Languages & Notations for Systems Biology

Luca Cardelli

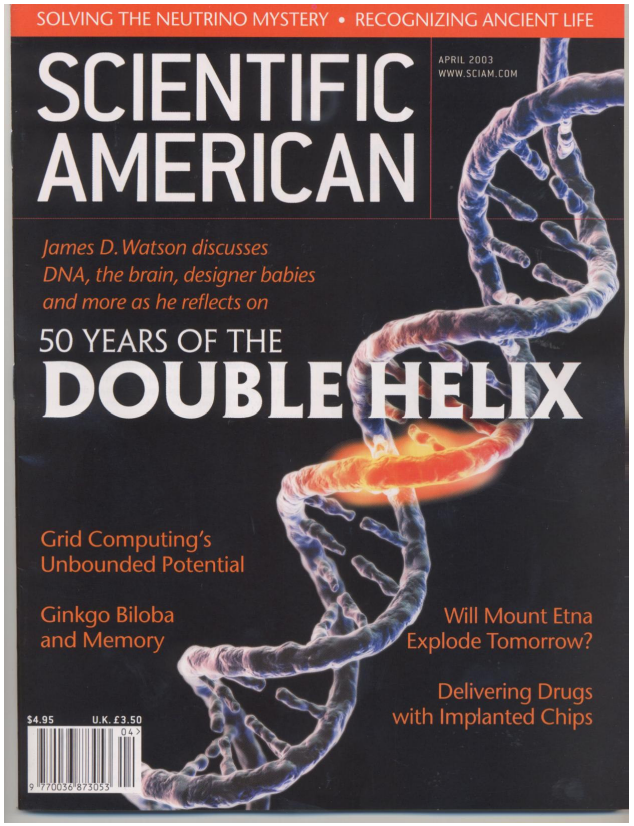
Microsoft Research
Cambridge UK

2004-12-14 Pisa
Galileo Galilei Colloquium

www.luca.demon.co.uk

50 Years of Molecular Cell Biology

- How cells work:
  - DNA stores information
  - DNA instructs Ribosomes to assemble Proteins
  - Proteins (>10000) do things:
    - Process signals, activate DNA
    - Catalyze reactions to produce substances
    - Control energy production and consumption
  - Bootstrapping still a mystery
    - Happened a long time ago; not understood, not essential.



Towards Systems Biology

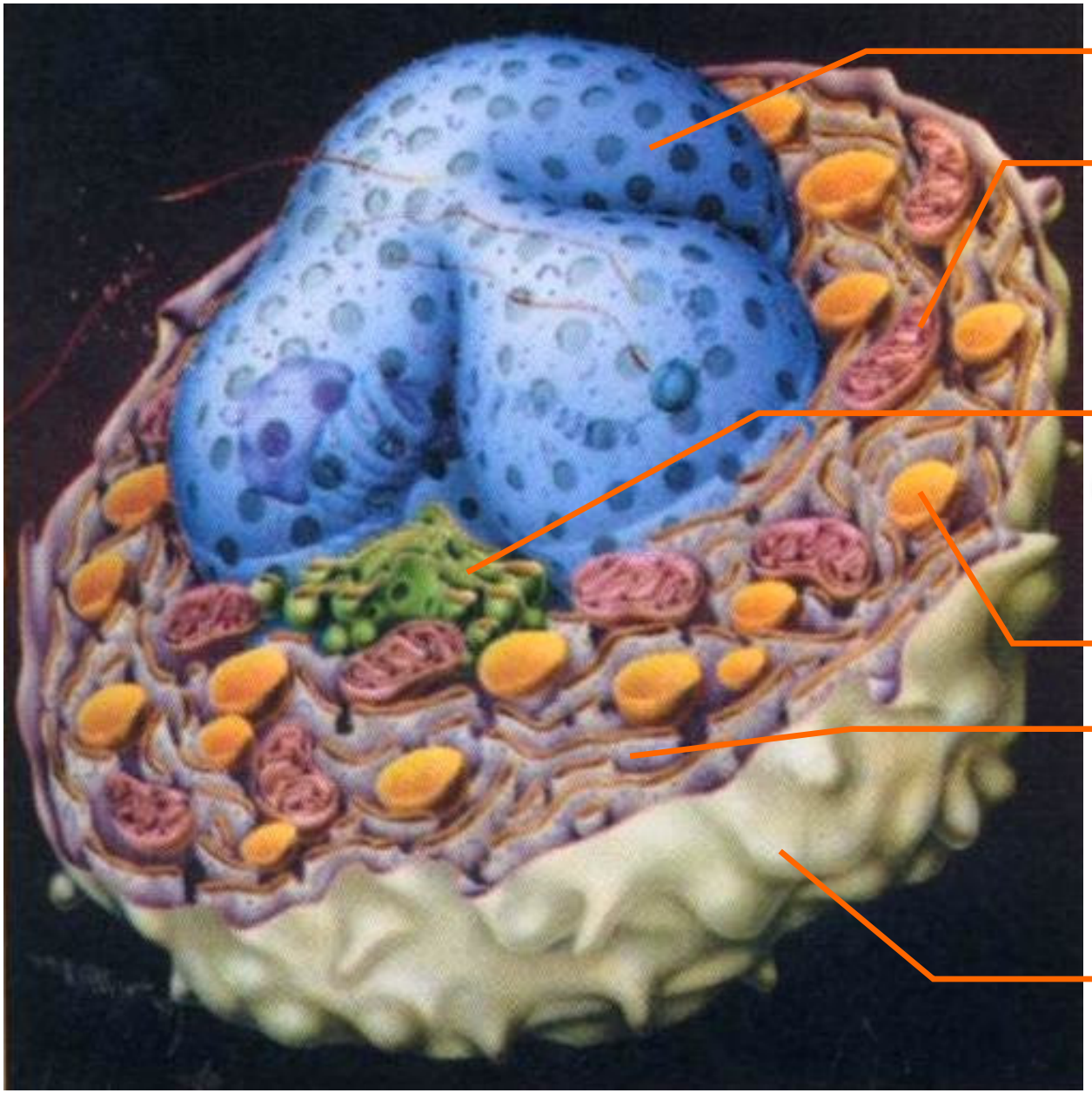
- Biologists now understand many of the cellular components, but do not yet understand how "the system" works.
  - Predictive biology and pharmacology still rare.
  - Synthetic biology still unreliable.
  - Massive data gathering and mining in progress (e.g. Genome projects); much yet to be understood.
- What kind of a system?
  - Based on digital information (DNA).
  - But how is information structured? How is it used?
  - How complex is the system?
  - Can we fix it when it breaks?

Structural Architecture

Eukaryotic Cell

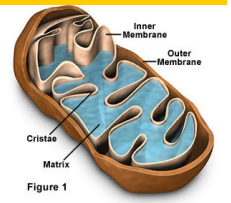
(10~100 trillion in human body)

Membranes everywhere

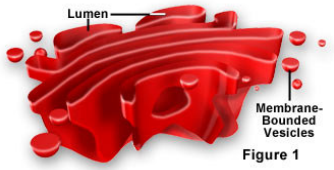


Nuclear membrane

Mitochondria

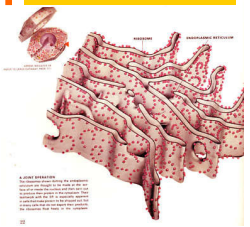


Golgi



Vesicles

E.R.



Plasma membrane (<10% of all membranes)

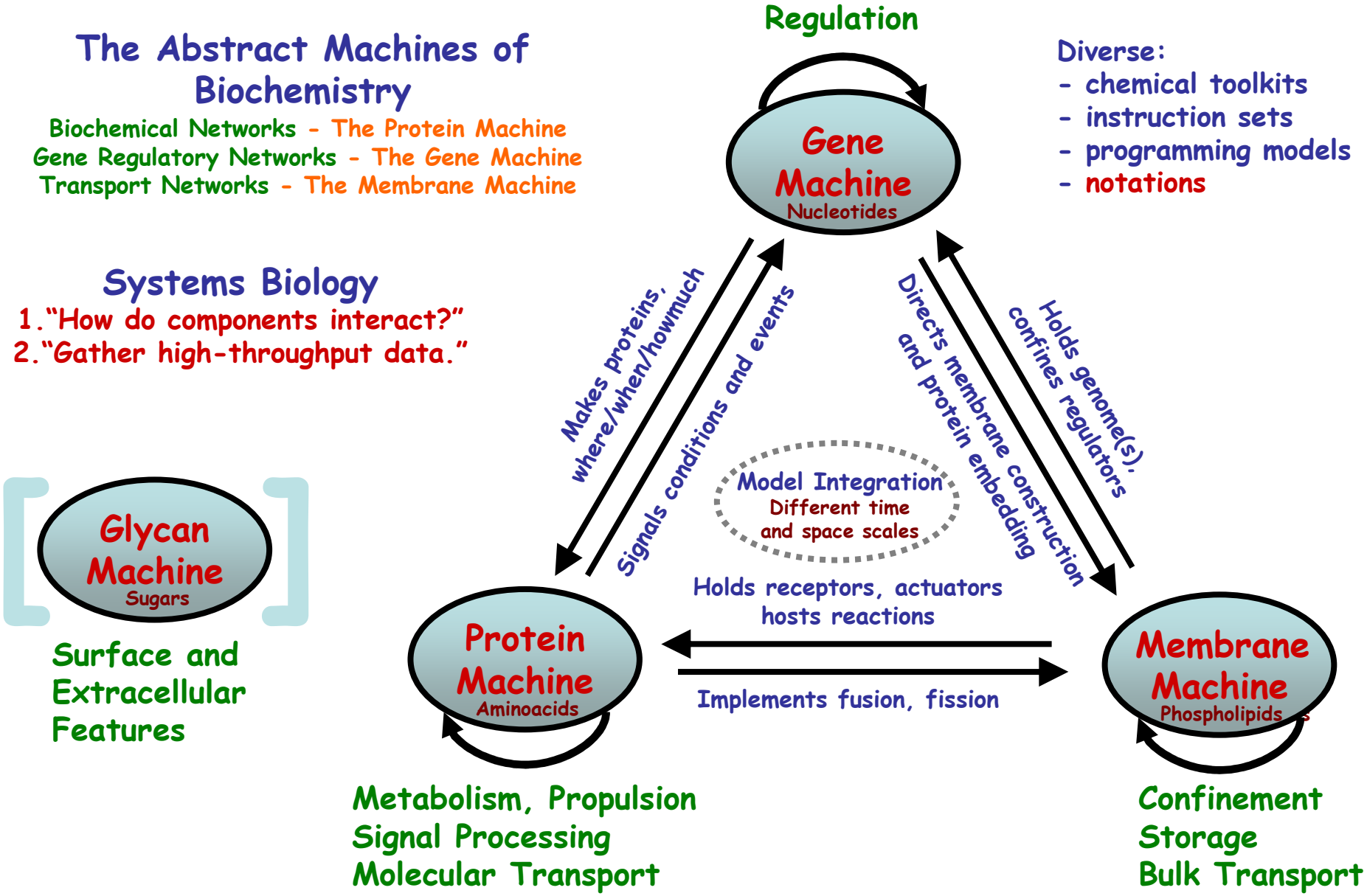
Functional Architecture

The Abstract Machines of Biochemistry

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

Systems Biology

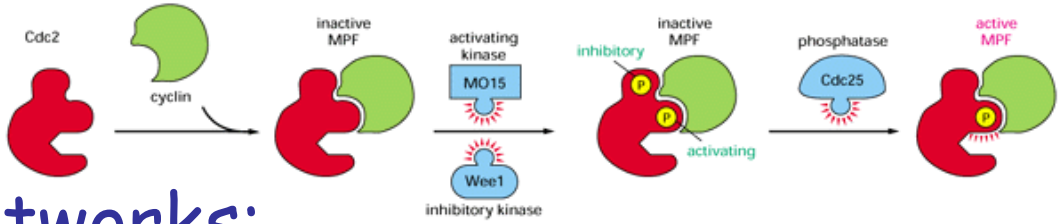
1."How do components interact?"
2."Gather high-throughput data."



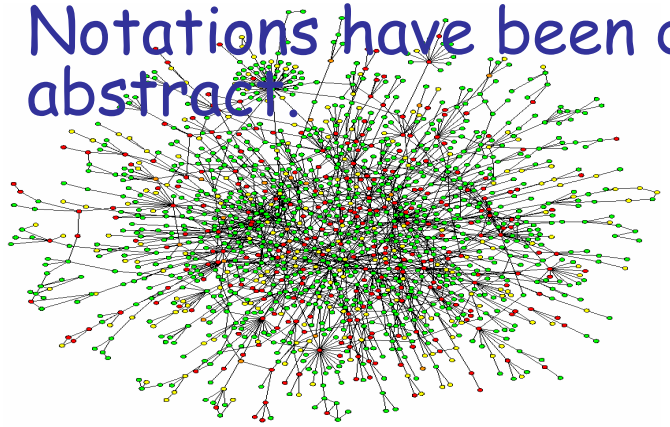
1. The Protein Machine

Very close to the atoms.

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



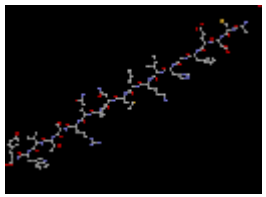
- Mapping out such networks:
  - In principle, it's "just" a very large set of chemical equations.
  - Notations have been developed to summarize and abstract.



