks the phosphatase active site. Binding of external SH2 ligands activates by disrupting the autoinhibitory interaction. (c) N-WASP contains an Enabled VASP homology 1 (EVH1) domain, a B motif, a GBD, a proline-rich segment (pro) and an output region (VCA) that alone binds the Arp2/3 complex and stimulates its actin nucleation activity. The B and GBD motifs are required to repress activity and, by current models, are thought to participate in intracomplex interactions (only the structure of the GBD intramolecular complex for WASP is known). GTP-bound Cdc42 and PIP$_2$ synergistically activate N-WASP.

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.
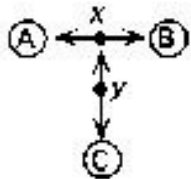
MIM: Molecular Interaction Maps (Kohn)

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.
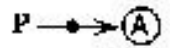
Covalent modification of protein **A**. The single-arrowed line indicates that A can exist in a phosphorylated state. The node represents the phosphorylated species.
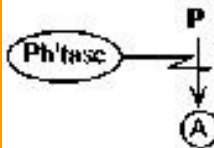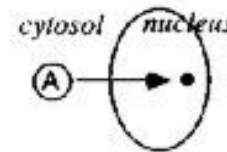
Cleavage of a covalent bond: dephosphorylation of **A** by a phosphatase.

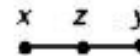Proteolytic cleavage at a specific site within a protein.

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

JDesigner

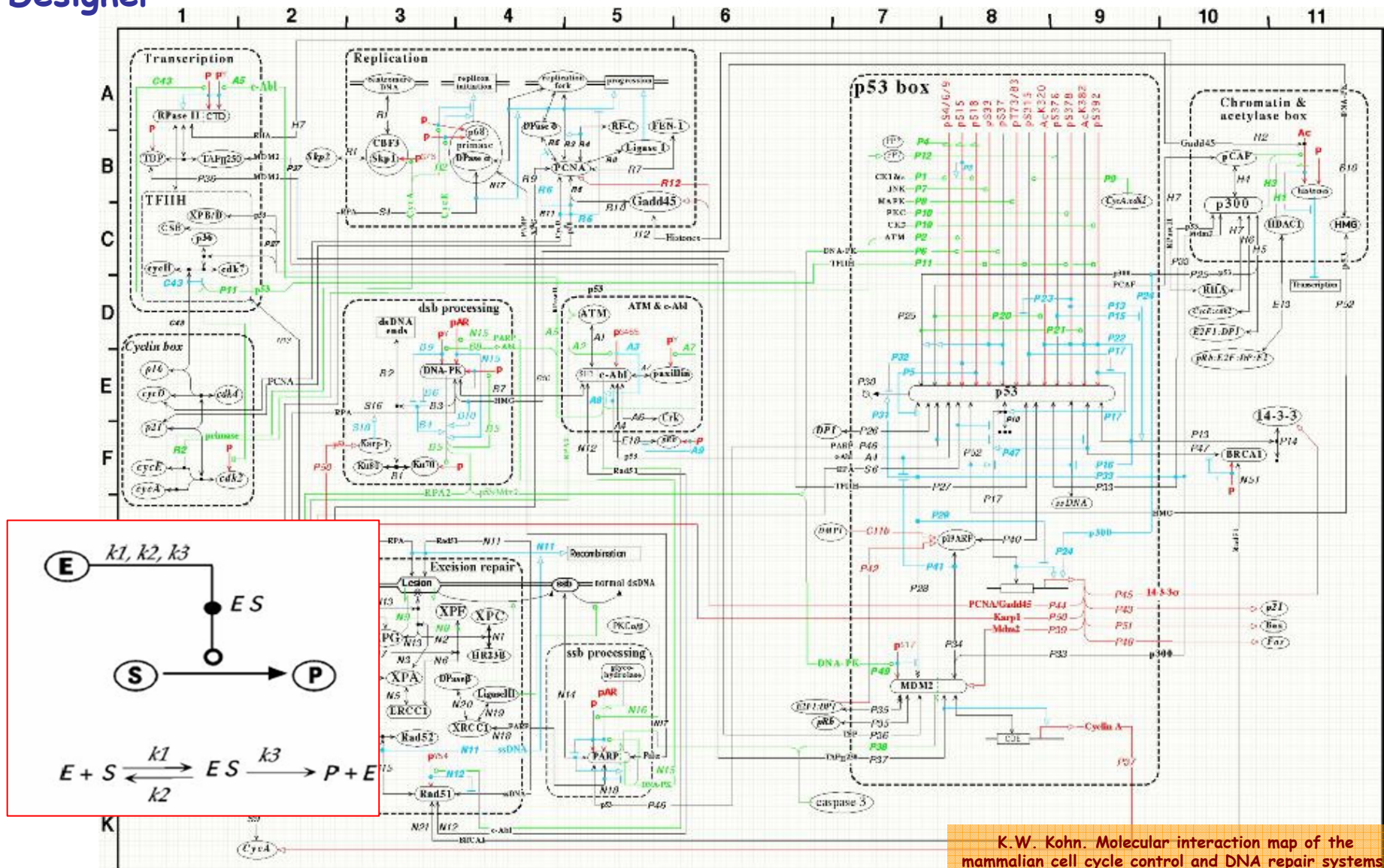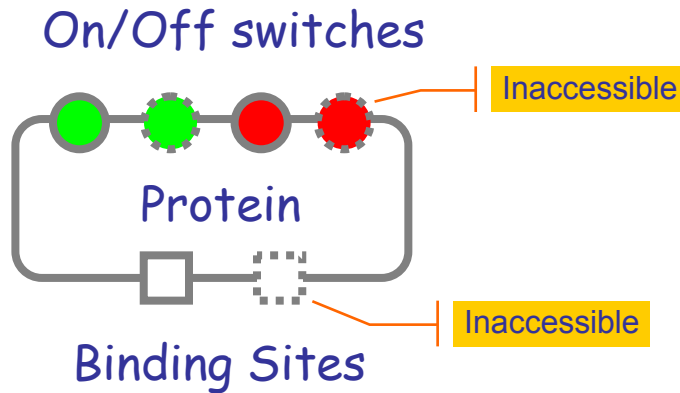## The p53-Mdm2 and DNA Repair Regulatory Network

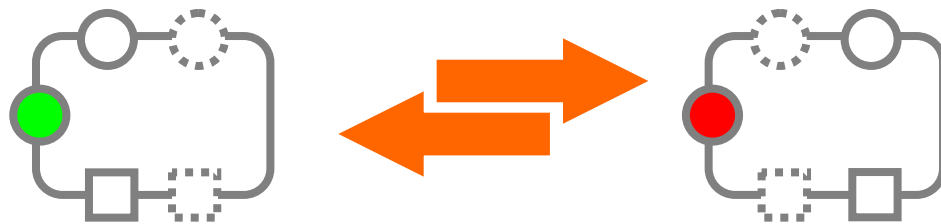Figure 6B: The p53-Mdm2 and DNA repair regulatory network (version 2p - May 19, 1999)

The Protein Machine "Instruction Set"

On/Off switches

Inaccessible

Protein

Inaccessible

Binding Sites

cf. BioCalculus [Kitano&Nagasaki], κ-calculus [Danos&Laneve]

Each protein has a structure
of binary switches and binding sites.
But not all may be always *accessible*.
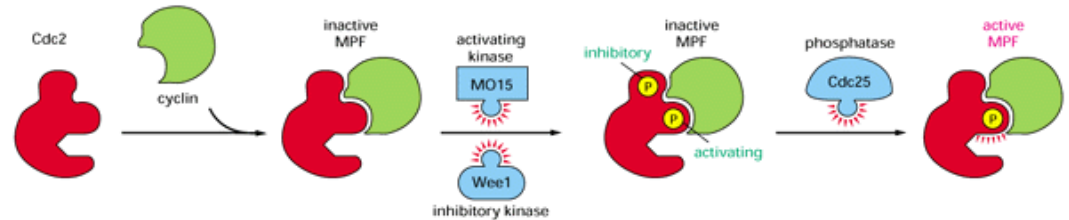
Switching of accessible switches.
- May cause other switches and
binding sites to become (in)accessible.
- May be triggered or inhibited by nearby specific
proteins in specific states.
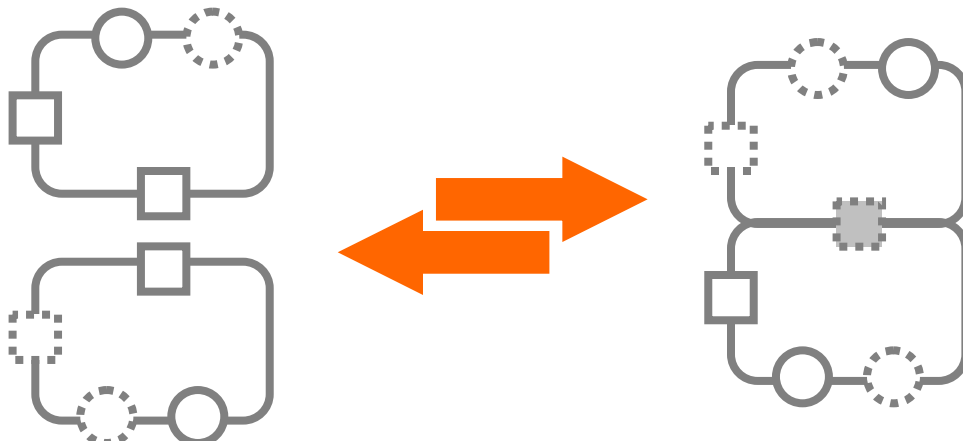
Binding on accessible sites.
- May cause other switches and
binding sites to become (in)accessible.
- May be triggered or inhibited by nearby specific
proteins in specific states.

Notations for the Protein Machine

- **Stochastic π-Calculus**
  - Priami (following Hillston's PEPA) formalizes a stochastic version of 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

MAPK Cascade

Ultrasensitivity in the mitogen-activated protein **cascade**, Chi-Ying F. Huang and James E. Ferrell, Jr., 1996, _Proc. Natl. Acad. Sci. USA_, 93, 10078-10083.

Biochemistry: Huang and Ferrell
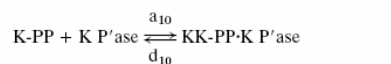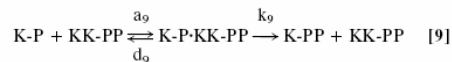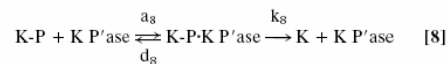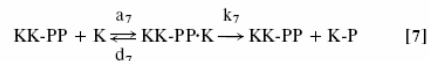
Proc. Natl. Acad. Sci. USA 93 (1996)

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

Reaction	Range of assumed K_m values	Range of effective Hill coefficients (nH) predicted for		
		MAPKKK	MAPKK	MAPK
1. MAPKKK → MAPKKK*	60–1500 nM	1.0	1.7	4.9
2. MAPKKK* → MAPKKK	60–1500 nM	1.0	1.7	4.9
3. MAPKK → MAPKK-P	60–1500 nM	1.0	1.3–2.3	4.0–5.1
4. MAPKK-P → MAPKK	60–1500 nM	1.0	1.5–1.9	3.6–6.7
5. MAPKK-P → MAPKK-PP	60–1500 nM	1.0	1.3–2.4	3.8–5.2
6. MAPKK-PP → MAPKK-P	60–1500 nM	1.0	1.7–1.8	4.1–6.4
7. MAPK → MAPK-P	60–1500 nM (300 nM[†])	1.0	1.7	3.7–6.2
8. MAPK-P → MAPK	60–1500 nM	1.0	1.7	4.3–5.2
9. MAPK-P → MAPK-PP	60–1500 nM	1.0	1.7	3.4–6.1
10. MAPK-PP → MAPK-P	60–1500 nM	1.0	1.7	4.7–5.1

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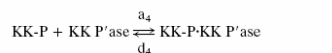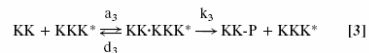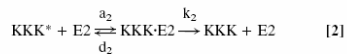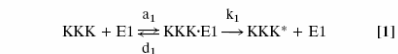[†]The K_m value for reaction 7 has been measured to be 300 nM for the phosphorylation of a mammalian MAPK by a MAPKK (N. Ahn, personal communication). All of the other K_m values were initially assumed to be 300 nM as well.

10 chemical reactions

INPUT (E1)

MAPKKK ⇄ MAPKKK*

E2

MAPKK ⇄ MAPKK-P ⇄ MAPKK-PP

MAPKK P'ase

MAPK ⇄ MAPK-P ⇄ MAPK-PP

MAPK P'ase

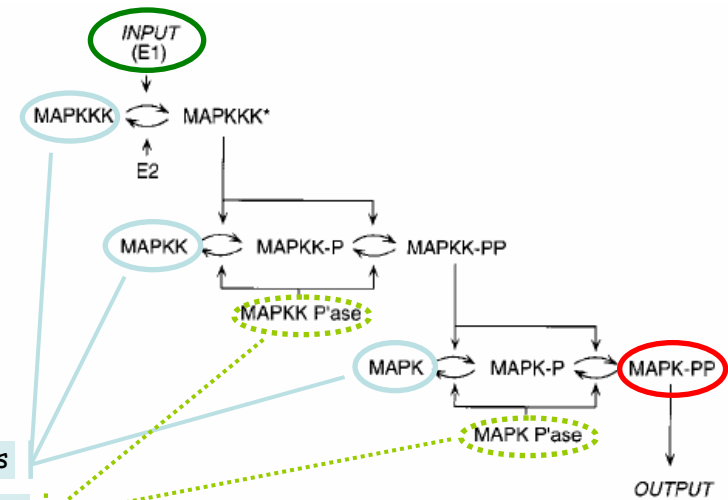OUTPUT

Reservoirs

Back Enzymes

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 MAPKK 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. MAPKKK* denotes activated MAPKKK. MAPKK-P and MAPKK-PP denote singly and doubly phosphorylated MAPKK, respectively. MAPK-P and MAPK-PP denote singly and doubly phosphorylated MAPK. P'ase denotes phosphatase.

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.

$$KKK + E1 \underset{d_1}{\overset{a_1}{\rightleftharpoons}} KKK{\cdot}E1 \overset{k_1}{\longrightarrow} KKK^* + E1 \quad [1]$$

$$KKK^* + E2 \underset{d_2}{\overset{a_2}{\rightleftharpoons}} KKK{\cdot}E2 \overset{k_2}{\longrightarrow} KKK + E2 \quad [2]$$

$$KK + KKK^* \underset{d_3}{\overset{a_3}{\rightleftharpoons}} KK{\cdot}KKK^* \overset{k_3}{\longrightarrow} KK{-}P + KKK^* \quad [3]$$

$$KK{-}P + KK\ P'ase \underset{d_4}{\overset{a_4}{\rightleftharpoons}} KK{-}P{\cdot}KK\ P'ase$$
$$\overset{k_4}{\longrightarrow} KK + KK\ P'ase \quad [4]$$

$$KK{-}P + KKK^* \underset{d_5}{\overset{a_5}{\rightleftharpoons}} KK{-}P{\cdot}KKK^* \overset{k_5}{\longrightarrow} KK{-}PP + KKK^* \quad [5]$$

$$KK{-}PP + KK\ P'ase \underset{d_6}{\overset{a_6}{\rightleftharpoons}} KK{-}PP{\cdot}KK\ P'ase$$
$$\overset{k_6}{\longrightarrow} KK{-}P + KK\ P'ase \quad [6]$$

$$KK{-}PP + K \underset{d_7}{\overset{a_7}{\rightleftharpoons}} KK{-}PP{\cdot}K \overset{k_7}{\longrightarrow} KK{-}PP + K{-}P \quad [7]$$

$$K{-}P + K\ P'ase \underset{d_8}{\overset{a_8}{\rightleftharpoons}} K{-}P{\cdot}K\ P'ase \overset{k_8}{\longrightarrow} K + K\ P'ase \quad [8]$$

$$K{-}P + KK{-}PP \underset{d_9}{\overset{a_9}{\rightleftharpoons}} K{-}P{\cdot}KK{-}PP \overset{k_9}{\longrightarrow} K{-}PP + KK{-}PP \quad [9]$$

$$K{-}PP + K\ P'ase \underset{d_{10}}{\overset{a_{10}}{\rightleftharpoons}} KK{-}PP{\cdot}K\ P'ase$$
$$\overset{k_{10}}{\longrightarrow} K{-}P + K\ P'ase \quad [10]$$

As 18 Ordinary Differential Equations
Plus 7 conservation equations

$$\frac{d}{dt}[KKK] = -a_1[KKK][E1] + d_1[KKK \cdot E1]$$
$$+ k_2[KKK^* \cdot E2] \qquad [11]$$

$$\frac{d}{dt}[KKK \cdot E1] = a_1[KKK][E1] - (d_1 + k_1)[KKK \cdot E1] \qquad [12]$$

$$\frac{d}{dt}[KKK^*] = -a_2[KKK^*][E2] + d_2[KKK^* \cdot E2]$$
$$+ k_1[KKK \cdot E1] + (k_3 + d_3)[KK \cdot KKK^*] - a_3[KKK^*][KK]$$
$$+ (k_5 + d_5)[KK \text{-} P \cdot KKK^*] - a_5[KK \text{-} P][KKK^*] \qquad [13]$$

$$\frac{d}{dt}[KKK^* \cdot E2] = a_2[KKK^*][E2] - (d_2 + k_2)[KKK^* \cdot E2] \qquad [14]$$

$$\frac{d}{dt}[KK] = -a_3[KK][KKK^*] + d_3[KK \cdot KKK^*]$$
$$+ k_4[KK \text{-} P \cdot KK P'ase] \qquad [15]$$

$$\frac{d}{dt}[KK \cdot KKK^*] = a_3[KK][KKK^*]$$
$$- (d_3 + k_3)[KK \cdot KKK^*] \qquad [16]$$

$$\frac{d}{dt}[KK \text{-} P] = -a_4[KK \text{-} P][KK P'ase] + d_4[KK \text{-} P \cdot KK P'ase]$$
$$+ k_3[KK \cdot KKK^*] + k_6[KK \text{-} PP \cdot KK P'ase]$$
$$+ d_5[KK \text{-} P \cdot KKK^*] - a_5[KK \text{-} P][KKK^*] \qquad [17]$$

$$+ d_5[KK \text{-} P \cdot KKK^*] - a_5[KK \text{-} P][KKK^*] \qquad [17]$$

$$\frac{d}{dt}[KK \text{-} P \cdot KK P'ase] = a_4[KK \text{-} P][KK P'ase]$$
$$- (d_4 + k_4)[KK \text{-} P \cdot KK P'ase] \qquad [18]$$

$$\frac{d}{dt}[KK \text{-} P \cdot KKK^*] = a_5[KK \text{-} P][KKK^*]$$
$$- (d_5 + k_5)[KK \text{-} P \cdot KKK^*] \qquad [19]$$

$$\frac{d}{dt}[KK \text{-} PP] = k_5[KK \text{-} P \cdot KKK^*] - a_6[KK \text{-} PP][KK P'ase]$$
$$+ d_6[KK \text{-} PP \cdot KK P'ase] - a_7[KK \text{-} PP][K]$$
$$+ (d_7 + k_7)[K \cdot KK \text{-} PP]$$
$$+ (d_9 + k_9)[K \text{-} P \cdot KK \text{-} PP]$$
$$- a_9[K \text{-} P][KK \text{-} PP] \qquad [20]$$

$$\frac{d}{dt}[KK \text{-} PP \cdot KK P'ase] = a_6[KK \text{-} PP][KK P'ase]$$
$$- (d_6 + K_6)[KK \text{-} PP \cdot KK P'ase] \qquad [21]$$

$$\frac{d}{dt}[K] = -a_7[K][KK \text{-} PP] + d_7[K \cdot KK \text{-} PP]$$
$$+ k_8[K \text{-} P \cdot K P'ase] \qquad [22]$$

$$\frac{d}{dt}[K \cdot KK \text{-} PP] = a_7[K][KK \text{-} PP] - (d_7 + k_7)[K \cdot KK \text{-} PP] \qquad [23]$$

$$\frac{d}{dt}[K \text{-} P] = k_7[K \cdot KK \text{-} PP] - a_8[K \text{-} P][K P'ase]$$
$$+ d_8[K \text{-} P \cdot K P'ase] - a_9[K \text{-} P][KK \text{-} PP]$$
$$+ d_9[K \text{-} P \cdot KK \text{-} PP] + k_{10}[K \text{-} PP \cdot K P'ase] \qquad [24]$$

$$\frac{d}{dt}[K \text{-} P \cdot K P'ase] = a_8[K \text{-} P][K P'ase]$$
$$- (d_8 + k_8)[K \text{-} P \cdot K P'ase] \qquad [25]$$

$$\frac{d}{dt}[K \text{-} P \cdot KK \text{-} PP] = a_9[K \text{-} P][KK \text{-} PP]$$
$$- (d_9 + k_9)[K \text{-} P \cdot KK \text{-} PP] \qquad [26]$$

$$\frac{d}{dt}[K \text{-} PP] = -a_{10}[K \text{-} PP][K P'ase]$$
$$+ d_{10}[K \text{-} P \cdot K P'ase] + k_9[K \text{-} P \cdot KK \text{-} PP] \qquad [27]$$

$$\frac{d}{dt}[K \text{-} PP \cdot K P'ase] = a_{10}[K \text{-} PP][K P'ase]$$
$$- (d_{10} + k_{10})[K \text{-} PP \cdot K P'ase] \qquad [28]$$

> 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)

$$[E1_{tot}] = [E1] + [KKK \cdot E1] \qquad [30]$$
$$[E2_{tot}] = [E2] + [KKK^* \cdot E2] \qquad [31]$$
$$[KK_{tot}] = [KK] + [KK \text{-} P] + [KK \text{-} PP] + [KK \cdot KKK^*]$$
$$+ [KK \text{-} P \cdot KKK^*] + [KK \text{-} P \cdot KK P'ase]$$
$$+ [KK \text{-} PP \cdot KK P'ase]$$
$$+ [KK \text{-} PP \cdot K] + [KK \text{-} PP \cdot K \text{-} P] \qquad [32]$$
$$[KK P'ase_{tot}] = [KK P'ase] + [KK P'ase \cdot KK \text{-} P]$$
$$+ [KK P'ase \cdot KK \text{-} PP] \qquad [33]$$
$$[K_{tot}] = [K] + [K \text{-} P] + [K \text{-} PP] + [KK \text{-} PP \cdot K]$$
$$+ KK \text{-} PP \text{-} K \text{-} P] + [K \text{-} P \cdot K P'ase] + [K \text{-} PP \cdot K P'ase] \qquad [34]$$
$$[K P'ase_{tot}] = [K P'ase] + [K \text{-} P \cdot K P'ase]$$
$$+ [K \text{-} PP \cdot K P'ase] \qquad [35]$$

These equations were solved numerically using the Runge–Kutta-based NDSolve algorithm in Mathematica (Wolfram Research, Champaign, IL). An annotated copy of the Mathematica code for the MAPK cascade rate equations can be obtained from J.E.F.

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

$$[KKK_{tot}] = [KKK] + [KKK^*] + [KKK \cdot E1]$$
$$+ [KKK^* \cdot E2]$$
$$+ [KKK^* \cdot K] + [KKK^* \cdot K \text{-} P] \qquad [29]$$

> in exactly one state

> Each molecule

The Circuit

(input)

E1

KKK ⇌ KKK* KK ⇌ KK-P ⇌ KK-PP K ⇌ K-P ⇌ K-PP

E2

KK-P'ase

K-P'ase

(output)
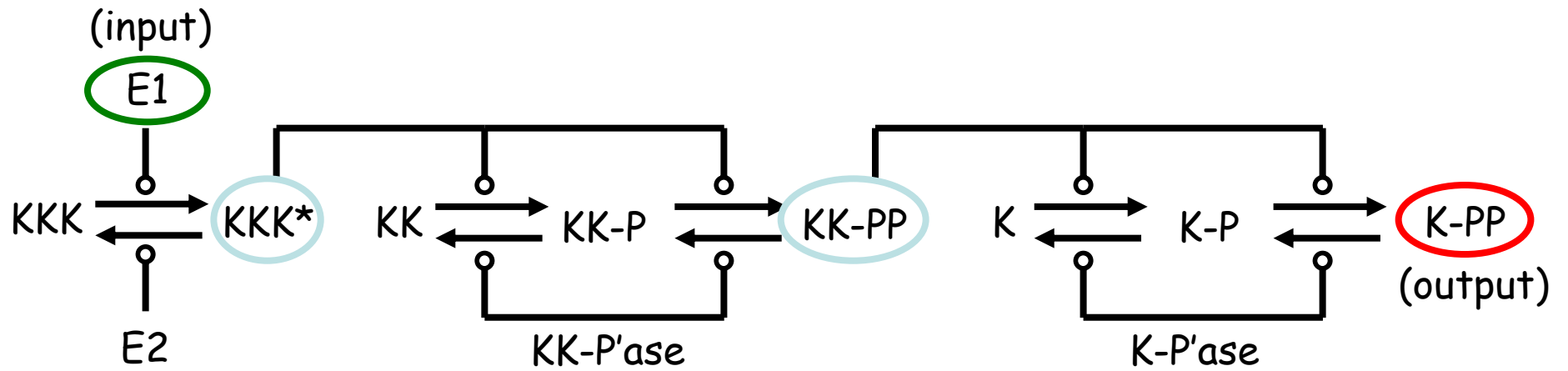
Enzymatic Reactions

Reaction View

E

(c,d,e)

S $\longrightarrow$ P

$\equiv$

E+S $\underset{d}{\overset{c}{\rightleftarrows}}$ ES $\overset{e}{\longrightarrow}$ P+E

intermediate complex

Interaction View

bind

unbind

react

E

a_c | u_d | k_e $\cdots\cdots>$ P

S

private bindings between one S and one E molecule

$S() \triangleq$ new $u@d$ new $k@e$
$\qquad !a_c(u,k); (!u_d; S() + !k_e; P())$

bind   unbind   react

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

$P() \triangleq \ldots$
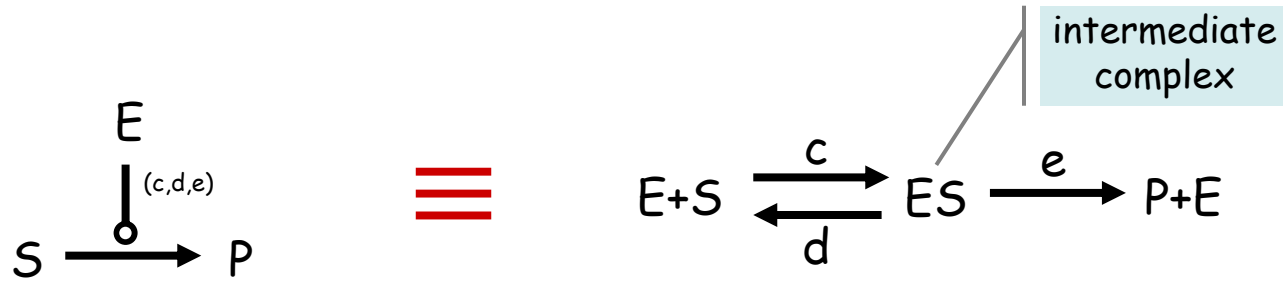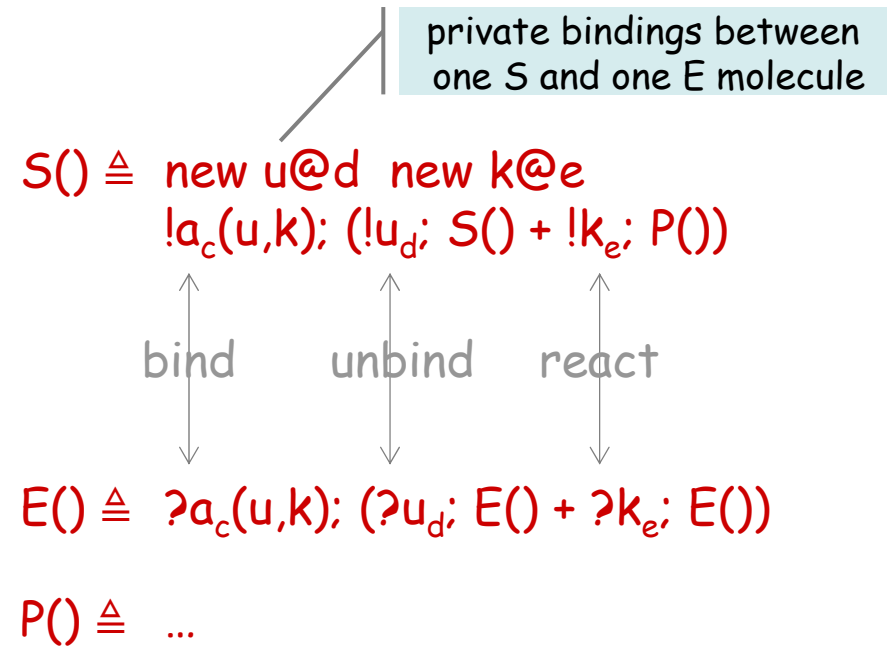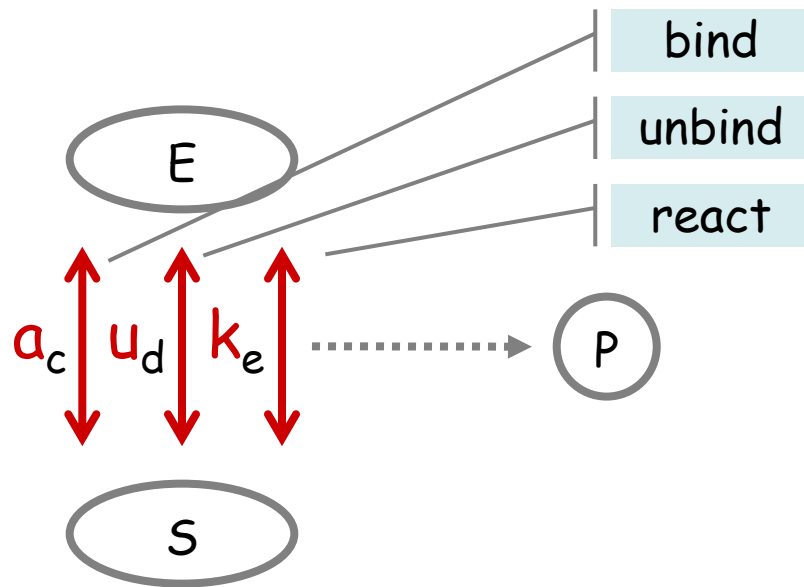
As 12 processes (in SPiM)