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

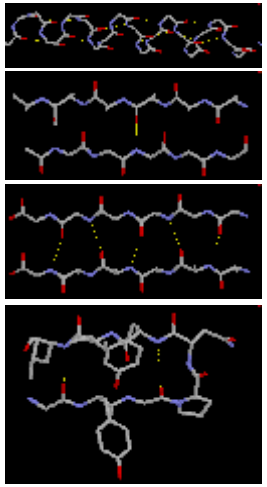
Protein Structure

Primary



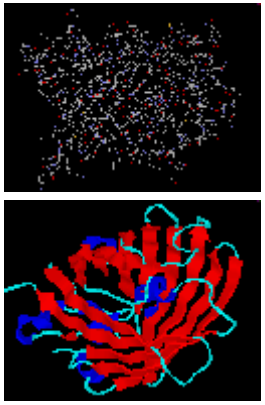
The 20 Aminoacids

Secondary



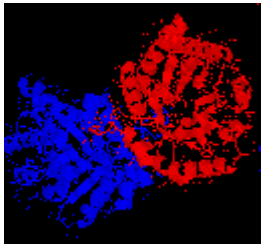
Alpha Helix, Beta Sheet

Tertiary



Green Fluorescent Protein

Quaternary



Triose Phosphate Isomerase

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

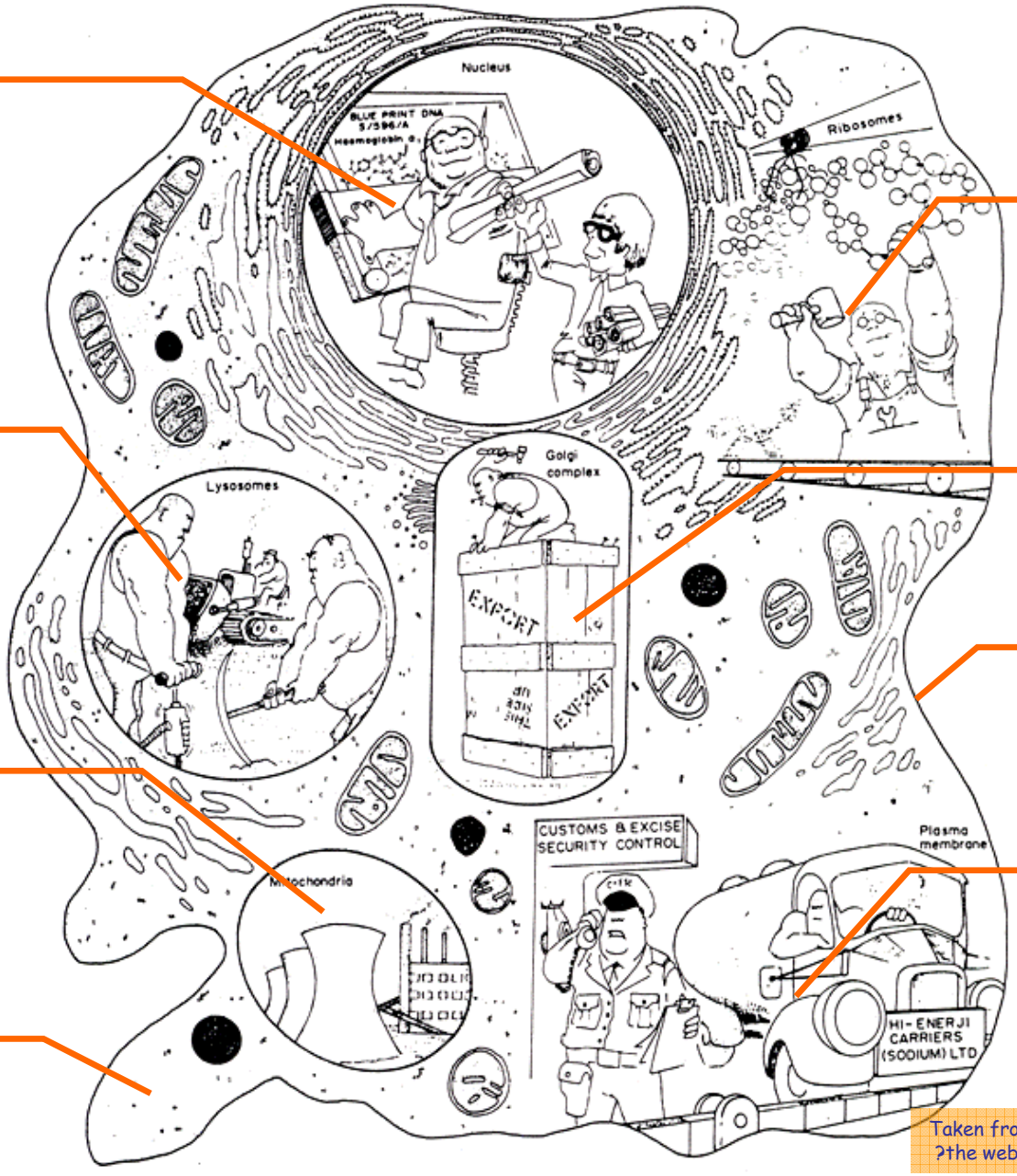
Protein Function

Regulation

Degradation

Metabolism

Movement



Assembly

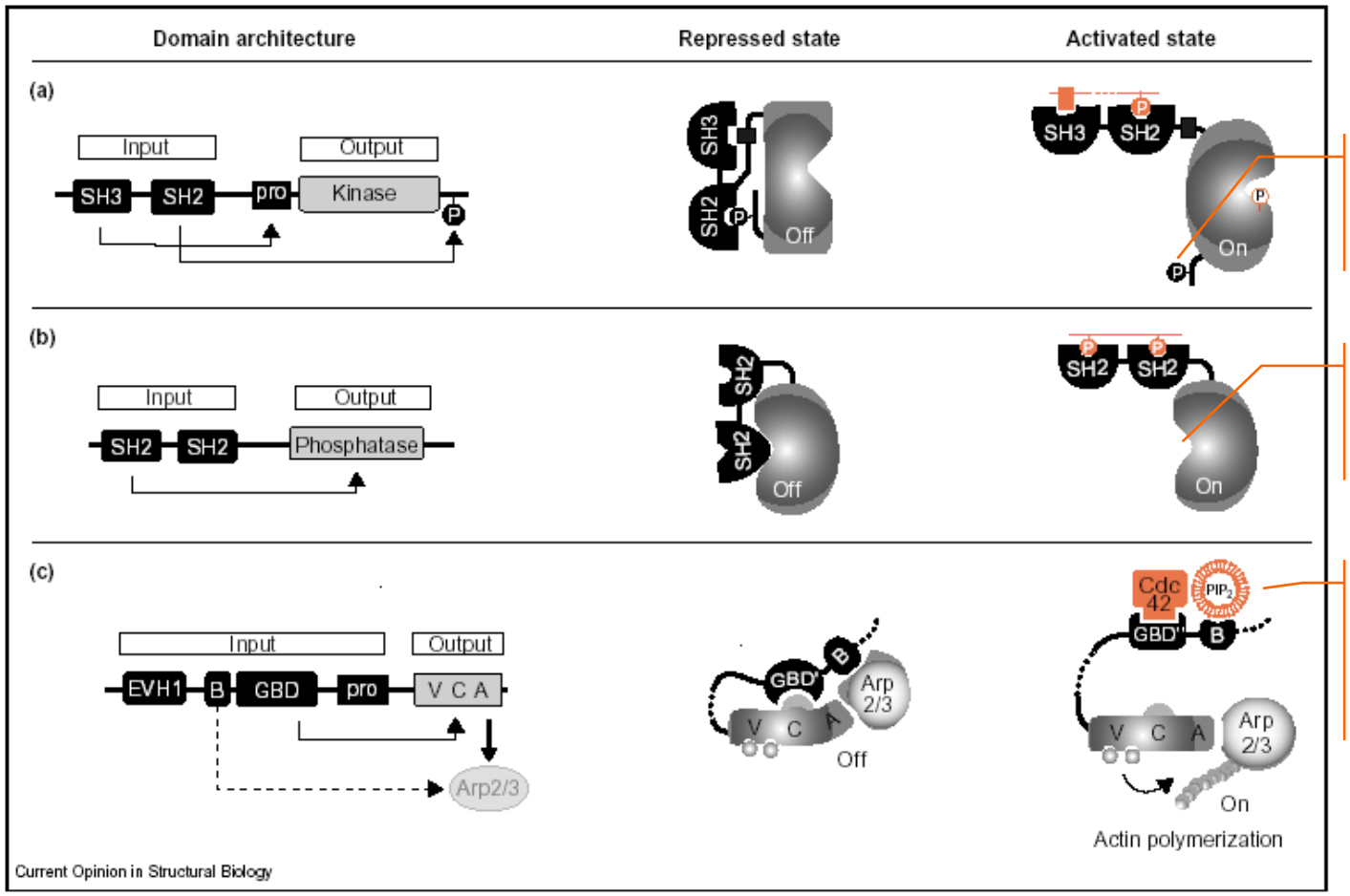
Transport

Structure

Signalling

Taken from ?the web?

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.

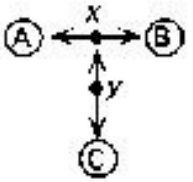
Humans have the same number of modular protein *domains* (building blocks) as worms, but twice the number of multi-domain proteins.

Taken from Wendell Lim

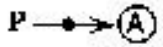
MIM: Molecular Interaction Maps (Kohn)

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.

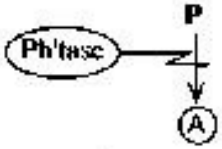
Asymmetric binding where protein **A** donates a peptide that binds to a receptor site or pocket on protein **B**.



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.



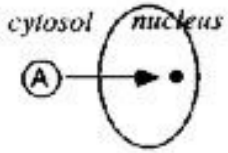
Covalent modification of protein **A**. The single-arrowed line indicates that **A** can exist in a phosphorylated state. The node represents the phosphorylated species.



Cleavage of a covalent bond: dephosphorylation of **A** by a phosphatase.

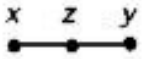
Proteolytic cleavage at a specific site within a protein.

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

Taken from Kurt W. Kohn

Molecular Interaction Maps

http://www.cds.caltech.edu/~hsauro/index.htm

JDesigner

The p53-Mdm2 and DNA Repair Regulatory Network

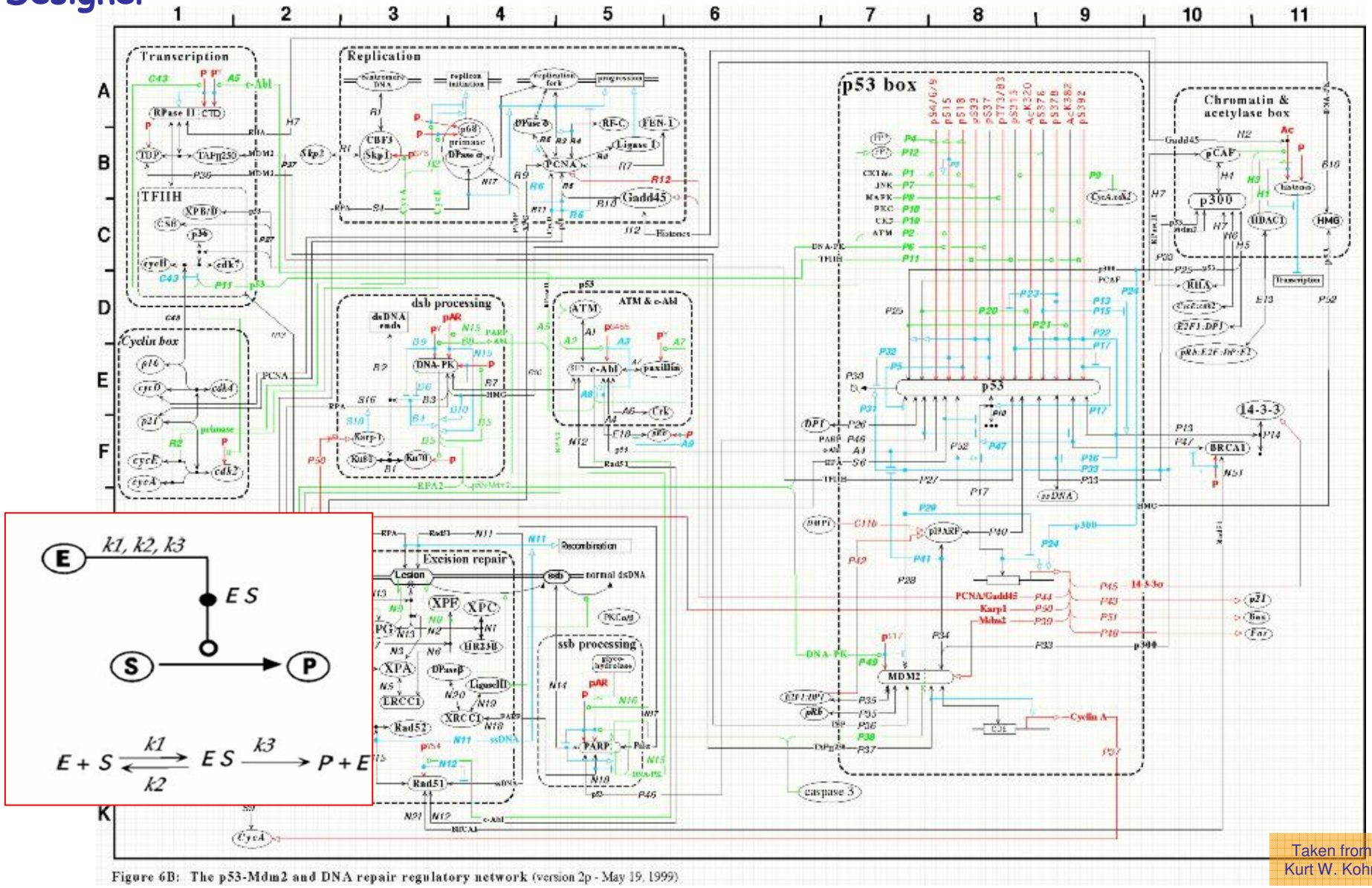


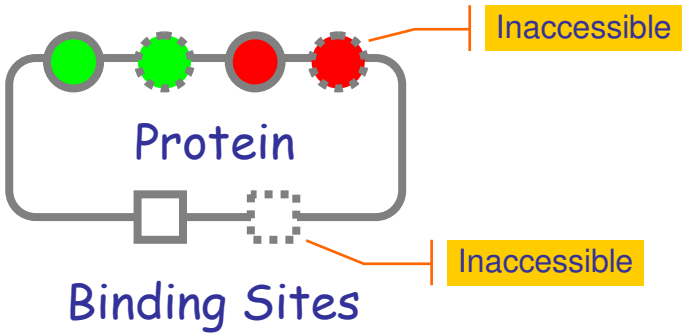
Figure 6B: The p53-Mdm2 and DNA repair regulatory network (version 2p - May 19, 1999)

Taken from Kurt W. Kohn

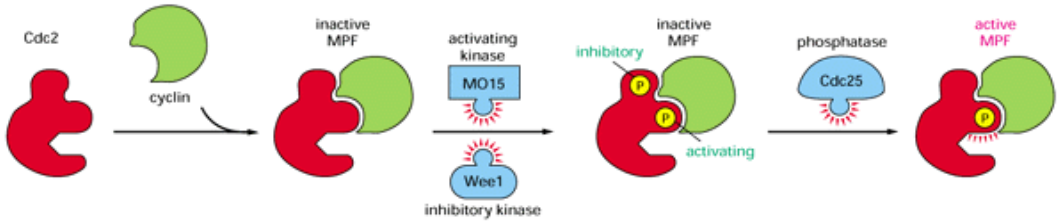
The Protein Machine "Instruction Set"

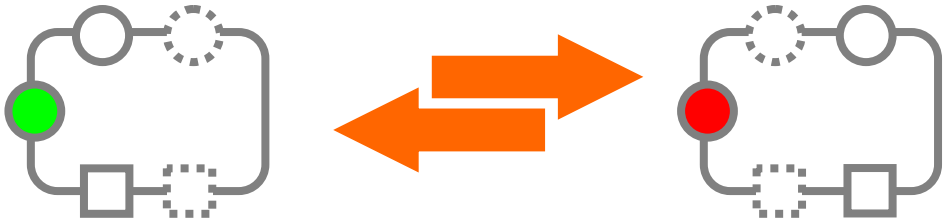
cf. BioCalculus [Kitano&Nagasaki], κ-calculus [Danos&Laneve]

On/Off switches



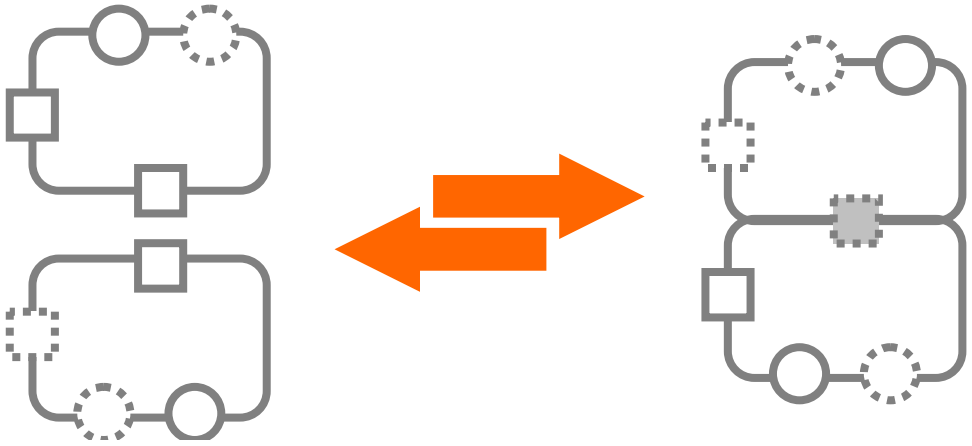
Each protein has a structure
of binary switches and binding sites.
But not all may be always *accessible*.





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.

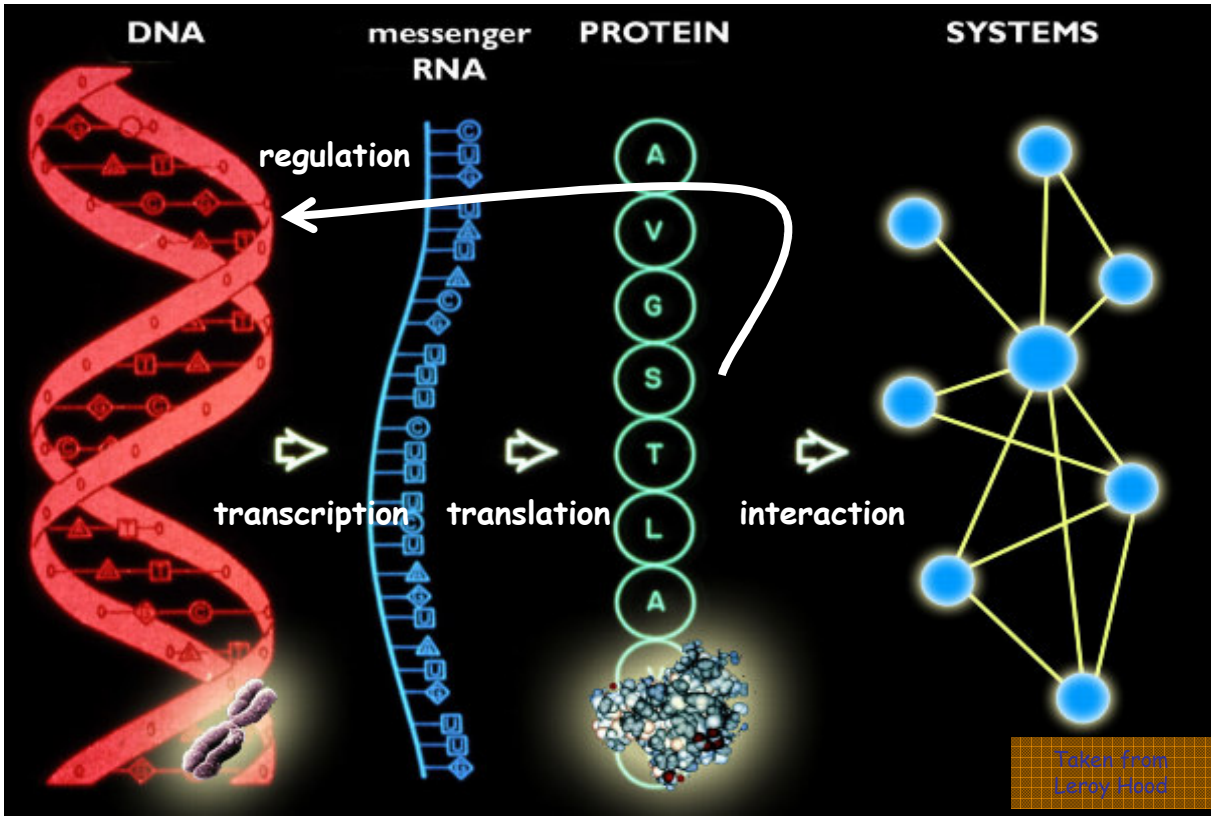
Notations for the Protein Machine

- Stochastic π-Calculus
  - Priami (following Hillston's PEPA) formalizes a stochastic version of π-calculus where channels have communication *rates*.
- BioSPi
  - Regev-Shapiro-Silverman propose modeling chemical interactions (exchange of electrons and small molecules) as "communication".
  - Standard stochastic simulation algorithms (Gillespie) can be used to run in-silico experiments.
  - Complex formation is encoded via π-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
    - Graph editors
    - Simulators (including simulation web services)
    - Databases

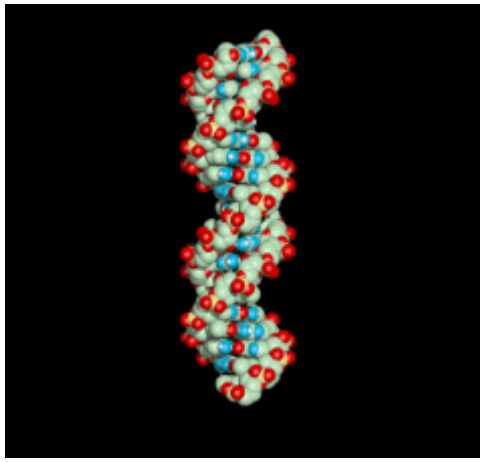
2. The Gene Machine

Pretty far from the atoms.

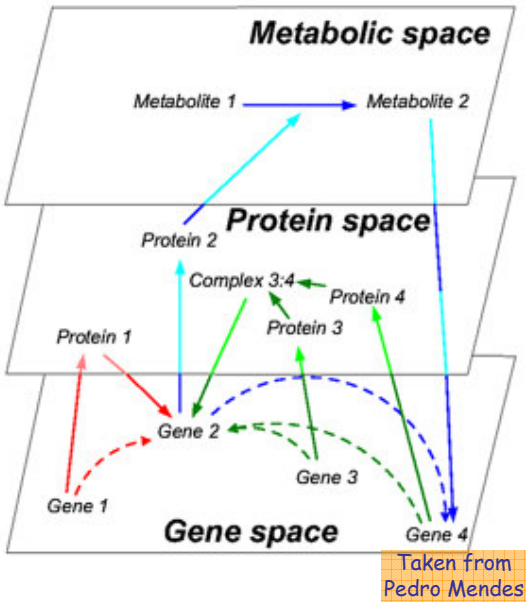
The "Central Dogma" of Molecular Biology

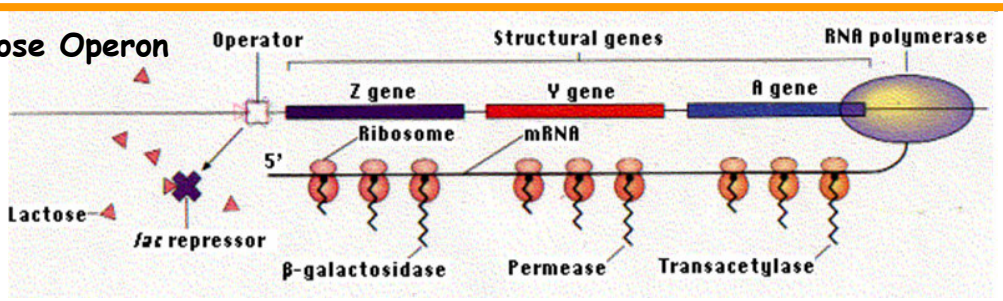


4-letter digital code | 4-letter digital code | 20-letter digital code | 50.000(?) shapes



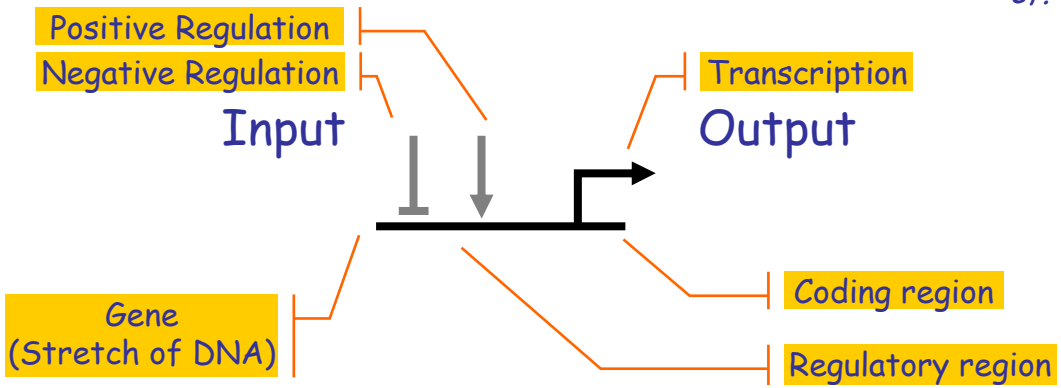
DNA Tutorial

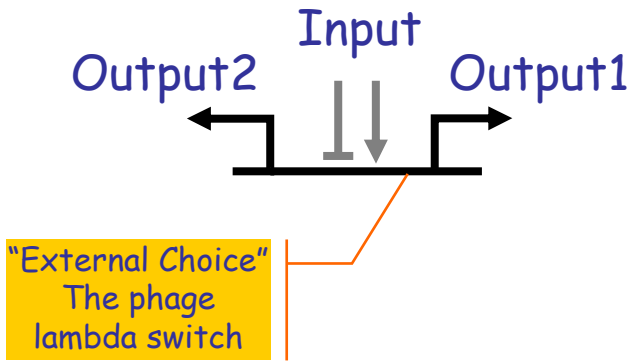




The Gene Machine "Instruction Set"

cf. Hybrid Petri Nets [Matsuno, Doi, Nagasaki, Miyano]





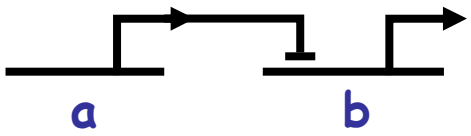
Regulation of a gene (positive and negative) influences transcription. The regulatory region has precise DNA sequences, but not meant for coding proteins: meant for binding regulators.