let KKK() =
  (new u1@d1:Release new k1@r1:React
  !a1(u1,k1); (do !u1;KKK() or !k1;KKKst()))   [1]substrate

KKK:E1 complex

and KKKst() =
  (new u2@d2:Release new k2@r2:React
  do !a2(u2,k2); (do !u2;KKKst() or !k2;KKK())   [2]substrate
  or ?a3(u3,k3); (do ?u3;KKKst() or ?k3;KKKst())   [3]kinase
  or ?a5(u5,k5); (do ?u5;KKKst() or ?k5;KKKst()))   [5]kinase

let E1() =
  ?a1(u1,k1); (do ?u1;E1() or ?k1;E1())   [1]enzyme

E1:KKK complex

let E2() =
  ?a2(u2,k2); (do ?u2;E2() or ?k2;E2())   [2]enzyme

let KK() =
  (new u3@d3:Release new k3@r3:React
  !a3(u3,k3); (do !u3;KK() or !k3;KK_P()))   [3]substrate

and KK_P() =
  (new u4@d4:Release new k4@r4:React
  new u5@d5:Release new k5@r5:React
  do !a4(u4,k4); (do !u4;KK_P() or !k4;KK())   [4]substrate
  or !a5(u5,k5); (do !u5;KK_P() or !k5;KK_PP()))   [5]substrate

and KK_PP() =
  (new u6@d6:Release new k6@r6:React
  do !a6(u6,k6); (do !u6;KK_PP() or !k6;KK_P())   [6]substrate
  or ?a7(u7,k7); (do ?u7;KK_PP() or ?k7;KK_PP())   [7]kinase
  or ?a9(u9,k9);   [9]kinase

One process for each component (12) including enzymes, but not for complexes.

and KKPse() =
  do ?a4(u4,k4); (do ?u4;KKPse() or ?k4;KKPse())   [4]phtase
  or ?a6(u6,k6); (do ?u6;KKPse() or ?k6;KKPse())   [6]phtase

let K() =
  (new u7@   No need for conservation equations: implicit in "choice" operator in the calculus.
  !a7(u7,k7)   [7]substrate

and K_P() =
  (new u8@d8:Release new k8@r8:React
  new u9@d9:Release new k9@r9:React
  do !a8(u8,k8); (do !u8;K_P() or !k8;K())   [8]substrate
  or !a9(u9,k9); (do !u9;K_P() or !k9;K_PP()))   [9]substrate

and K_PP() =
  (new u10@d10:Release new k10@r10:React
  !a10(u10,k10); (do !u10;K_PP() or !k10;K_P()))   [10]substrate

and KPse() =
  do ?a8(u8,k8); (do ?u8;KPse() or ?k8;KPse())   [8]phtase
  or ?a10(u10,k10); (do ?u10;KPse() or ?k10;KPse())   [10]phtase

… and 30 Interaction Channels

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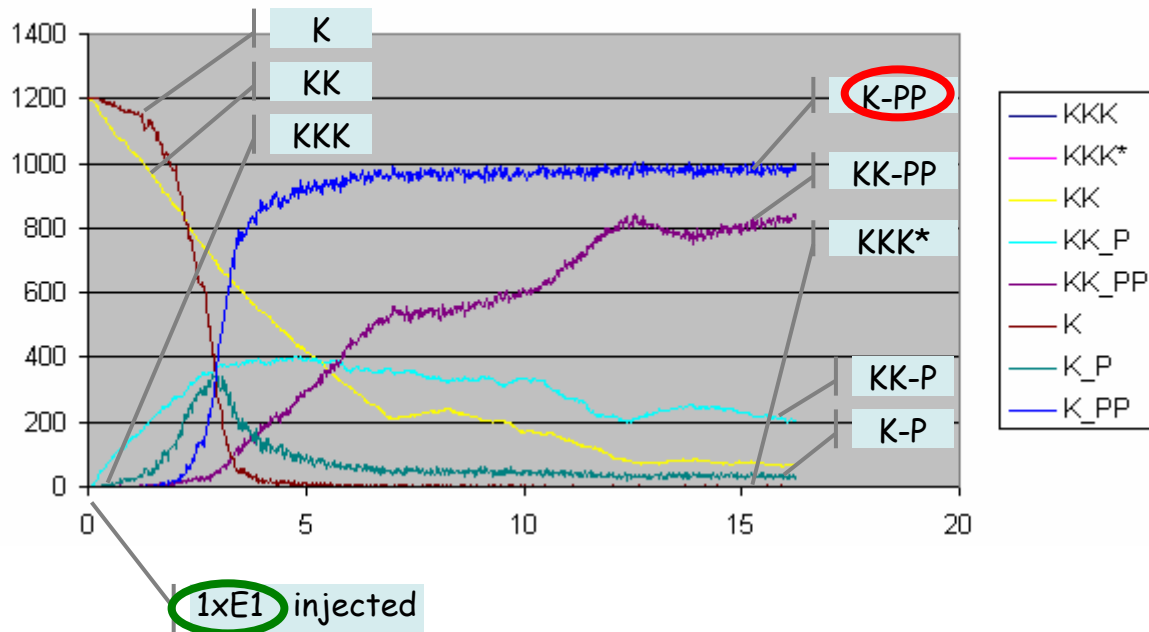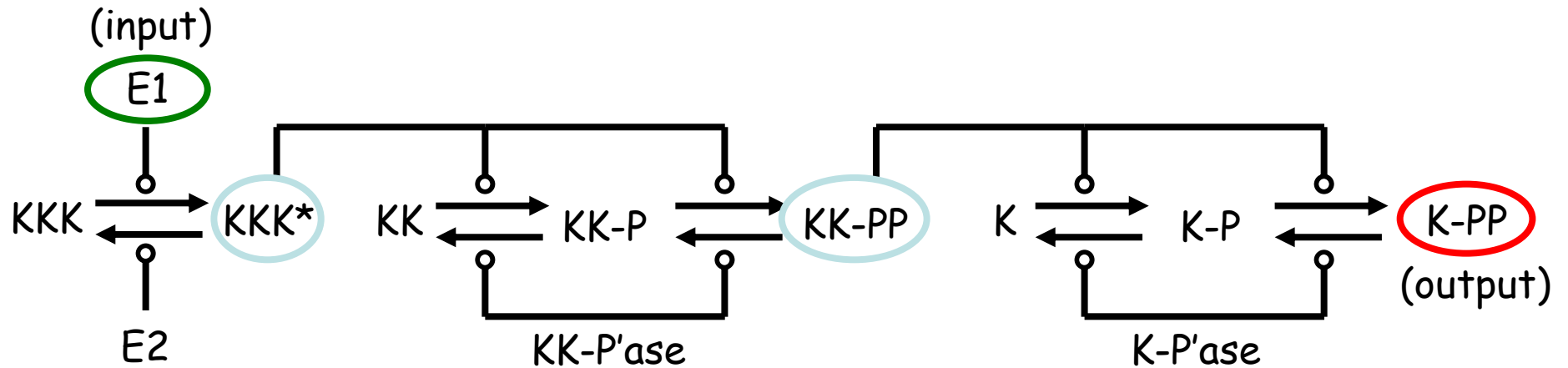type React = chan()
type Bond = chan(Release,React)

new a1@1.0:Bond val d1=1.0 val r1=1.0
new a2@1.0:Bond val d2=1.0 val r2=1.0
new a3@1.0:Bond val d3=1.0 val r3=1.0
new a4@1.0:Bond val d4=1.0 val r4=1.0
new a5@1.0:Bond val d5=1.0 val r5=1.0
new a6@1.0:Bond val d6=1.0 val r6=1.0
new a7@1.0:Bond val d7=1.0 val r7=1.0
new a8@1.0:Bond val d8=1.0 val r8=1.0
new a9@1.0:Bond val d9=1.0 val r9=1.0
new a10@1.0:Bond val d10=1.0 val r10=1.0

…

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()

$a_i(u_i,k_i)$: release (u_i@d_i) and react (k_i@r_i) channels passed over bond (a_i) channel. (No behavior attached to channels except interaction rate.)
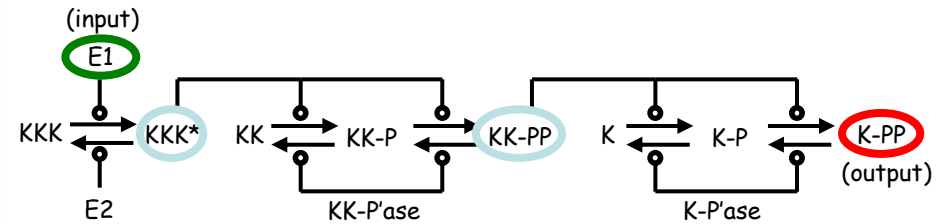
MAPK Cascade Simulation in SPiM

(input)

E1

KKK → KKK* KK ⇌ KK-P ⇌ KK-PP K ⇌ K-P ⇌ K-PP

E2 (output)

KK-P'ase K-P'ase

K
KK
K-PP
KKK
KK-PP
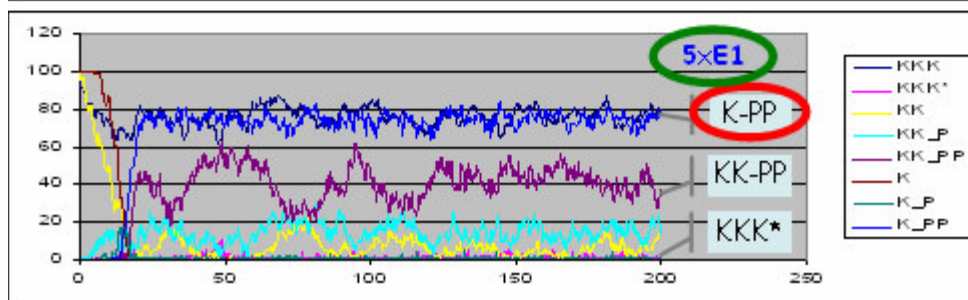KKK*
KK-P
K-P

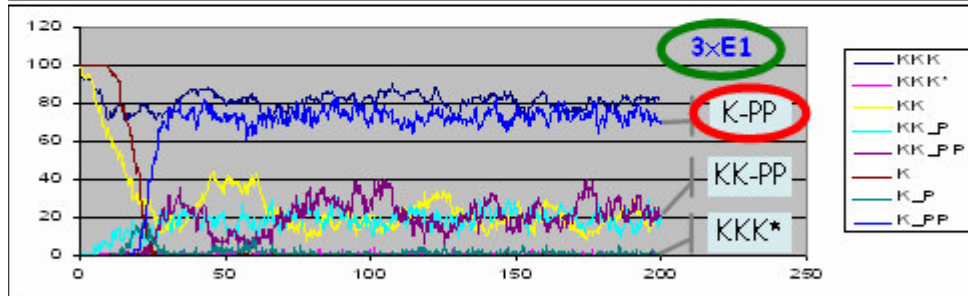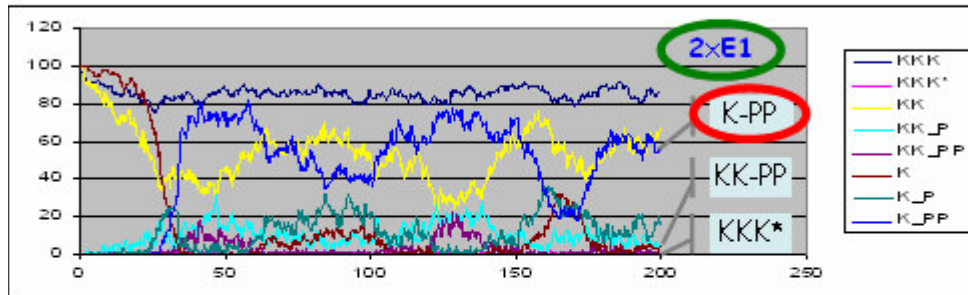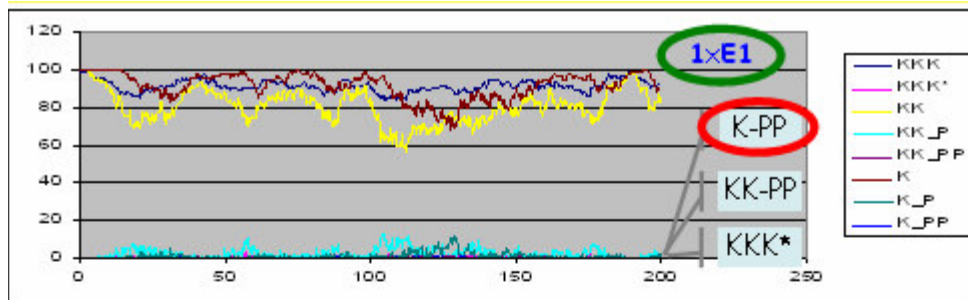KKK
KKK*
KK
KK_P
KK_PP
K
K_P
K_PP