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

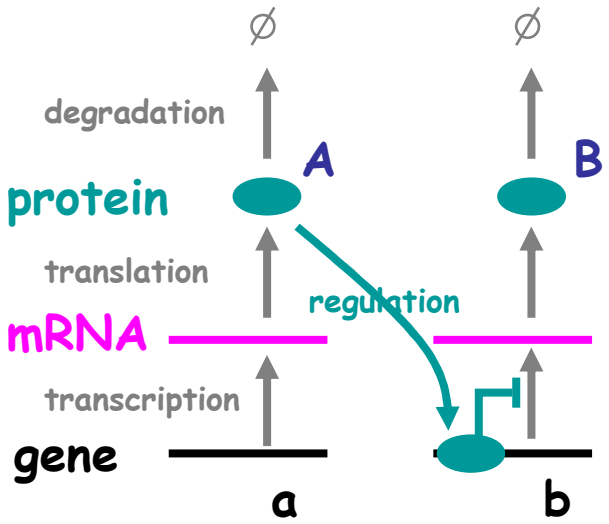
Human (and mammalian) Genome Size
3Gbp (Giga base pairs) 750MB @ 4bp/Byte (CD)
Non-repetitive: 1Gbp 250MB
In genes: 320Mbp 80MB
Coding: 160Mbp 40MB
Protein-coding genes: 30,000-40,000

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



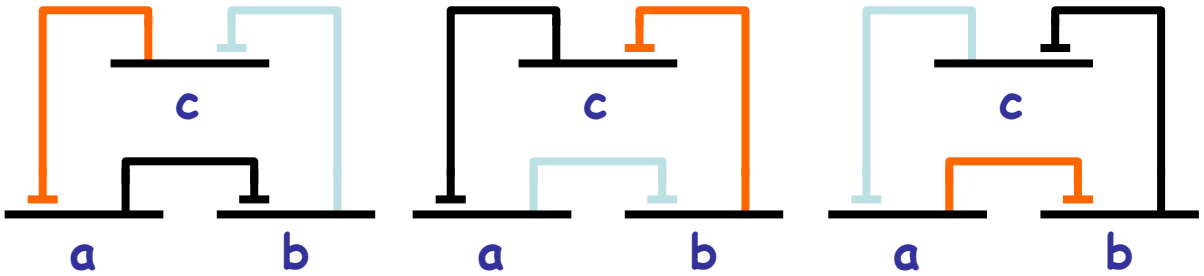
Is a shorthand for:



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

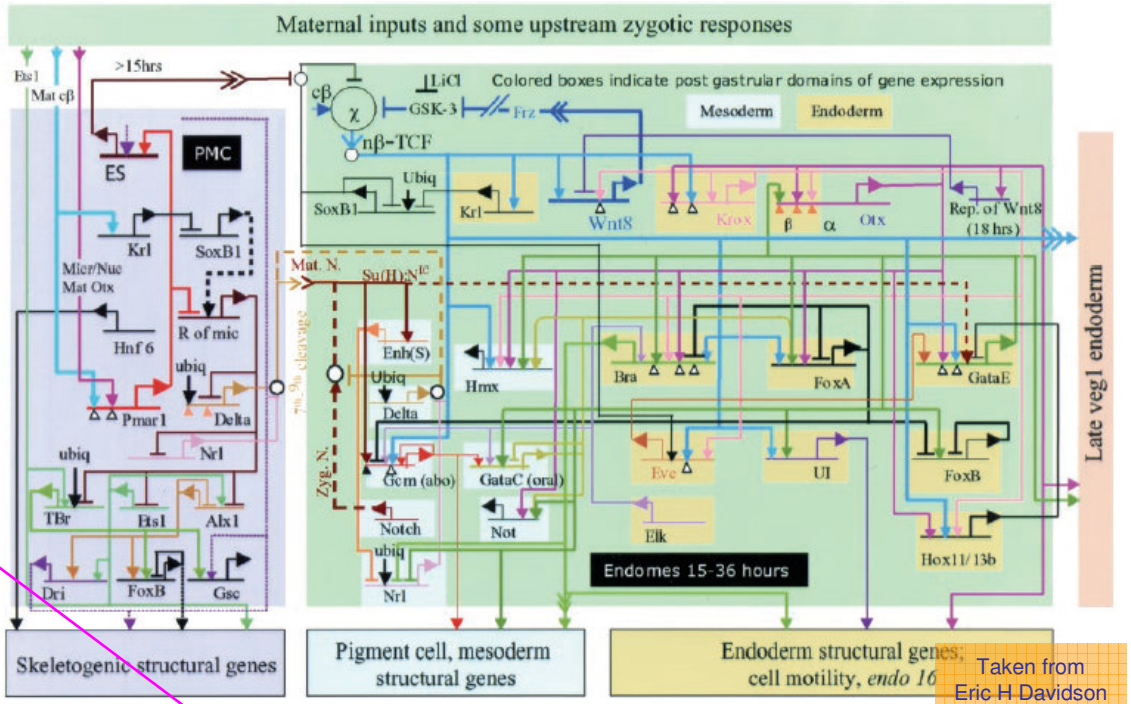


Expressed
Repressed
Expressing

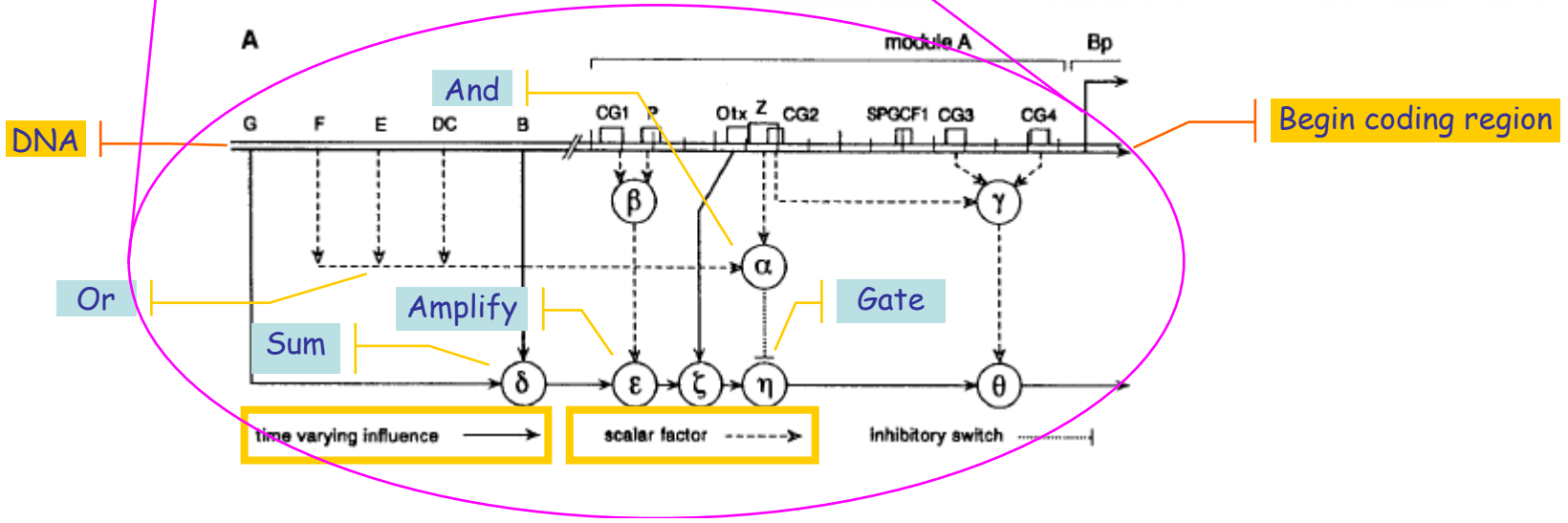
Gene Regulatory Networks

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

NetBuilder

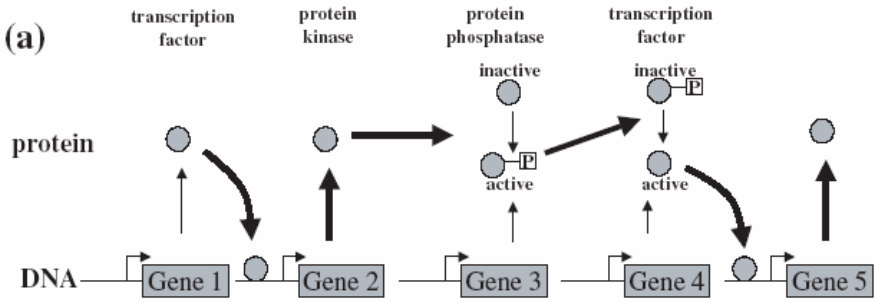


Taken from Eric H Davidson



Indirect Gene Effects

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



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

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

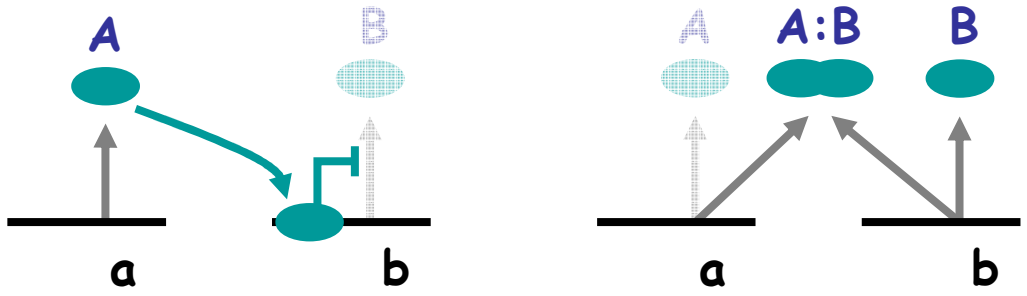
(c)
Aspect of gene activity: phosphorylation state
Genetic perturbation: gene deletion

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

Taken from Andreas Wagner

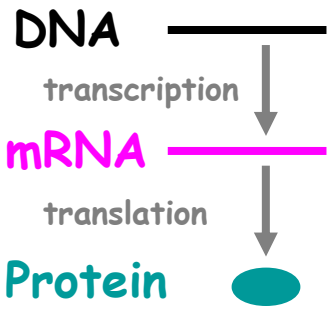
Fig. 1. The importance of specifiying gene activity when reconstructing genetic networks. (a) A hypothetical biochemical pathway involving two transcription factors, a protein kinase, and a protein phosphatase, as well as the genes encoding them. See text for details. (b) Shown is a list of perturbation 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].

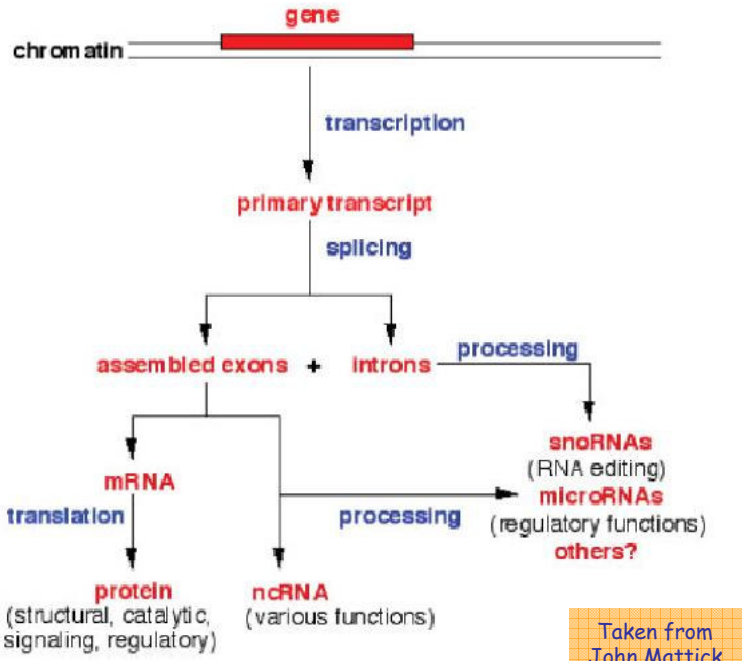


Structure of the Coding Region

The Central Dogma



Challenging the Dogma (in higher organisms)

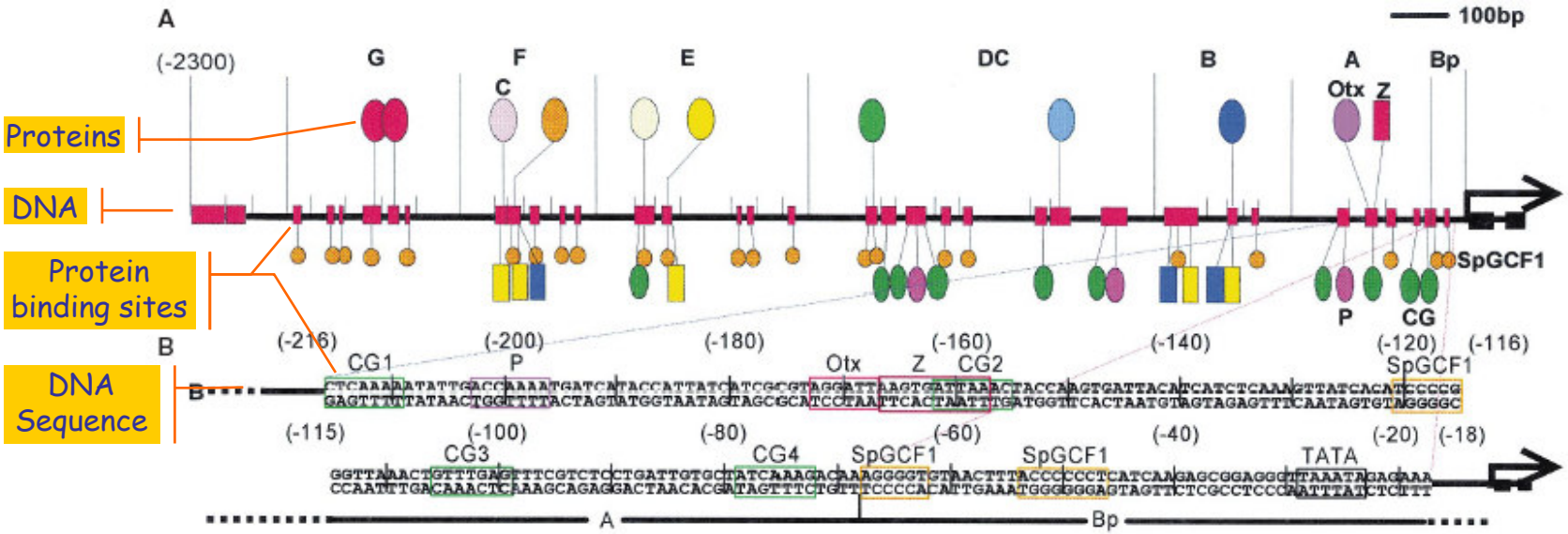


Taken from John Mattick

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

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



2300bp!

> average protein

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:

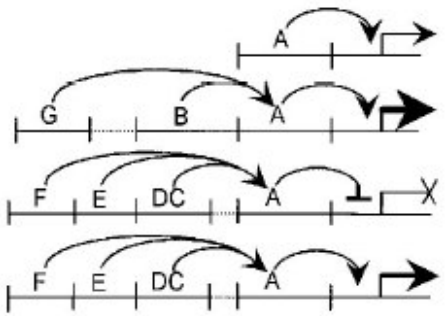
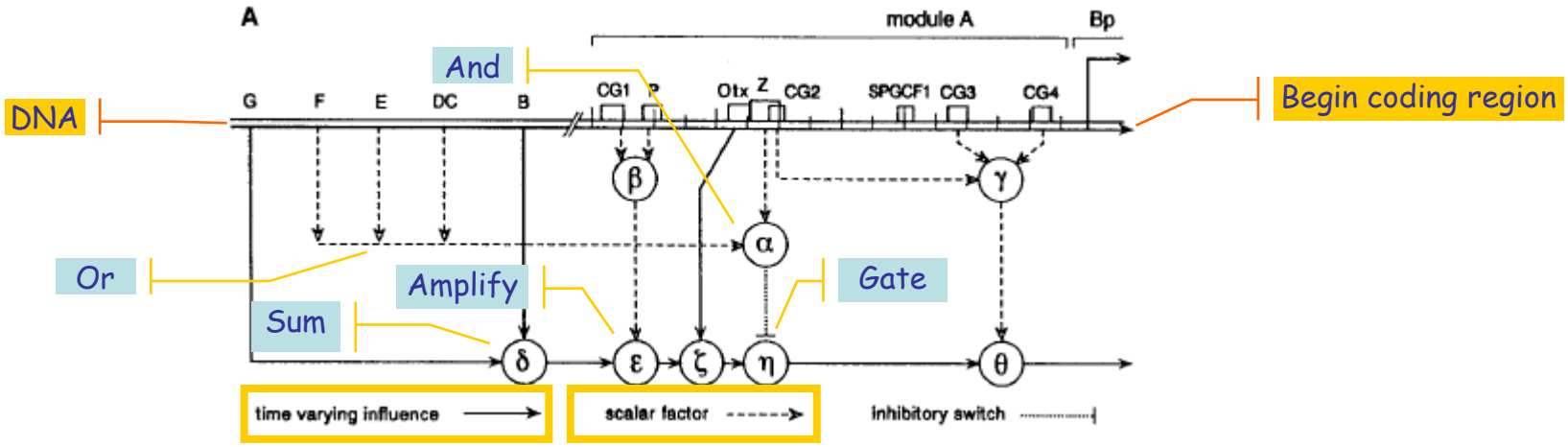


Fig. 1. *Endo16* cis-regulatory system and interactive roles of module A. (**A**) Diversity of protein binding sites and organization into modular subregions [modified from (*7*)]. Specific DNA binding sites are indicated as red blocks; modular subregions are denoted by letters G to A (Bp, basal promoter). 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.

Taken from Eric H Davidson

Function of a Regulatory Region



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
$\alpha = 1$	
else $\alpha = 0$	
if (P = 1 and CG_1 = 1)	Both P and CG_1 needed for synergistic link with module B
$\beta = 2$	
else $\beta = 0$	
if (CG_2 = 1 and CG_3 = 1 and CG_4 = 1)	Final step up of system output
$\gamma = 2$	
else $\gamma = 1$	
$\delta(t) = B(t) + G(t)$	Positive input from modules B and G
$\varepsilon(t) = \beta^{*}\delta(t)$	Synergistic amplification of module B output by CG_1-P subsystem
if ($\varepsilon(t) = 0$)	Switch determining whether Otx site in module A, or upstream modules (i.e., mainly module B), will control level of activity
$\xi(t) = Otx(t)$	
else $\xi(t) = \varepsilon(t)$	
if ($\alpha = 1$)	Repression function inoperative in endoderm but blocks activity elsewhere
$\eta(t) = 0$	
else $\eta(t) = \xi(t)$	
$\Theta(t) = \gamma^{*}\eta(t)$	Final output communicated to BTA

Taken from
Eric H Davidson

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 process.

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.

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.

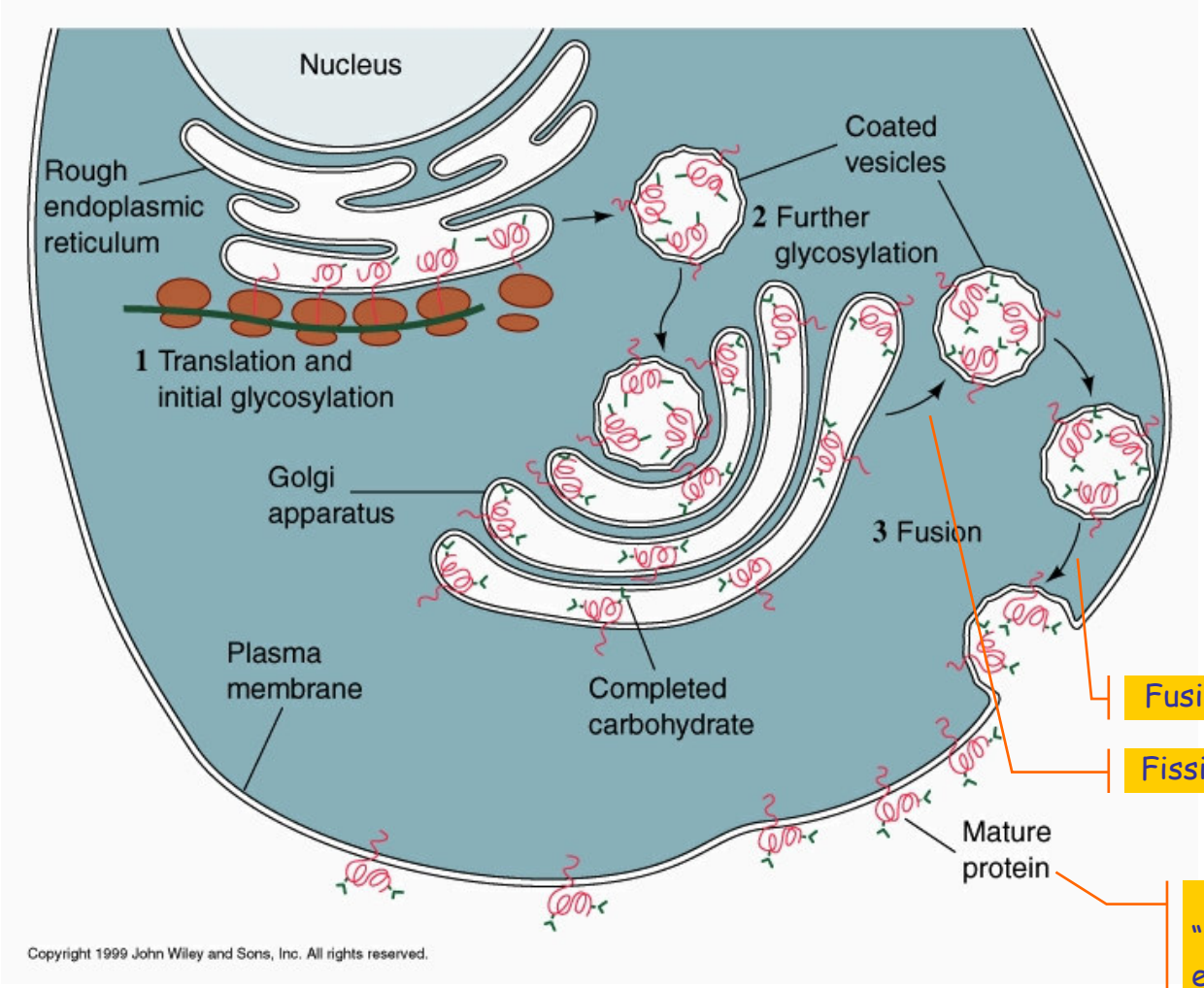
Stochastic degradation. Degradation is fundamental for system stability, and at the same time can lead to sudden instability and detection of concentration levels.

Notations for the Gene Machine

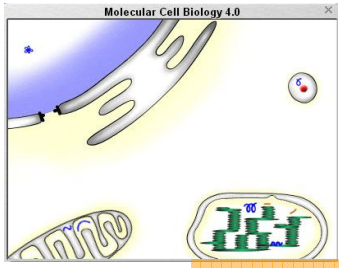
- Many of the same techniques as for the Protein Machine apply.
  - Process Calculi, Petri Nets, Term-Rewriting Systems…
- But the "programming model" is different.
  - Asynchronous stochastic control.
  - Biologically poorly understood.
  - Network "motifs" are being analyzed.
- Specific techniques:
  - Hybrid Petri Nets
    - [Matsuno, Doi, Nagasaki, Miyano] Gene Regulation
    - Genomic Object Net www.genomicobject.net
- Gene Regulation Diagrams
- Mixed Gene-Protein Diagrams

3. The Membrane Machine

Very far from the atoms.



Molecular transport and transformation through dynamic compartment fusion and fission.



Taken from MCB CD

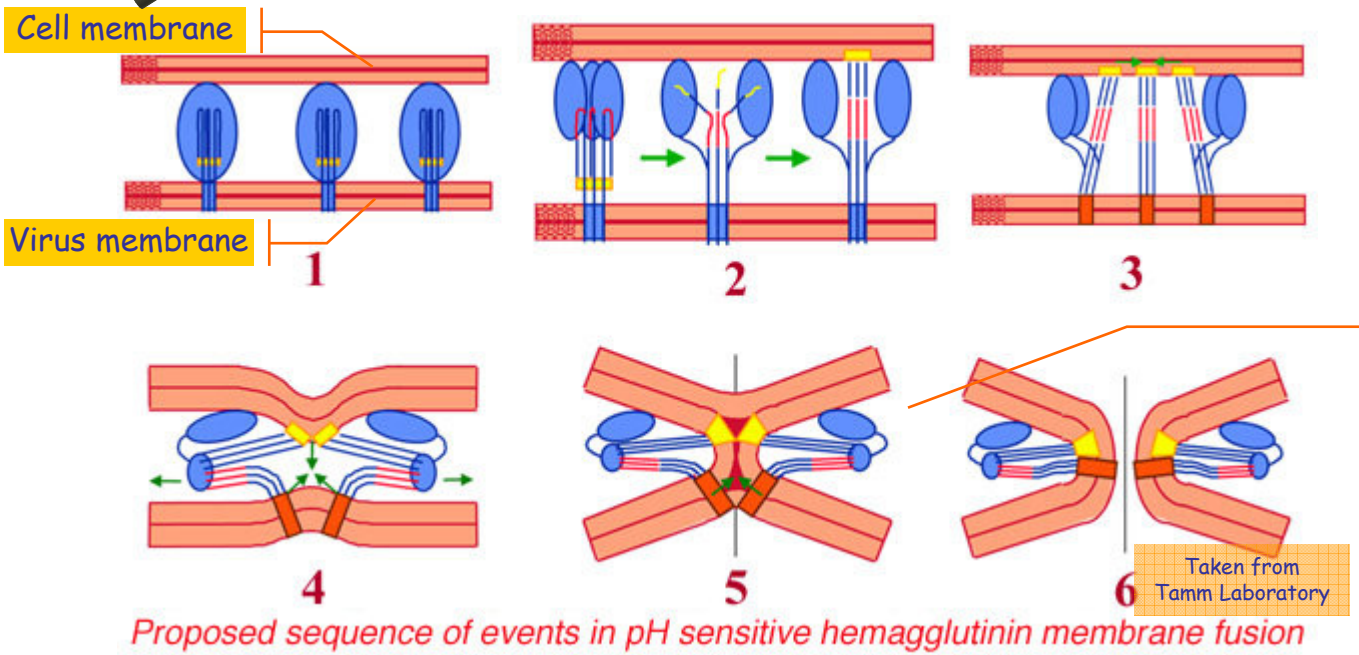
Fusion
Fission
} **The Instruction Set**