1xE1 injected

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:

1xE2 (0.3 nM)
1xKKPase (0.3 nM)
120xKPase (120 nM)
3xKKK (3 nM)
1200xKK (1.2 uM)
1200xK (1.2 uM)

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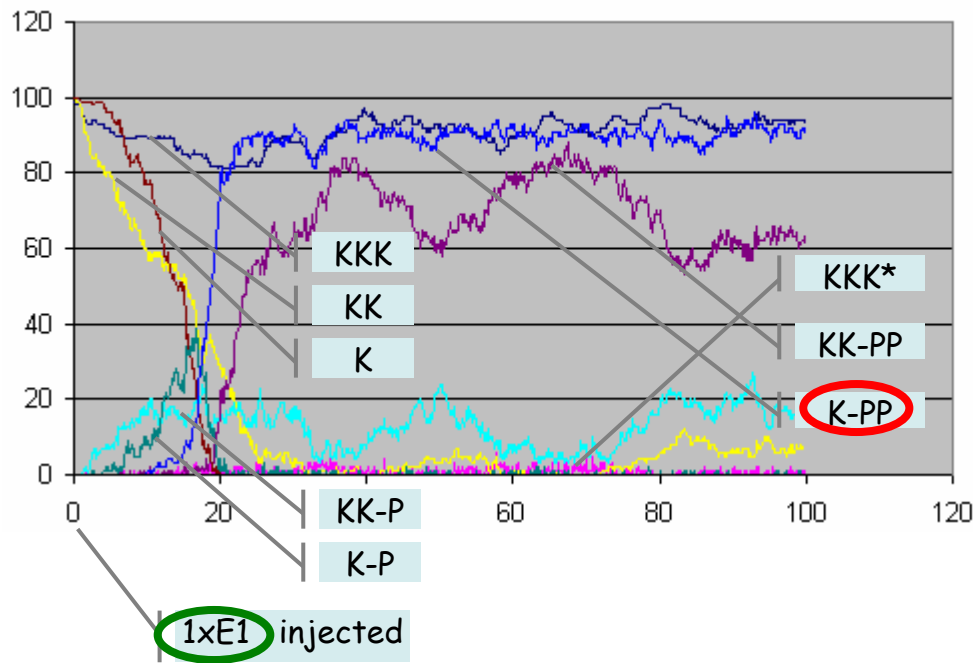(Kmx = (dx + rx) / ax, Km = 300 nM)

1xE1

MAPK Cascade Simulation in SPiM

All coefficients 1.0 !!!
100×KKK, 100×KK, 100×K,
13×E2, 13×KKPse, 13×KPse.
n×E1 as indicated
(1×E1 is not sufficient to produce an output)

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

All coefficients 1.0 !!!
100xKKK, 100xKK, 100xK,
5xE2, 5xKKPse, 5xKPse.

Input is 1xE1.
Output is 90xK-PP (ultrasensitivity).

2006-04-03   31

2. The Gene Machine

Pretty far from the atoms.

The "Central Dogma" of Molecular Biology
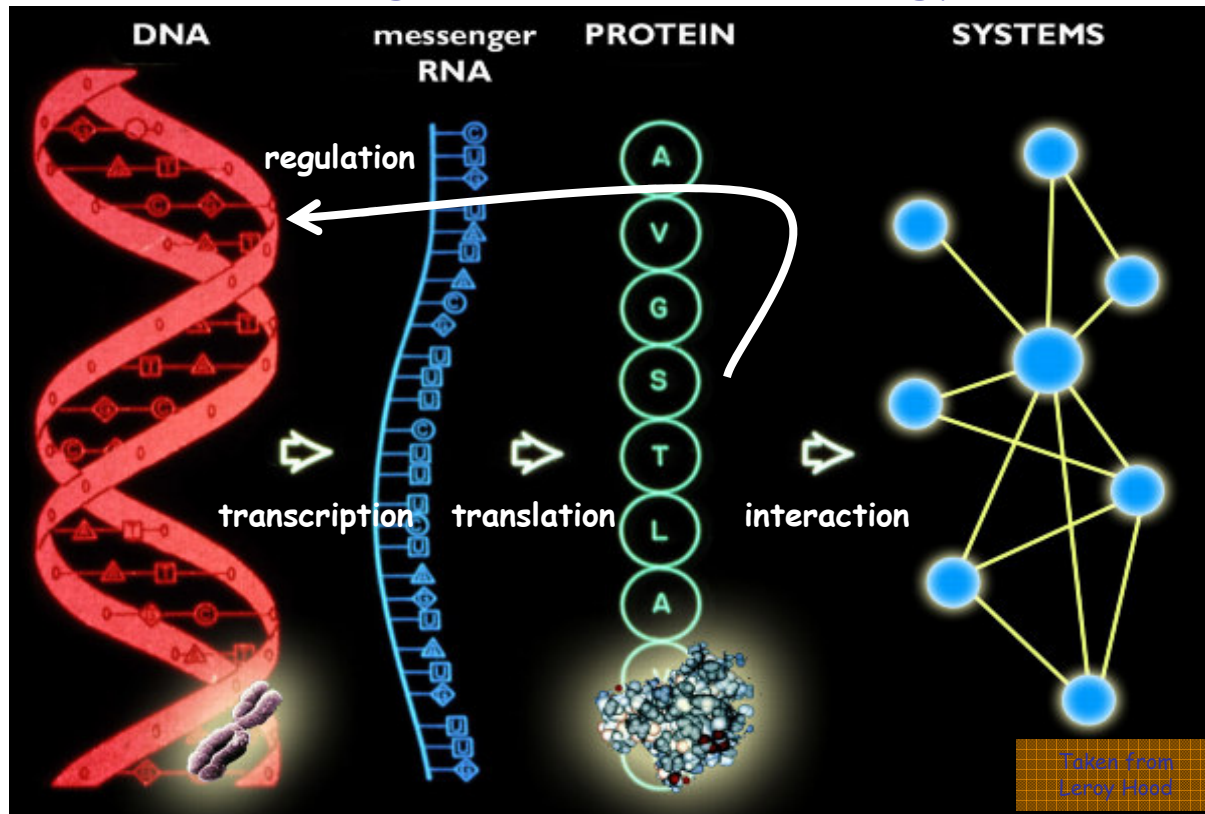
DNA	messenger RNA	PROTEIN	SYSTEMS
4-letter digital code	4-letter digital code	20-letter digital code	50.000(?) shapes

regulation — transcription — translation — interaction

Taken from Leroy Hood

DNA Tutorial

Metabolic space — Protein space — Gene space
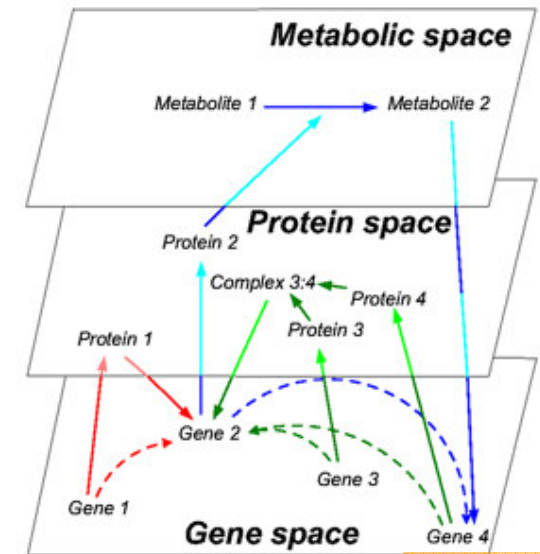
Taken from Pedro Mendes

Lactose Operon

The Gene Machine "Instruction Set"

Positive Regulation
Negative Regulation
Input
Transcription
Output
Gene (Stretch of DNA)
Coding region
Regulatory region
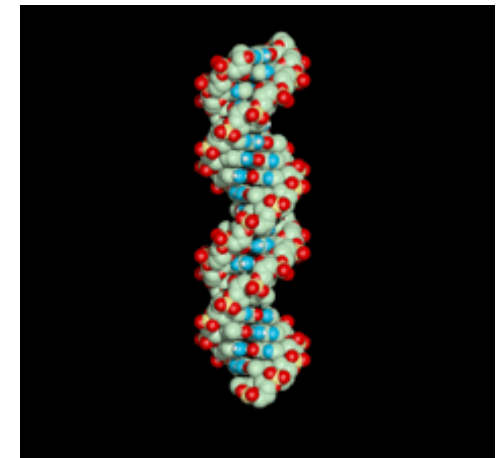
Output2
Input
Output1
"External Choice" The phage lambda switch
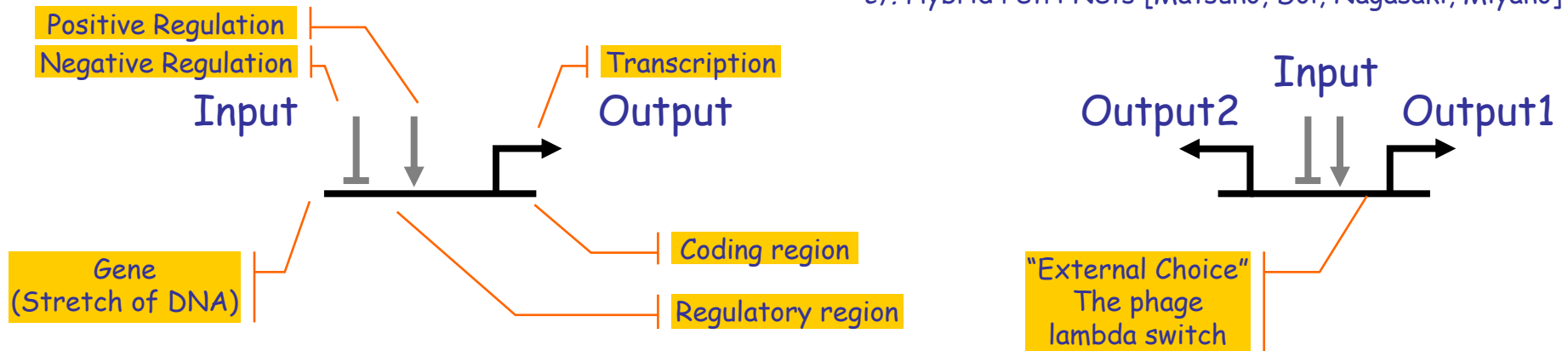
Regulation of a gene (positive and negative) influences transcription. The regulatory region has precise DNA sequences, but not meant for coding proteins: meant for binding regulators.

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
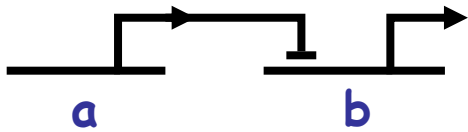
Is a shorthand for:

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.
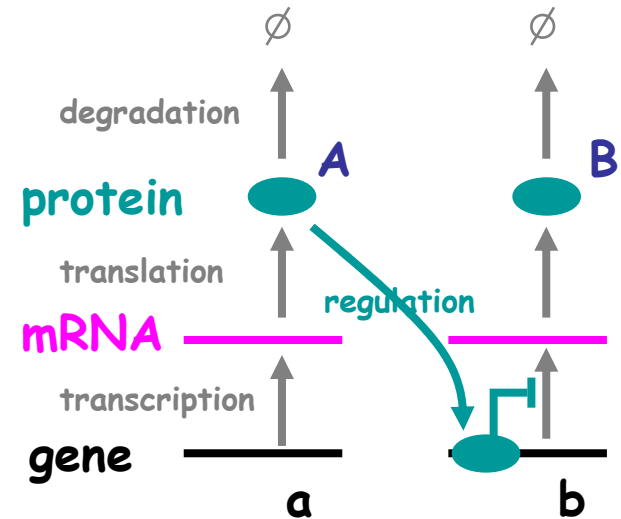
Ex: Bistable Switch

Ex: Oscillator

Expressed
Repressed
Expressing

2006-04-03   34

Indirect Gene Effects

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

(a) transcription factor, protein kinase, protein phosphatase, transcription factor

protein

DNA — Gene 1 — Gene 2 — Gene 3 — Gene 4 — Gene 5

(b)
Aspect of gene activity: mRNA expression
Genetic perturbation: gene deletion

G1:	G2, G5
G2:	G5
G3:	G5
G4:	G5
G5:	

(c)
Aspect of gene activity: phosphorylation state
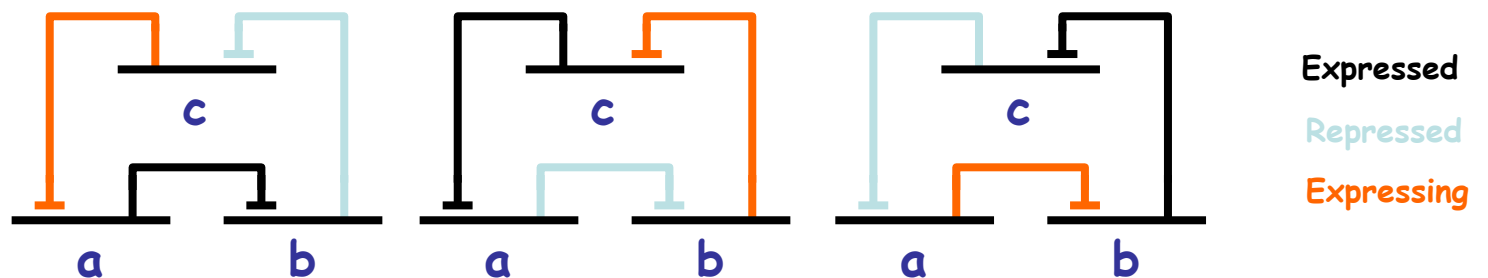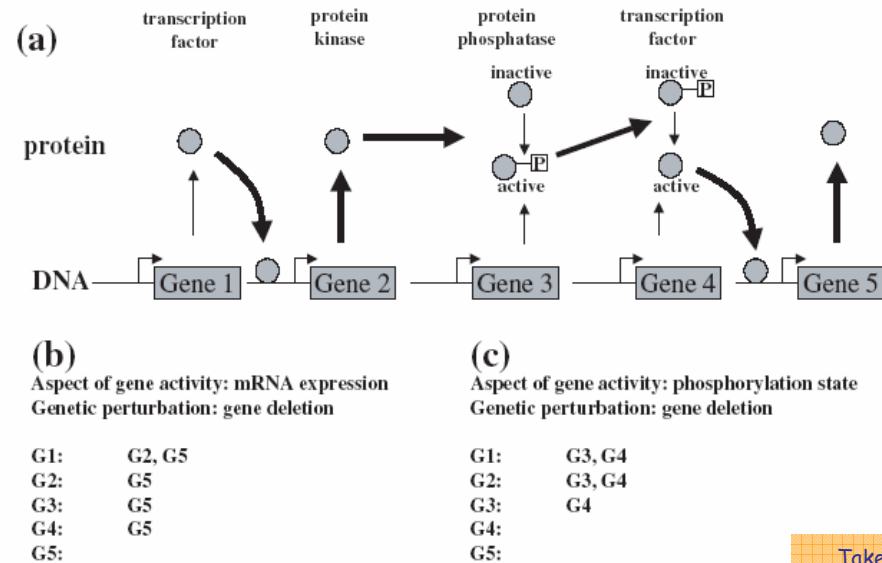Genetic perturbation: gene deletion

G1:	G3, G4
G2:	G3, G4
G3:	G4
G4:	
G5:	

Taken from Andreas Wagner

Fig. 1. The importance of specifiying 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 colon 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).