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" [MBP p.609]

Membrane Fusion

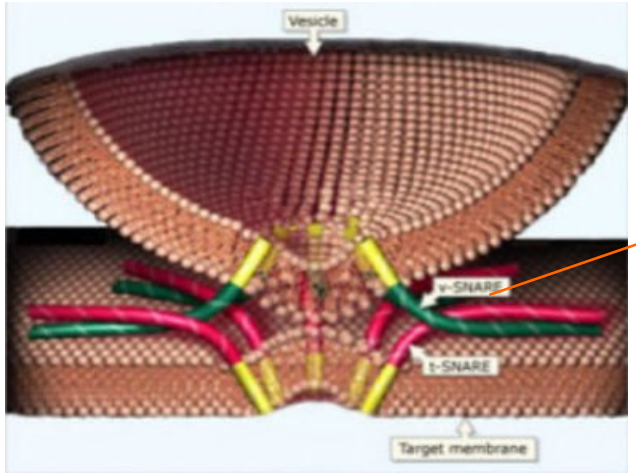
Positive curvature to Negative curvature transition in 3D



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]

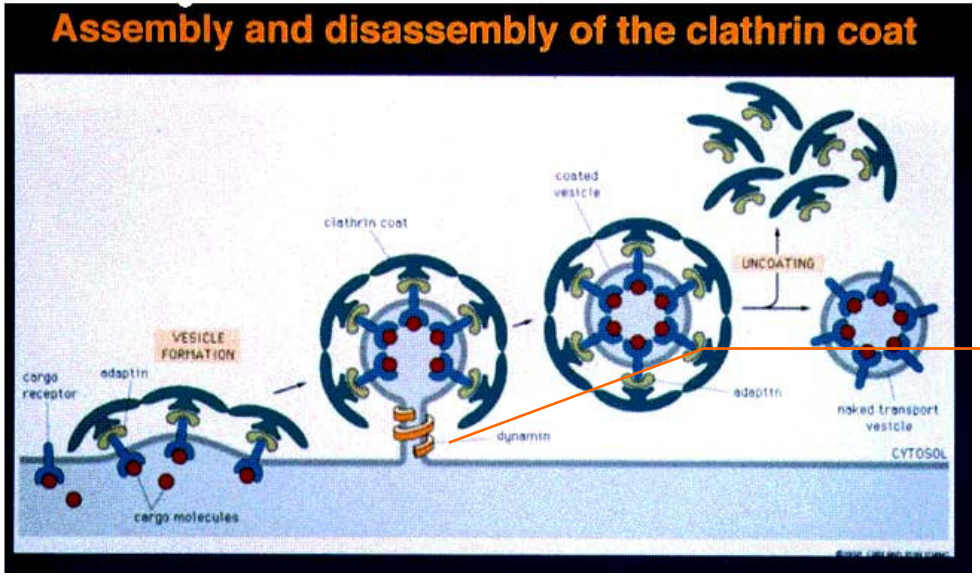


Cooperative fusion (vesicle)

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

Membrane Fission

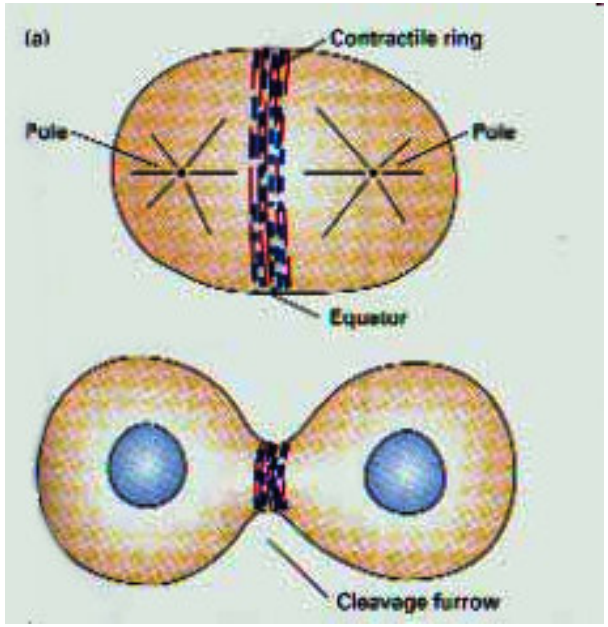
Negative curvature to Positive curvature transition in 3D



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]

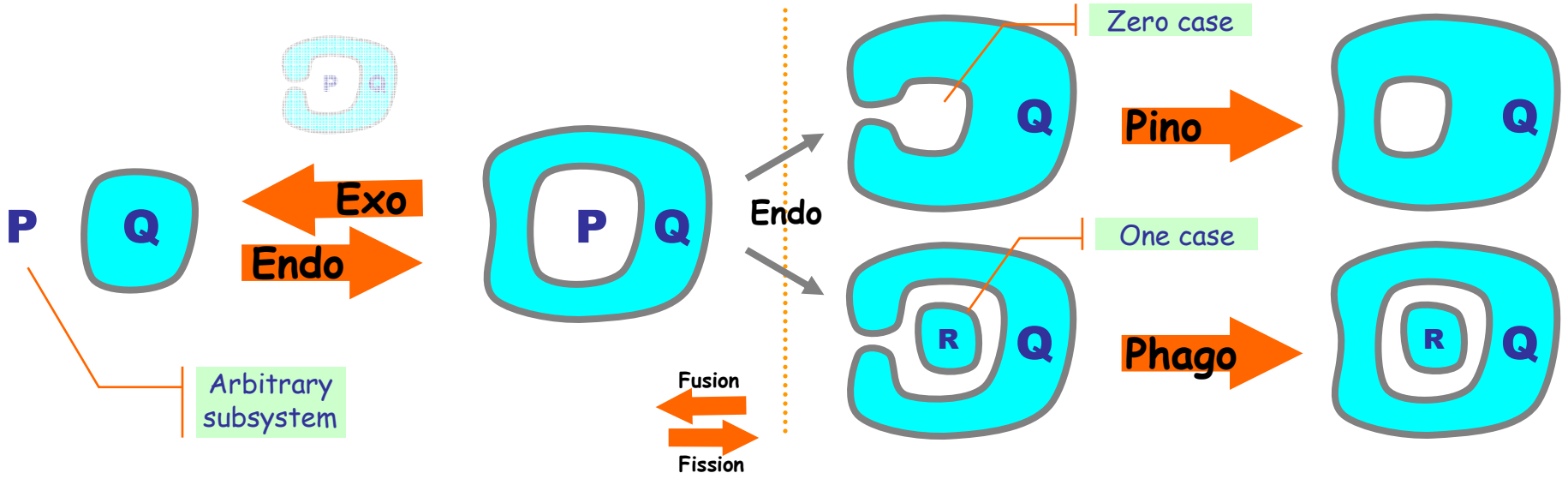


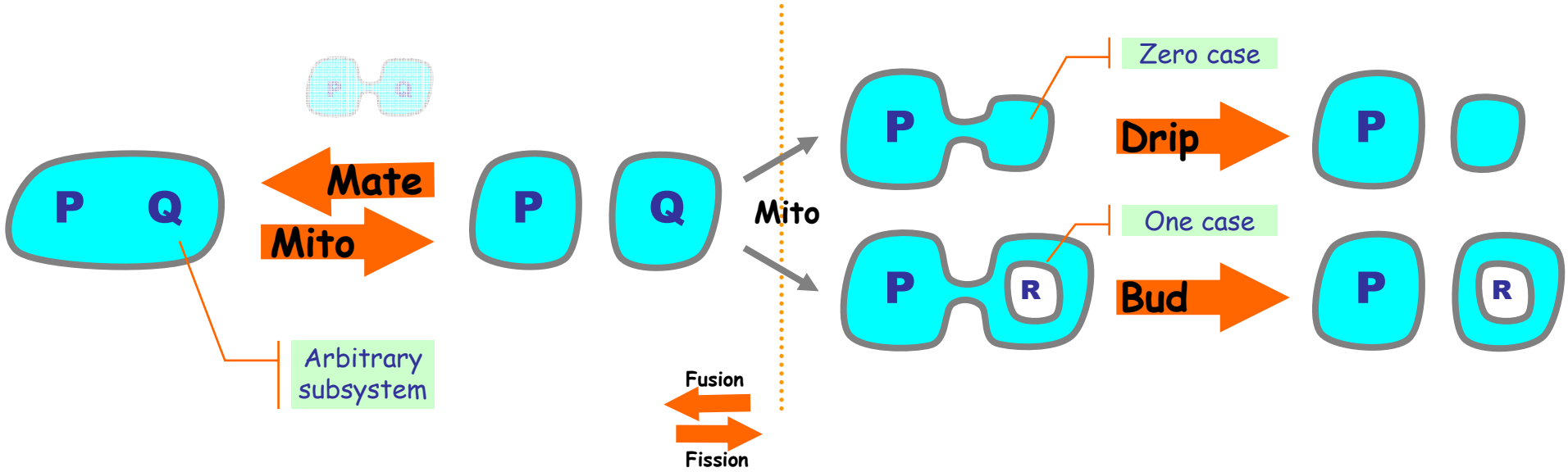
Cytokinesis (Mitosis)

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…

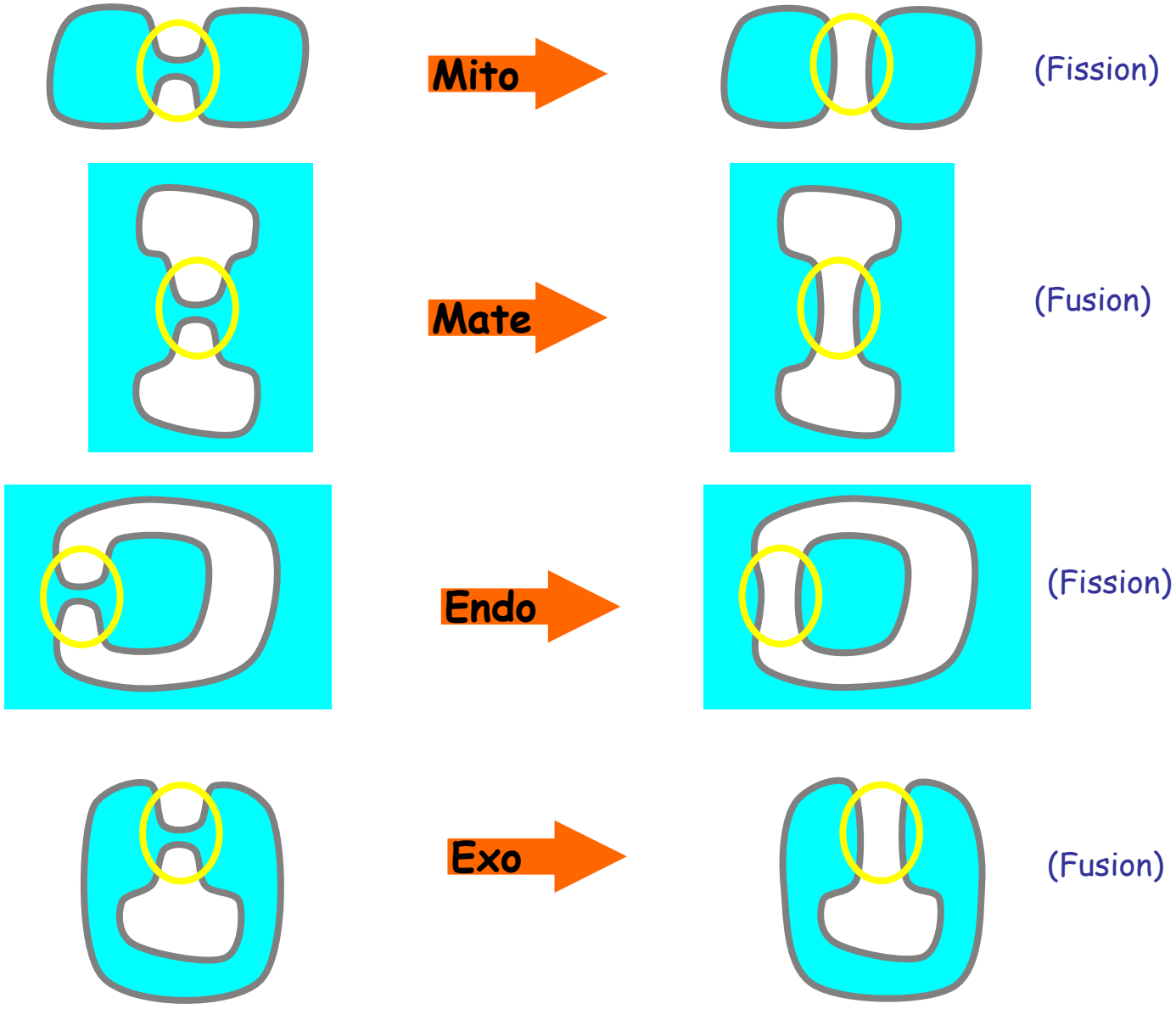
The Membrane Machine "Instruction Set"