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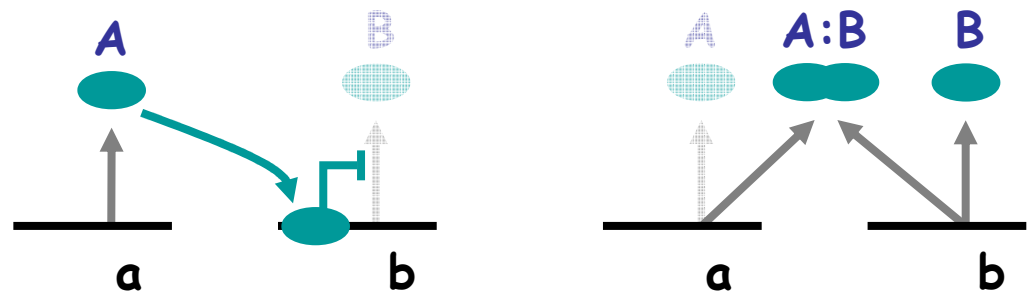
Structure of the Coding Region

The Central Dogma

DNA ——————

transcription

mRNA ——————

translation
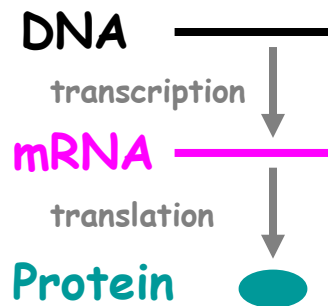
Protein ⬤

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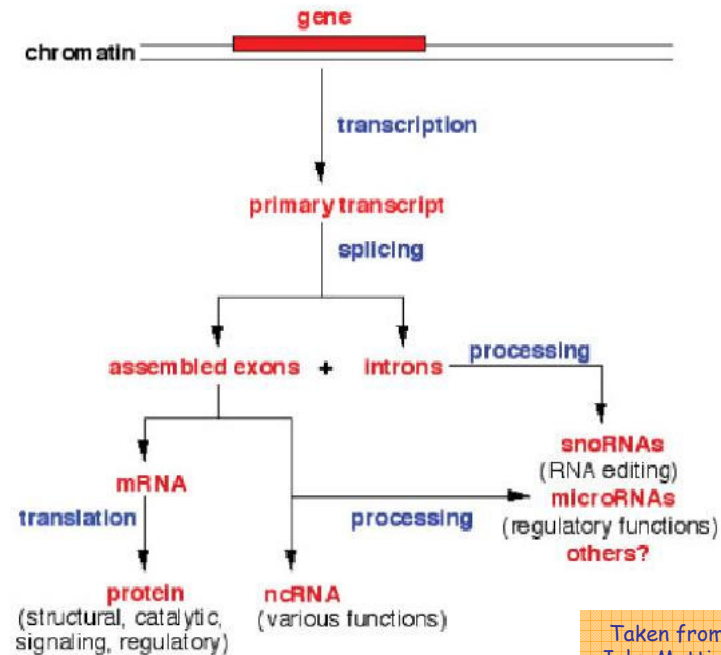- Fold-up like a protein
- Act like an enzyme
- Regulate other transcribed RNA
- Direct protein editing
– ...

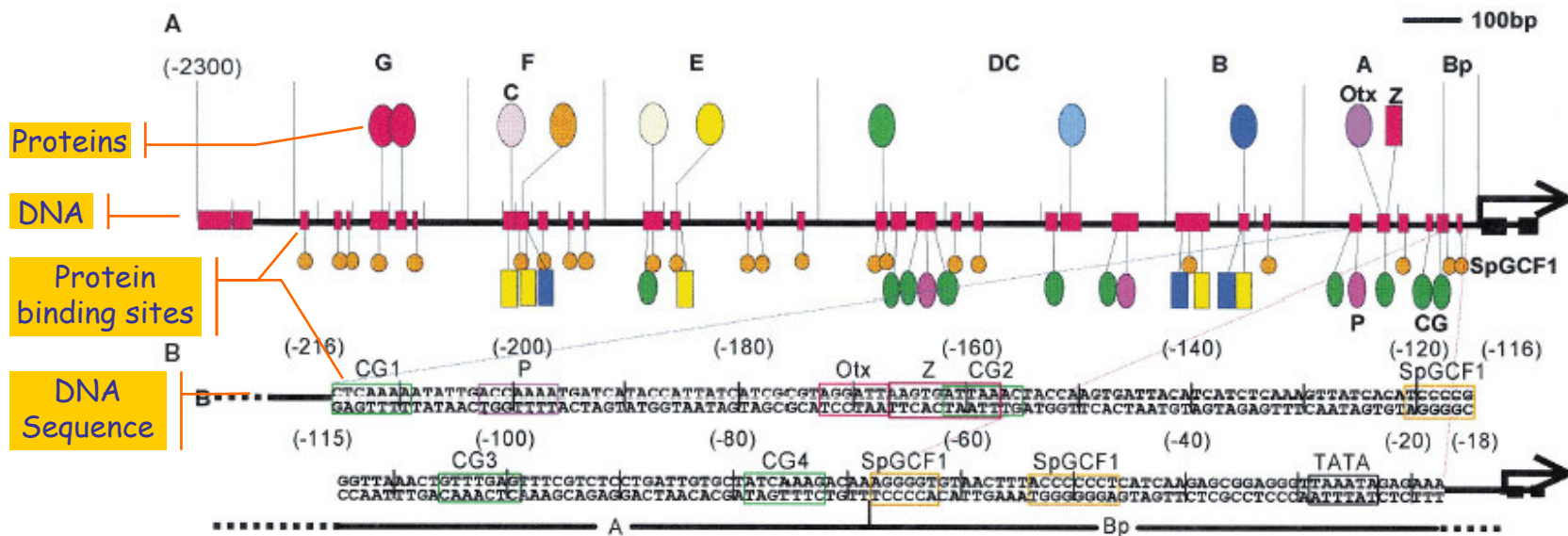## Challenging the Dogma (in higher organisms)

Taken from John Mattick

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

Proteins

DNA

Protein binding sites

DNA Sequence

2300bp!
> average protein

A

(-2300) G C F E DC B A Bp 100bp

Otx Z

SpGCF1

B

(-216) (-200) (-180) (-160) (-140) (-120) (-116)
CG1 P Otx Z CG2 SpGCF1

CTCAAAAATATTGACCAAAATGATCATACCATTATCATCGCGTAGGATTAAGTGATTAAACTACCAAGTGATTACATCATCTCAAAGTTATCACATCCCCG
GAGTTTTTATAACTGGTTTTACTAGTATGGTAATAGTAGCGCATCCTAATTCACTAATTTGATGGTTCACTAATGTAGTAGAGTTTCAATAGTGTAGGGGC

(-115) (-100) (-80) (-60) (-40) (-20) (-18)
CG3 CG4 SpGCF1 SpGCF1 TATA

GGTTAAACTGTTTGAGTTTCGTCTCCTGATTGTGCTATCAAAGACAAAGGGGTGTAACTTTACCCCCCTCATCAAGAGCGGAGGGTTAAATAGAGAAA
CCAATTTGACAAACTCAAAGCAGAGGACTAACACGATAGTTTCTGTTTCCCCACATTGAAATGGGGGGAGTAGTTCTCGCCTCCCAATTTATCTCTTT

P CG SpGCF1

A Bp

C 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):

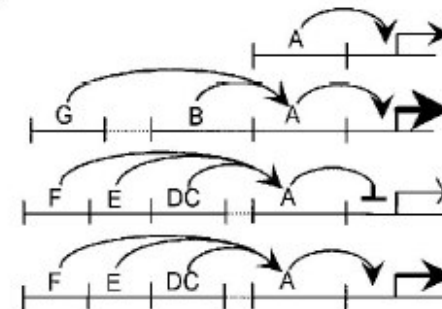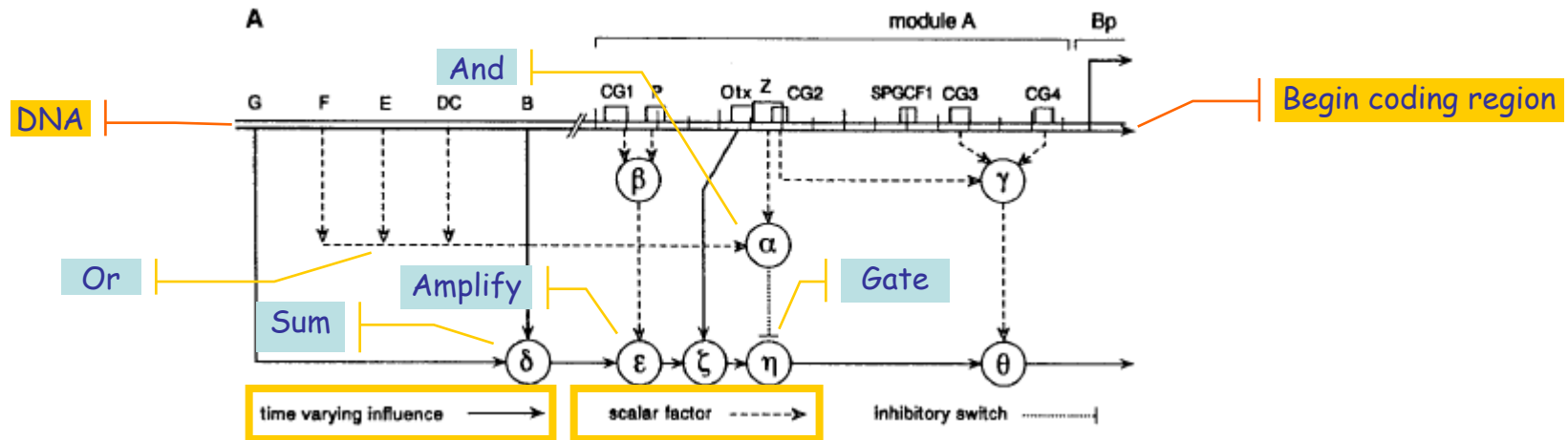Modules E, F and DC with LiCl treatment:

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 CG$_3$ and CG$_4$ sites as well as Bp has no endoderm-

specific activity and services 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 F, E, and DC. These functions are repression of expression in nonendodermal domains and enhancement of expression in response to LiCl.
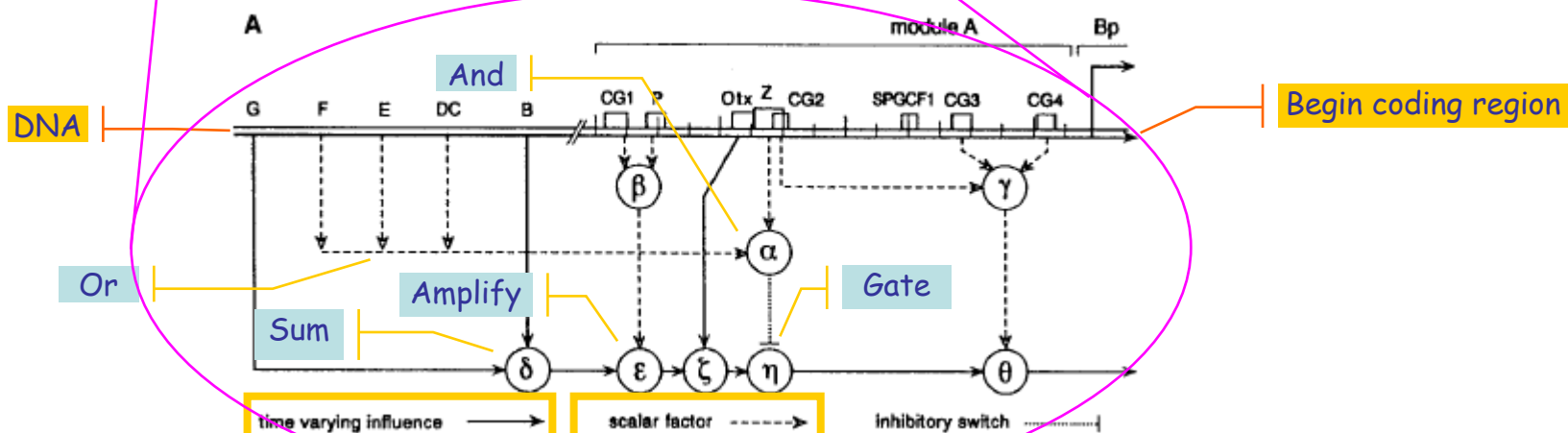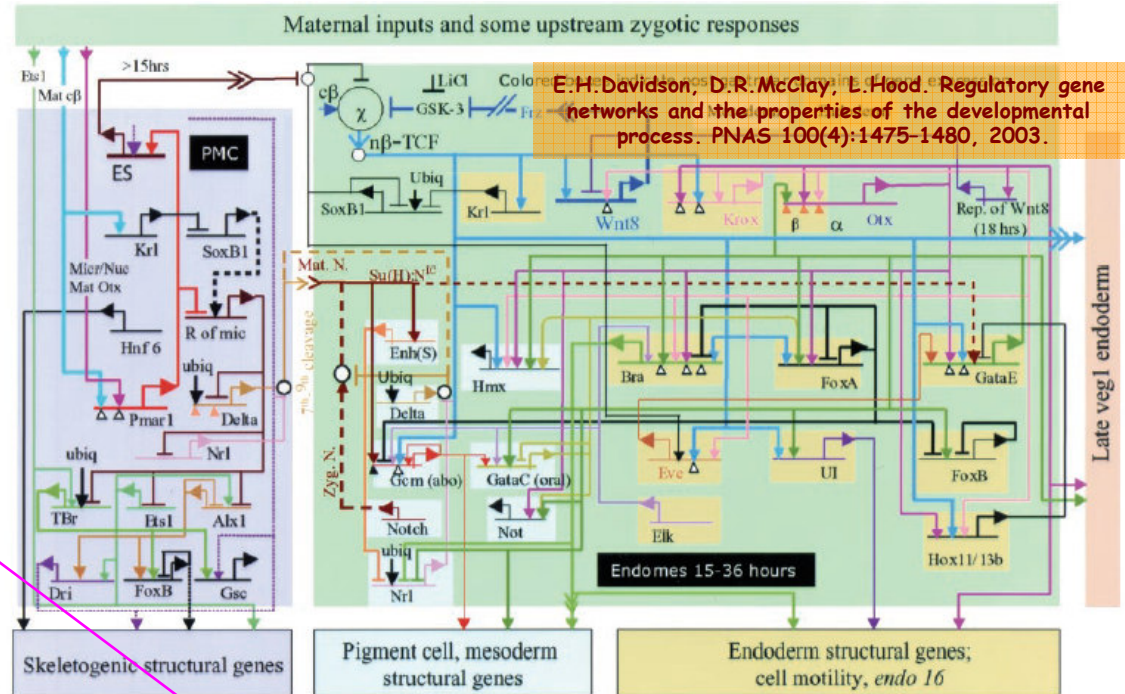
Function of a Regulatory Region

A

module A — Bp

DNA

Or, Sum, And, Amplify, Gate

Begin coding region

time varying influence → scalar factor ----→ inhibitory switch ···········

B

if (F = 1 or E = 1 or CD = 1) and (Z = 1) Repression functions of modules F, E, and DC mediated by Z site
 α = 1
else α = 0

if (P = 1 and CG₁ = 1) Both P and CG₁ needed for synergistic link with module B
 β = 2
else β = 0

if (CG₂ = 1 and CG₃ = 1 and CG₄ = 1) Final step up of system output
 γ = 2
else γ = 1

δ(t) = B(t) + G(t) Positive input from modules B and G

ε(t) = β*δ(t) Synergistic amplification of module B output by CG₁-P subsystem

if (ε(t) = 0) Switch determining whether Otx site in module A, or upstream modules (i.e., mainly module B), will control level of activity
 ξ(t) = Otx(t)
else ξ(t) = ε(t)

if (α = 1) Repression function inoperative in endoderm but blocks activity elsewhere
 η(t) = 0
else η(t) = ξ(t)

Θ(t) = γ*η(t) Final output communicated to BTA

C-H.Yuh, H.Bolouri, E.H.Davidson. Genomic Cis-Regulatory Logic: Experimental and Computational Analysis of a Sea Urchin Gene. Science 279:1896-1902, 1998

Gene Regulatory Networks

NetBuilder

E.H.Davidson, D.R.McClay, L.Hood. Regulatory gene networks and the properties of the developmental process. PNAS 100(4):1475-1480, 2003.

Maternal inputs and some upstream zygotic responses

Skeletogenic structural genes

Pigment cell, mesoderm structural genes

Endoderm structural genes; cell motility, *endo 16*

DNA

And

Or

Sum

Amplify

Gate

Begin coding region

C-H.Yuh, H.Bolouri, E.H.Davidson. Genomic Cis-Regulatory Logic: Experimental and Computational Analysis of a Sea Urchin Gene. Science 279:1896-1902, 1998

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

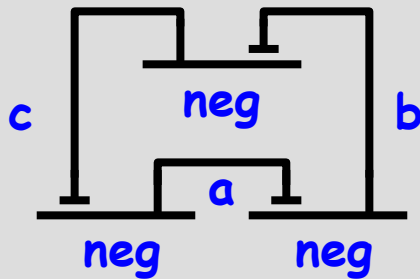- Mixed Gene-Protein Diagrams

Gene Gates and Circuits

A gene gate

a ⊣⊢→ b

neg
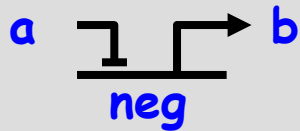
$neg(a,b) \triangleq$
$?a_r; \tau_\eta; neg(a,b) +$
$\tau_\varepsilon; (tr(b) \mid neg(a,b))$

$tr(p) \triangleq (!p_r; tr(p)) + \tau_\delta$

A genetic circuit (engineered in E.Coli)

c   neg   b

neg   a ⊣ neg

neg   neg

$neg(a,b) \mid$
$neg(b,c) \mid$
$neg(c,a)$

The stochastic-π program

```
val dk  = 0.001    (* Decay rate *)
val inh = 0.001    (* Inhibition rate *)
val cst = 0.1      (* Constitutive rate *)

let tr(p:chan()) =
    do !p; tr(p) or delay@dk

let neg(a:chan(), b:chan()) =
    do ?a; delay@inh; neg(a,b)
    or delay@cst; (tr(b) | neg(a,b))

(* The circuit *)
val bnd = 1.0        (* Protein binding rate *)
new a@bnd:chan() new b@bnd:chan() new c@bnd:chan()
run (neg(c,a) | neg(a,b) | neg(b,c))
```
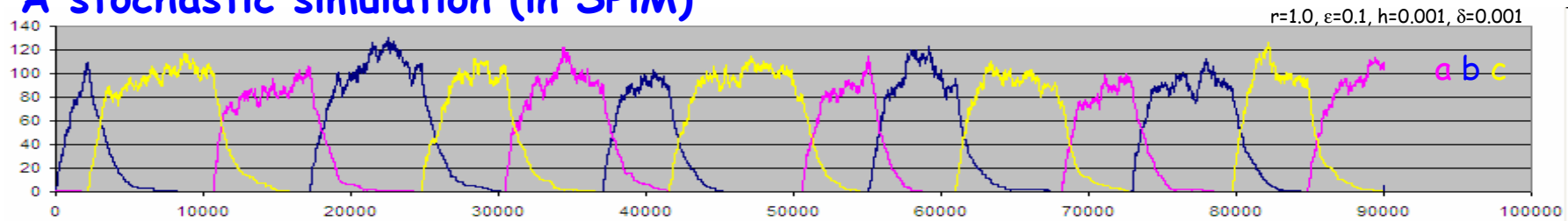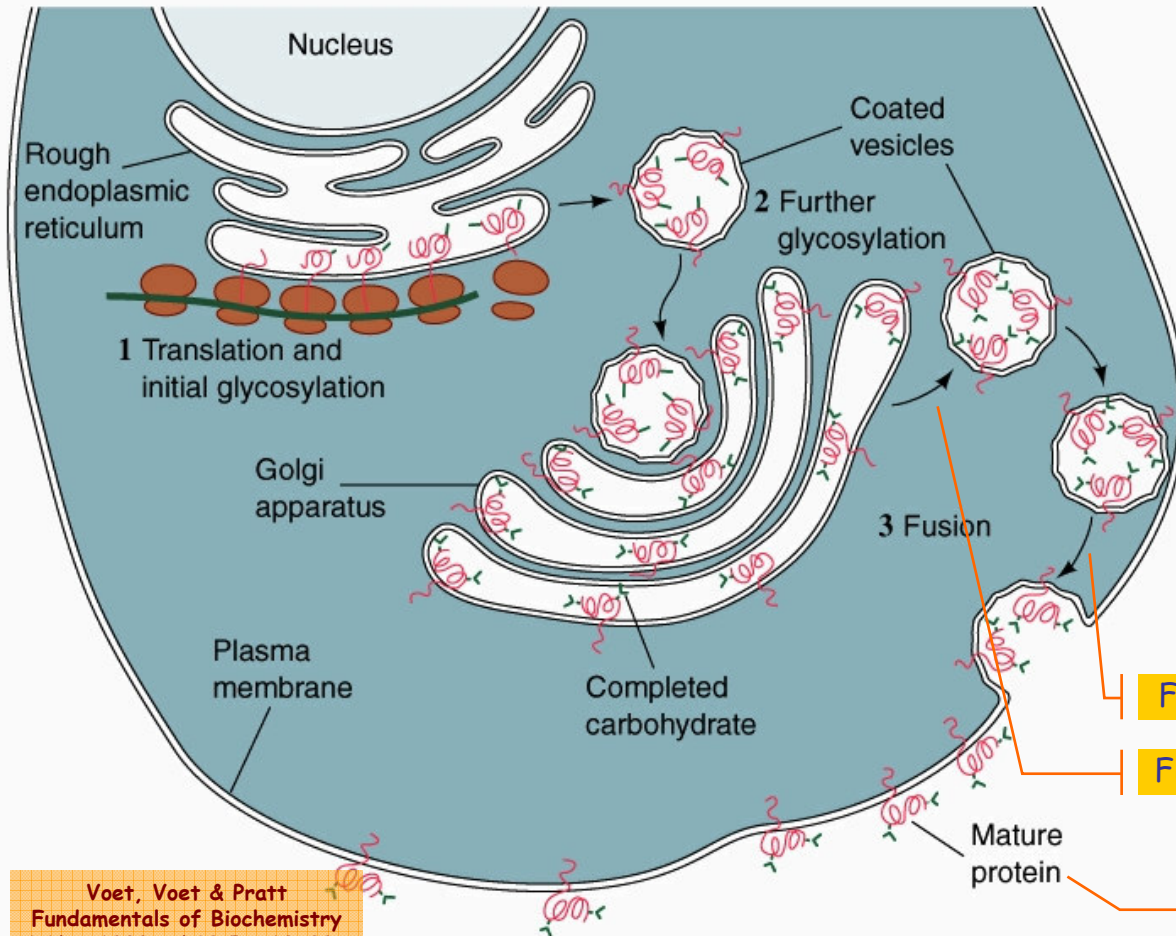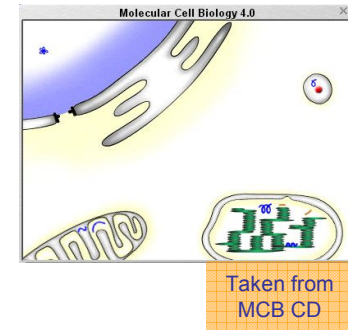
A stochastic simulation (in SPiM)

r=1.0, ε=0.1, h=0.001, δ=0.001

a b c

3. The Membrane Machine
Very far from the atoms.

Nucleus

Rough endoplasmic reticulum

Coated vesicles

2 Further glycosylation

1 Translation and initial glycosylation

Golgi apparatus

3 Fusion

Plasma membrane

Completed carbohydrate

Mature protein

Molecular transport and transformation through dynamic compartment fusion and fission.

Molecular Cell Biology 4.0

Taken from MCB CD

Fusion

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]

Membrane Fusion

Cell membrane

Virus membrane

1 **2** **3**

4 **5** **6**
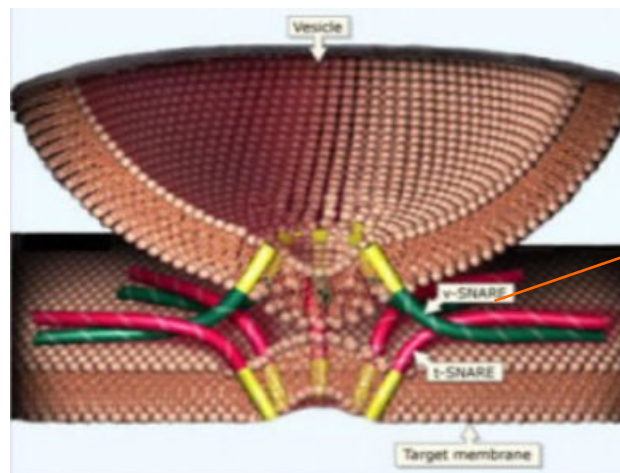
Taken from Tamm Laboratory
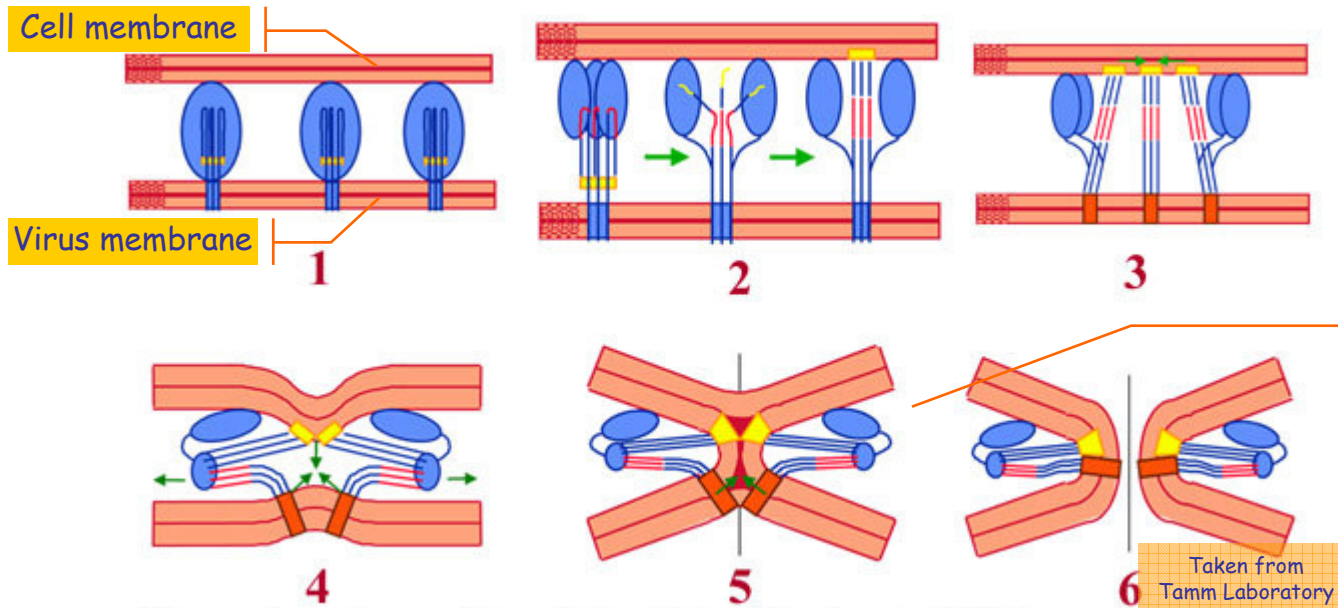
Proposed sequence of events in pH sensitive hemagglutinin membrane fusion