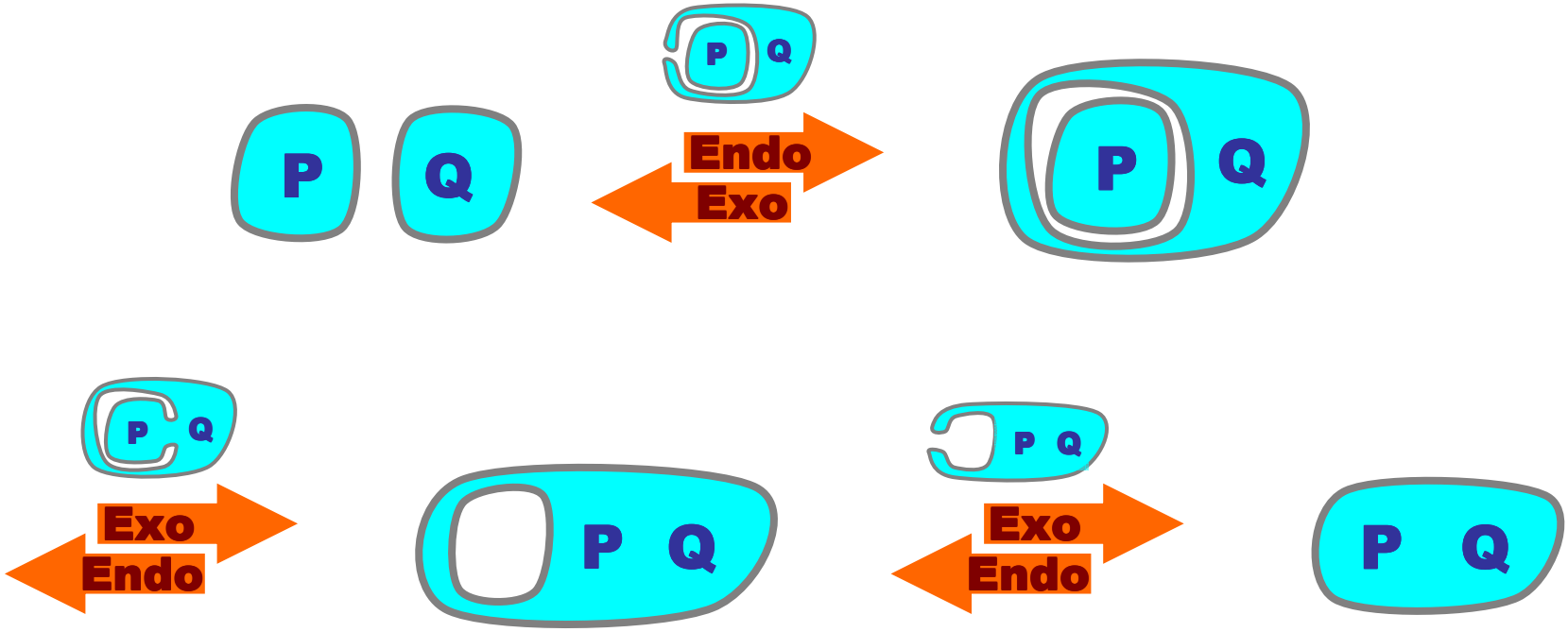


Locally Implementable!