Aggressive fusion (virus)

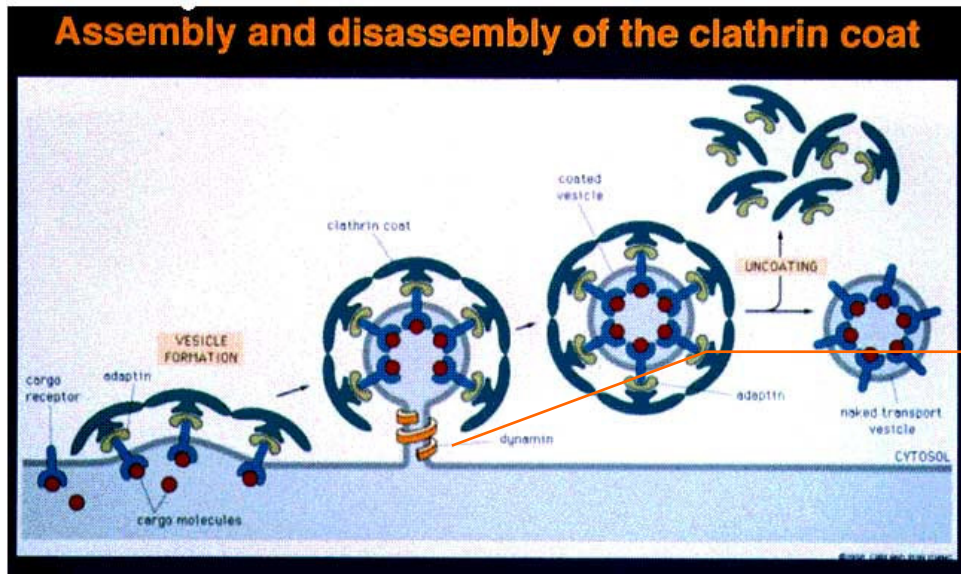By unknown mechanisms, the exoplasmic leaflets of the two membranes fuse" [MCB p745]

Cooperative fusion (vesicle)

Vesicle

v-SNARE

t-SNARE

Target membrane

"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

coated vesicle

clathrin coat

UNCOATING

VESICLE FORMATION

cargo receptor

adaptin

adaptin

naked transport vesicle

cargo molecules
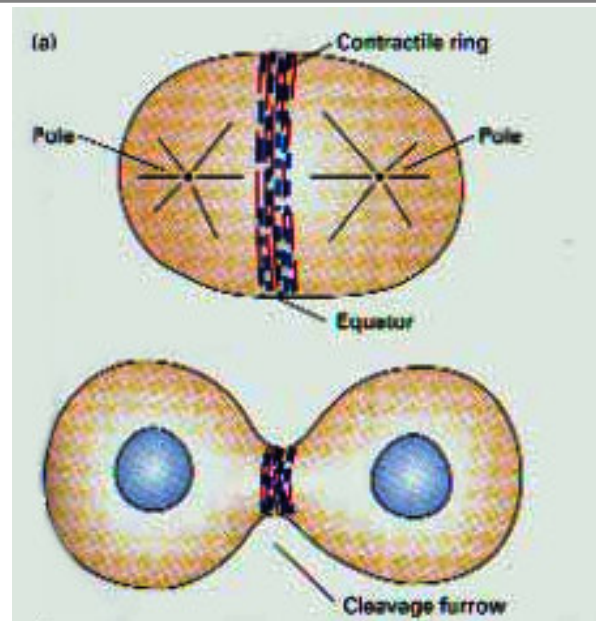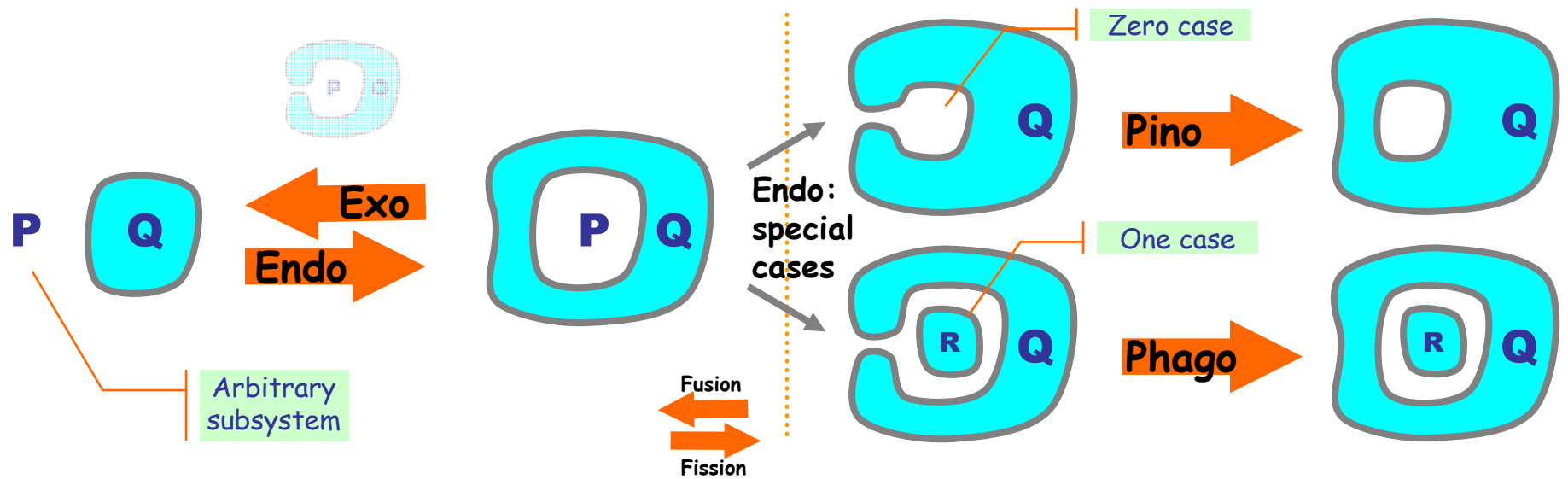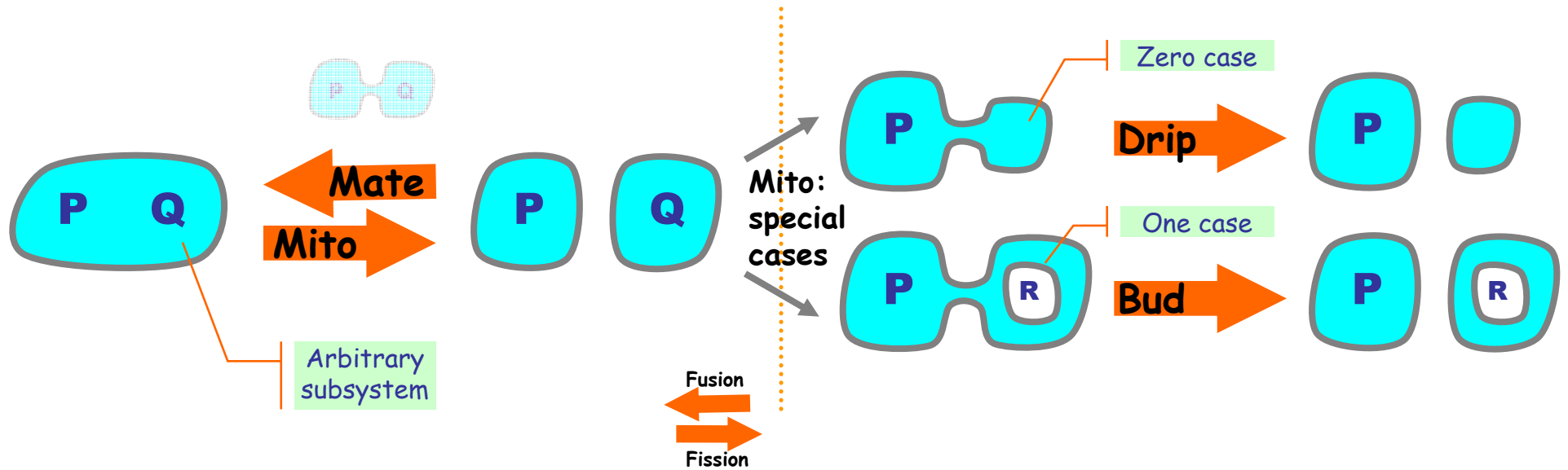
dynamin

CYTOSOL

Vesicle Formation

Movie by Allison Bruce

"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]
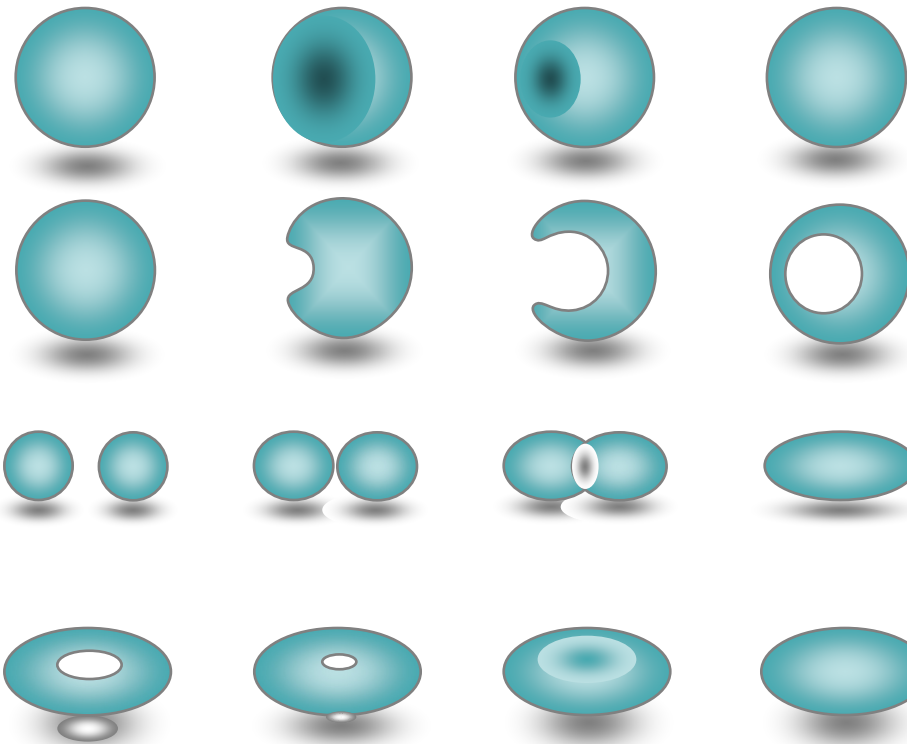

(a) Contractile ring

Pole

Pole

Equator

Cleavage furrow

Cytokinesis (Mitosis)

The Membrane Machine "Instruction Set"

Mate / **Mito** — Arbitrary subsystem

P Q ⟷ P Q

Mito: special cases

Zero case: P — **Drip** — P

One case: P R — **Bud** — P R

Fusion / Fission

Exo / **Endo** — Arbitrary subsystem

P Q ⟷ P Q

Endo: special cases

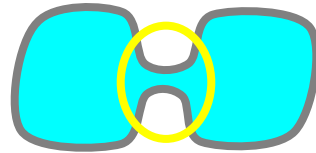Zero case: Q — **Pino** — Q
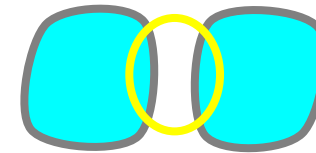
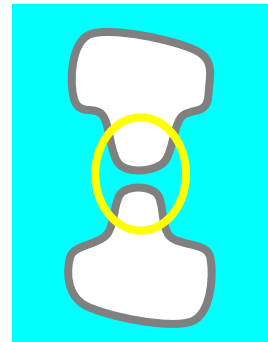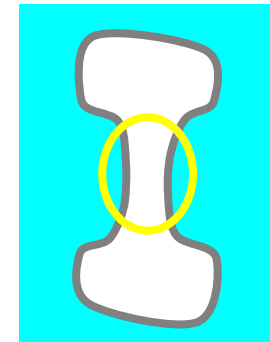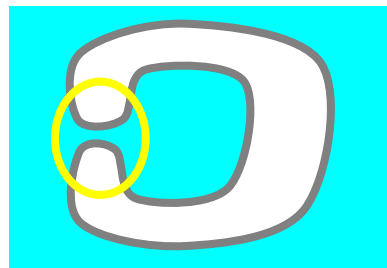One case: R Q — **Phago** — R Q

Fusion / Fission

… in 3D

Locally Implementable!

Global Views

Mito → (Fission)

Mate → (Fusion)
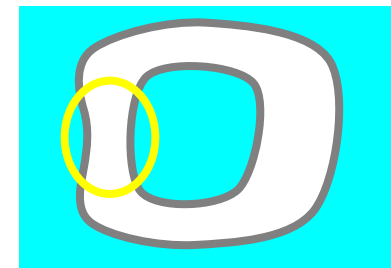
Endo → (Fission)

Exo → (Fusion)

Switch
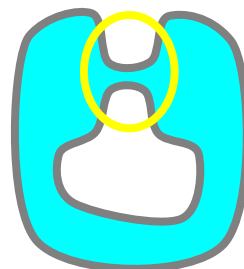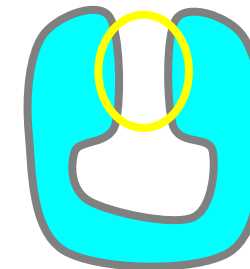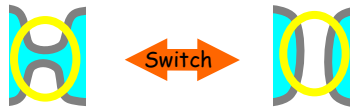
Same Local View!

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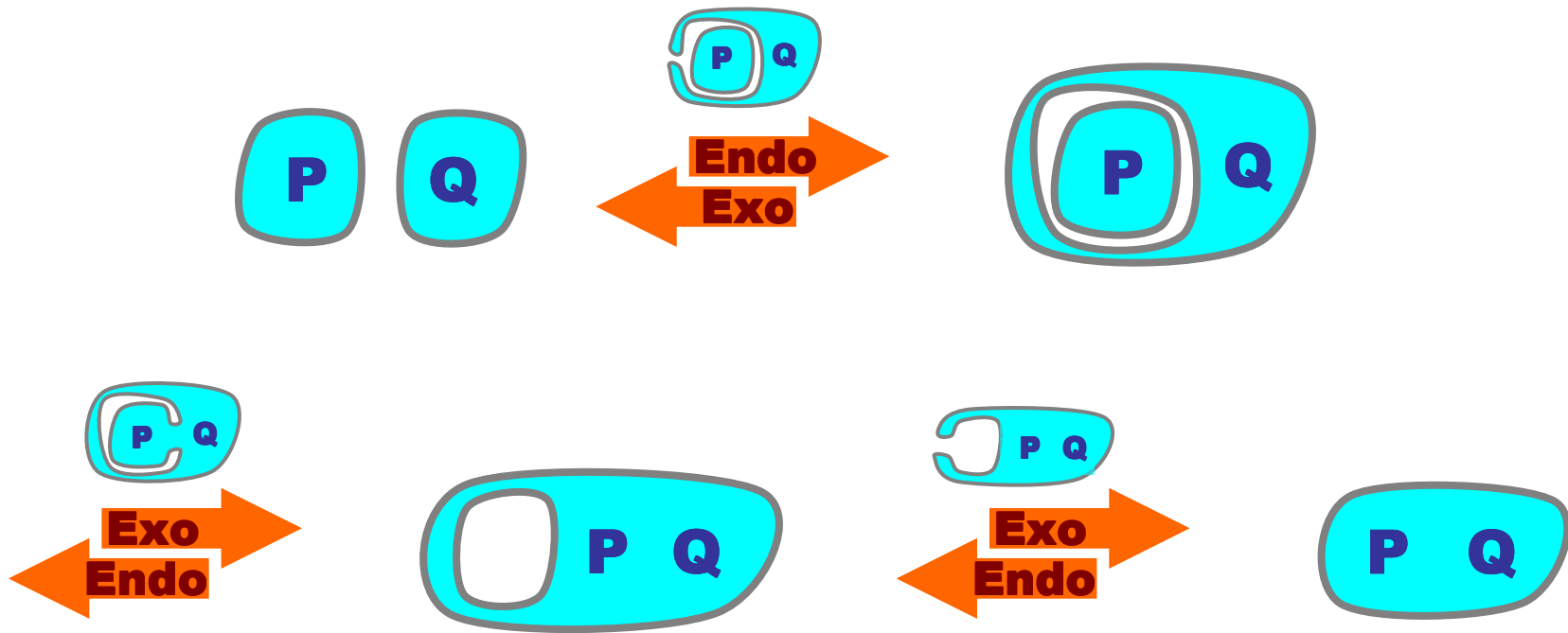
Notations for the Membrane Machine

- "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

Nucleus

Rough endoplasmic reticulum

Coated vesicles

2 Further glycosylation

1 Translation and initial glycosylation

Golgi apparatus

3 Fusion

Plasma membrane

Completed carbohydrate

Mature protein

Copyright 1999 John Wiley and Sons, Inc. All rights reserved.
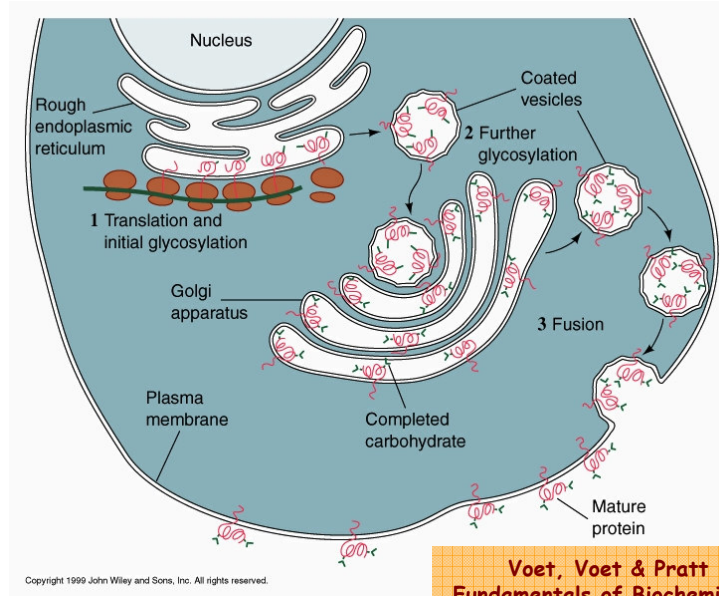
Voet, Voet & Pratt
Fundamentals of Biochemistry
Wiley 1999. Ch10 Fig 10-22.

LDL-Cholesterol Degradation

Cholesterol esters
Phospholipid monolayer
Apo-B protein
LDL particle
LDL receptor
Clathrin
Plasma membrane
Coated pit
Coated vesicle
Clathrin triskelions
Uncoated vesicle (early endosome)
Sorting vesicle (late endosome)
Separation of LDL particle from receptor due to lowered pH
Recycling of receptors to plasma membrane
pH = 5.0
Transport vesicle
Degradation of LDL to:
Amino acids
Cholesterol
Fatty acids

H.Lodish et al.
Molecular Cell Biology.
fourth Edition p.730.

Viral Replication

Virus
RNA
Capsid
Membrane
Envelope protein
} Nucleocapsid
Phago
Cytosol
Assembly
Translation
Budding
Mate
Disassembly
RNA
Exo
RNA Replication
Translation
Bud
Endosome
Nucleus
Endoplasmic Reticulum
(via Golgi)
Exo
Drip
Vesicle
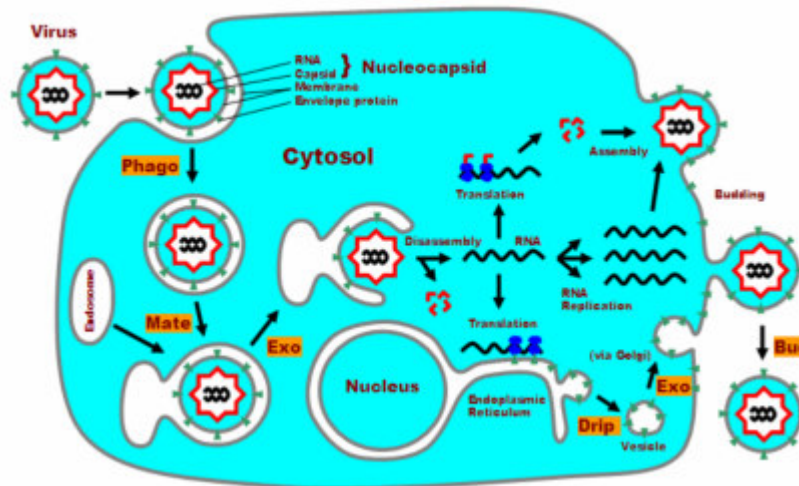
Adapted from: B.Alberts et al.
Molecular Biology of the Cell
third edition p.279.

Abstract Machines of Systems Biology

The "hardware" (biochemistry) is fairly well understood.
But what is the "software" that runs on these machines?

Regulation

Gene Machine
Nucleotides

Gene Regulatory Networks

Functional Architecture
Diverse
- chemical toolkits
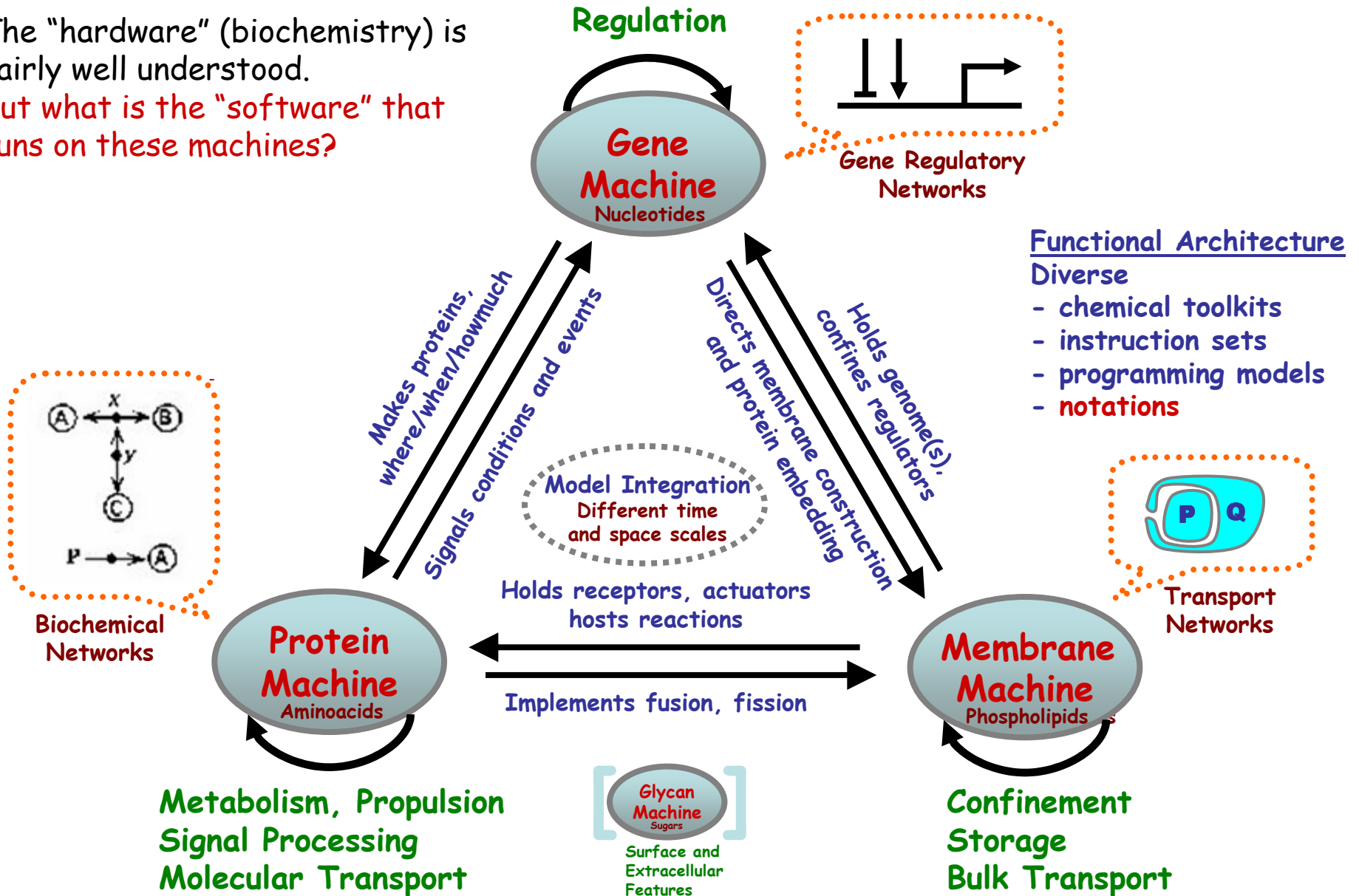- instruction sets
- programming models
- notations

Makes proteins, where/when/howmuch

Signals conditions and events

Directs membrane construction and protein embedding

Holds genome(s), confines regulators

Model Integration
Different time and space scales

Holds receptors, actuators hosts reactions

Biochemical Networks

Transport Networks

Protein Machine
Aminoacids

Implements fusion, fission

Membrane Machine
Phospholipids

P   Q

Metabolism, Propulsion
Signal Processing
Molecular Transport

Glycan Machine
Sugars

Surface and Extracellular Features

Confinement
Storage
Bulk Transport

Conclusions
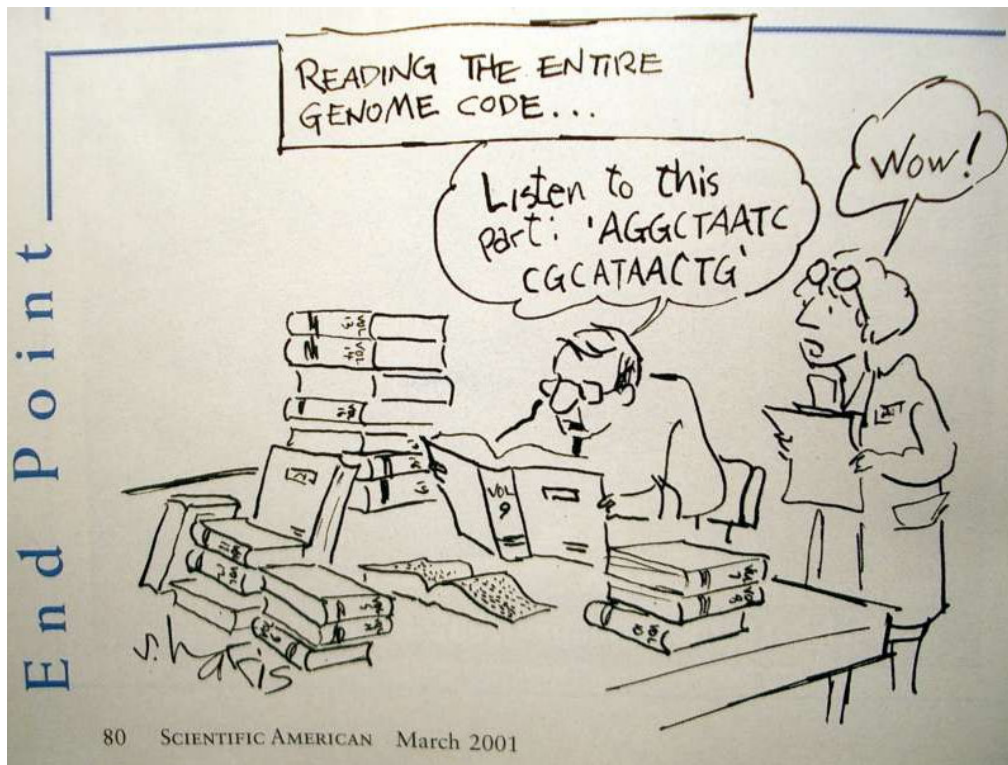
Q: "The data are accumulating and the computers are humming, what we are lacking are the words, the grammar and the syntax of a new language..."

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

[MCB] Molecular Cell Biology, Freeman.

[MBC] Molecular Biology of the Cell, Garland.

[Ptashne] A Genetic Switch.

[Davidson] Genomic Regulatory Systems.

[Milner] Communicating and Mobile Systems: the Pi-Calculus.

[Regev] Computational Systems Biology: A Calculus for Biomolecular Knowledge (Ph.D. Thesis).

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