Global Views

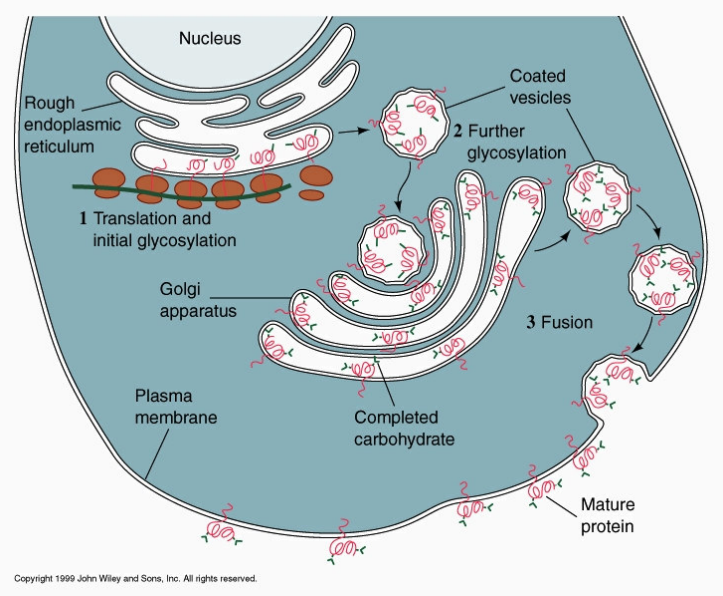


Mito/Mate by 3 Endo/Exo

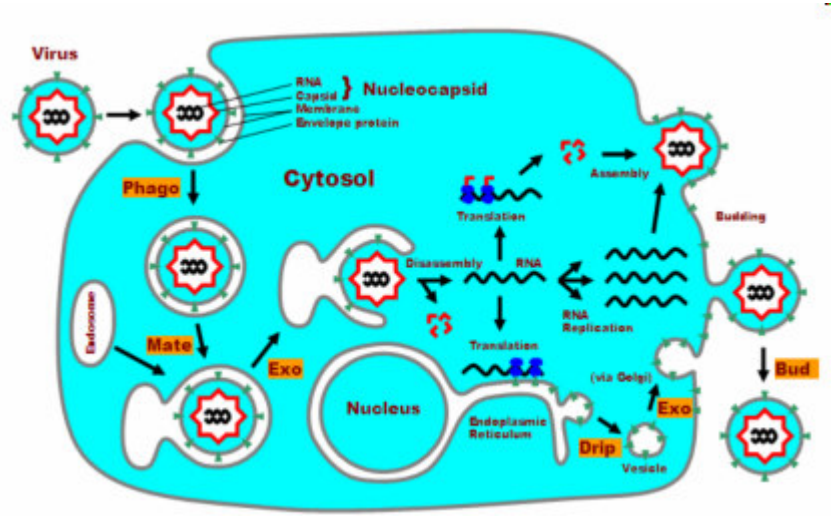


Membrane Algorithms

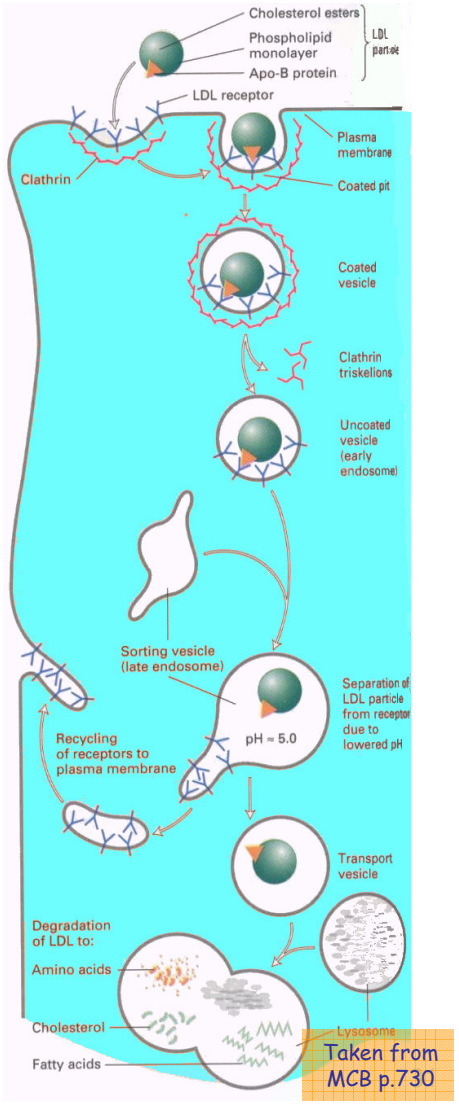
Protein Production and Secretion



Viral Replication

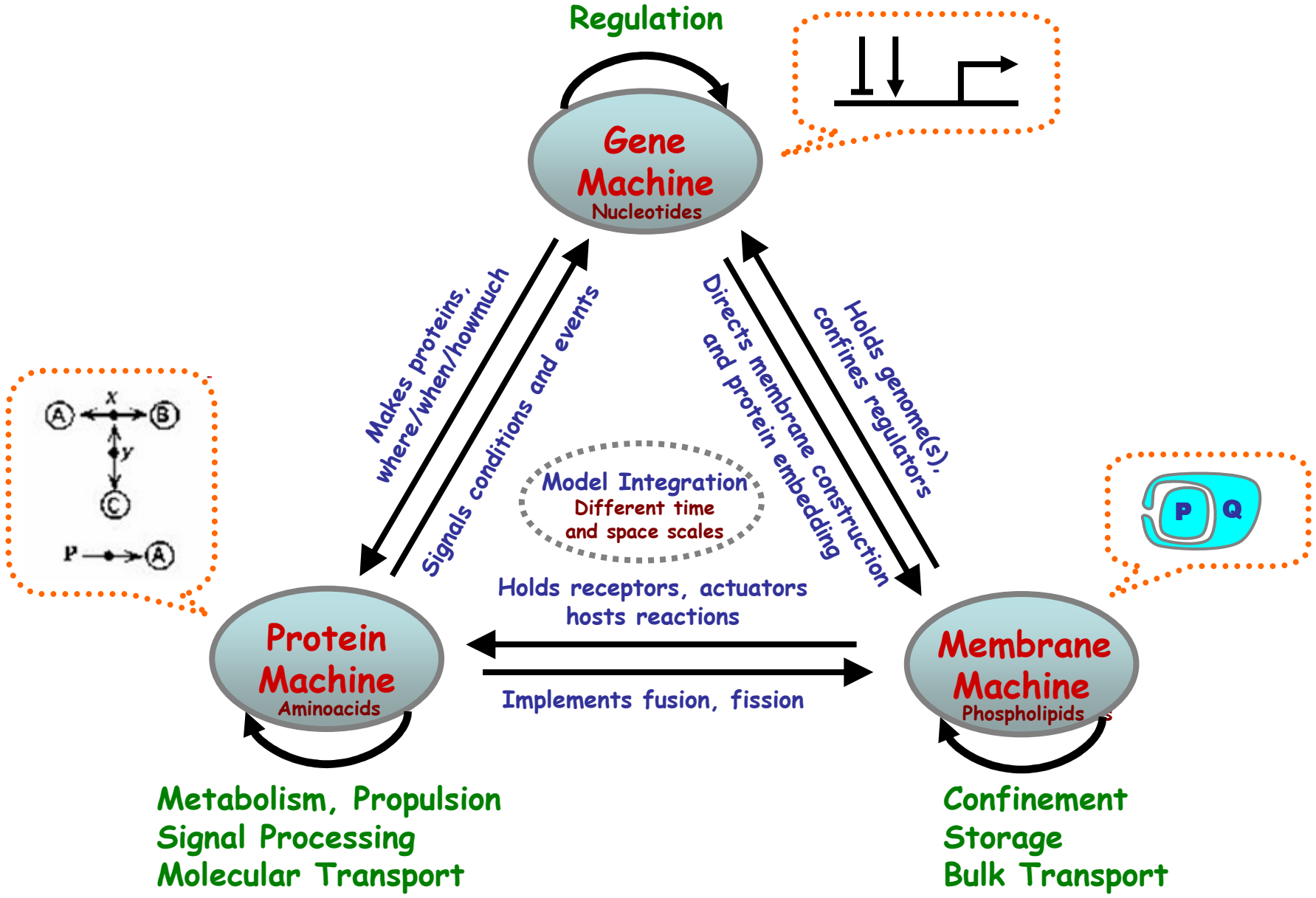


LDL-Cholesterol Degradation



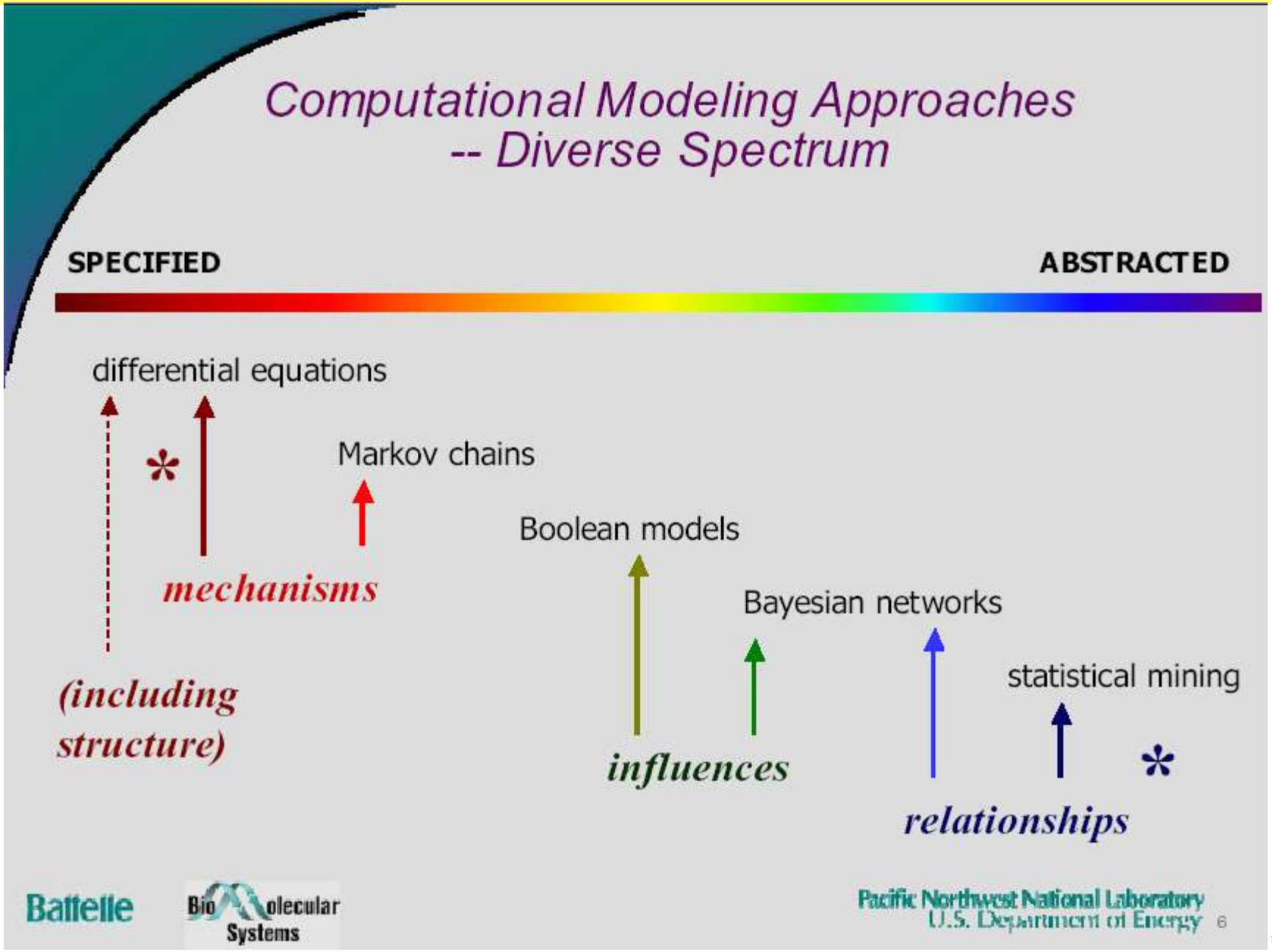
Taken from MCB p.730

Abstract Machines of Molecular Biology

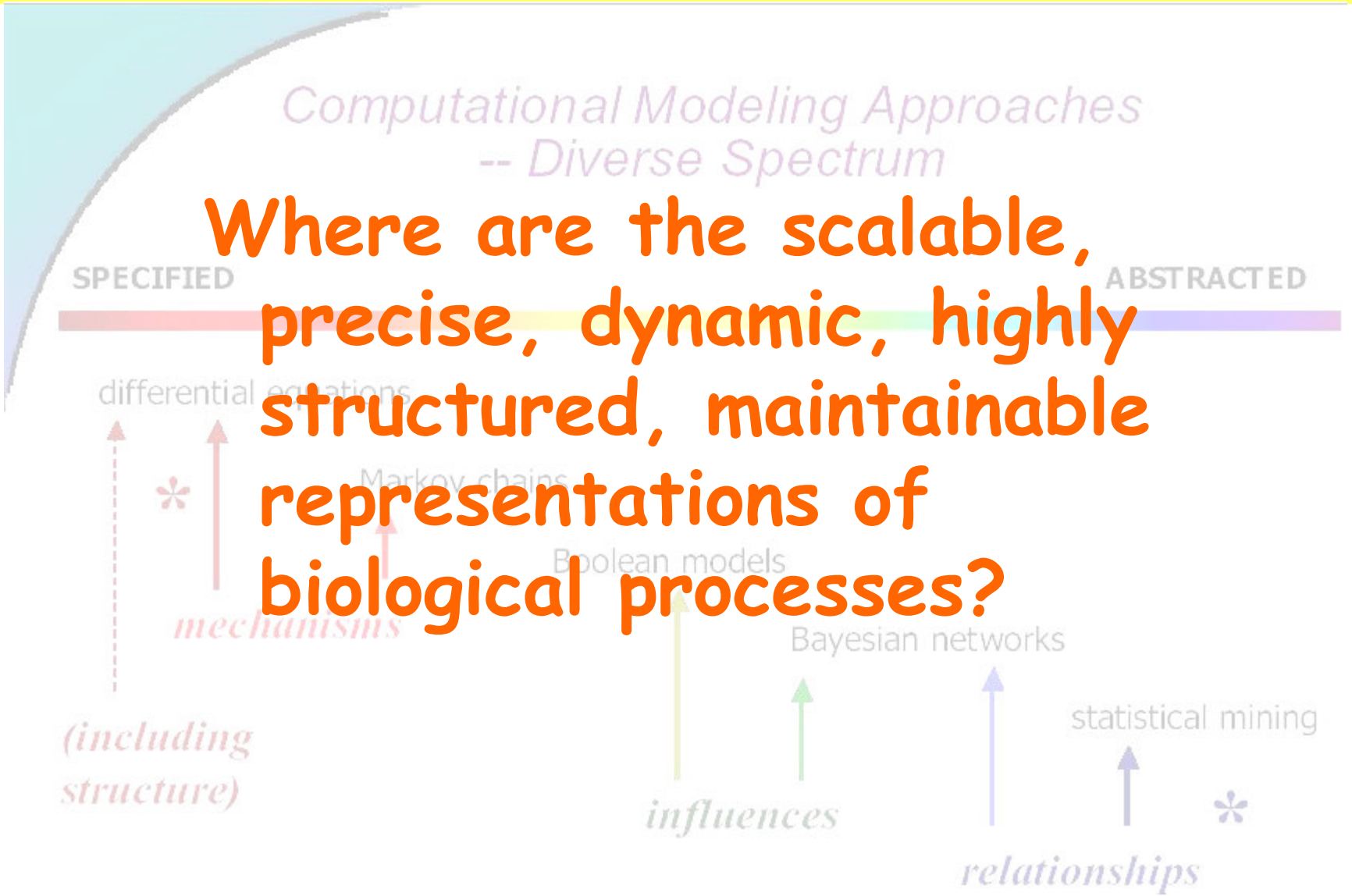


Stochastic Process Calculi

A Frequently-Seen Slide



A Frequently-Seen Slide

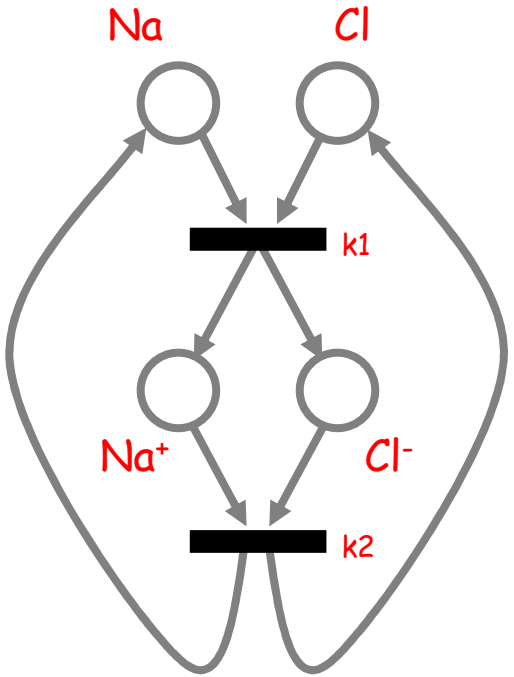


Where are the scalable, precise, dynamic, highly structured, maintainable representations of biological processes?

Chemistry vs. π-calculus

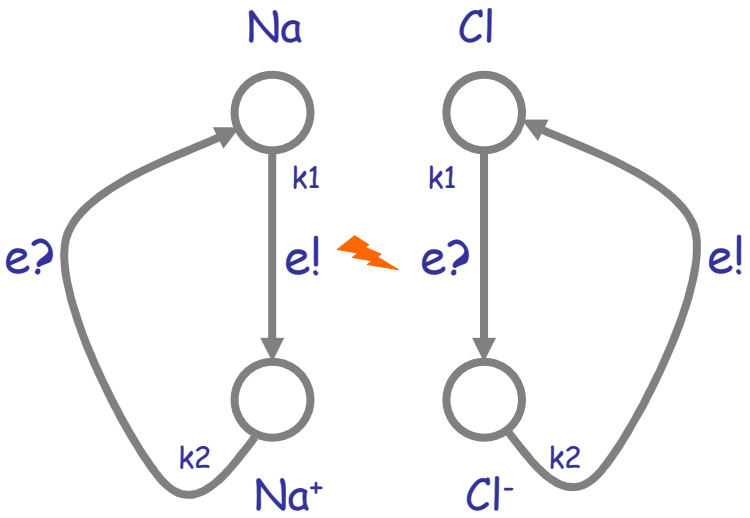
A process calculus (chemistry, or SBML)

$$Na + Cl \rightarrow_{k1} Na^+ + Cl^-$$
$$Na^+ + Cl^- \rightarrow_{k2} Na + Cl$$



(Can be converted to a CTMC)

A compositional graphical representation, and the corresponding calculus.



(Can be converted to a CTMC)

The same "model"

$$Na = e_{k1}!.\ \overbrace{e_{k2}?.\ Na}^{Na^+}$$
$$Cl = e_{k1}?.\ \underbrace{e_{k2}!.\ Cl}_{Cl^-}$$

A different process calculus (π)

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

Stochastic π-calculus Executive Summary

- A process calculus:
  - The modular representation of concurrent (and stochastic) processes of all kinds.
  - Cuts down to CTMCs in the finite case (not always), then standard tools are applicable.
  - Can be given friendly automata-like scalable graphical syntax (work in progress).
  - Is directly executable (e.g. via Gillespie).
  - Is analyzable (large body of literature, at least in the non-stochastic case).

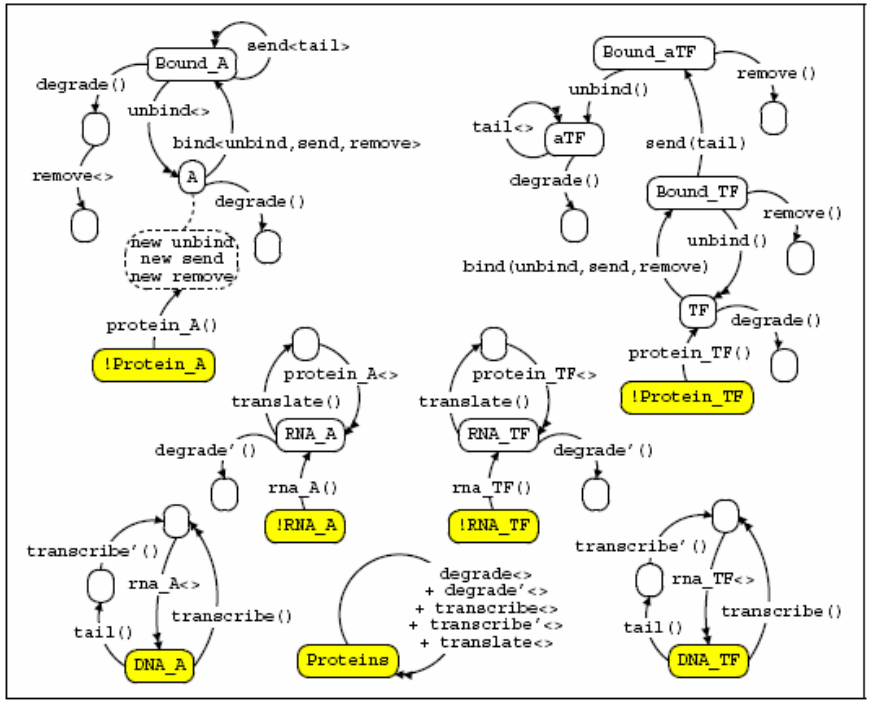


Figure 2. Regulating Gene Expression by Positive Feedback [9]

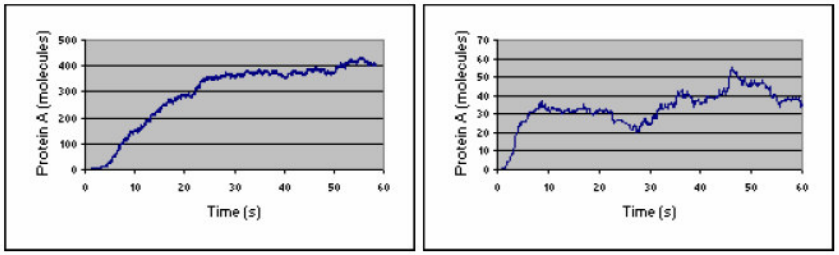


Figure 3. Protein A molecules v.s. time in presence (left) and absence (right) of TF

Regev-Shapiro: "Molecules as Computation"

Molecule	Process
Interaction capability	Channel
Interaction	Communication
Modification	State change

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

This mapping works well both for the "protein machine" (synchronous communication) and the "gene machine" (asynchronous communication). But is not enough for the "membrane machine".

π-calculus

Syntax

$$\begin{array}{rcll} \pi & ::= & x(y) & \text{receive } y \text{ along } x \\ & & \overline{x}\langle y\rangle & \text{send } y \text{ along } x \end{array}$$

$$P ::= 0 \mid \textstyle\sum_{i\in I} \pi_i.P_i \mid [x=y]\ P \mid P_1|P_2 \mid (\text{new } x)P \mid !P$$

Structural congruence

Renaming of bound variables

$$\begin{array}{rcll} x(y).P & = & x(z).(\{z/y\}\,P) & \text{if } z \notin FN(P) \\ (\text{new } y).P & = & (\text{new } z).(\{z/y\}\,P) & \text{if } z \notin FN(P) \end{array}$$

Structural congruence laws

$$\begin{array}{rcll} P|Q & \equiv & Q|P & \text{commutativity of parallel composition} \\ (P|Q)|R & \equiv & P|(Q|R) & \text{associativity of parallel composition} \\ P+Q & \equiv & Q+P & \text{commutativity of summation} \\ (P+Q)+R & \equiv & P+(Q+R) & \text{associativity of summation} \\ (\text{new } x)0 & \equiv & 0 & \text{restriction of inert processes} \\ (\text{new } x)(\text{new } y)P & \equiv & (\text{new } y)(\text{new } x)P & \text{polyadic restriction} \\ ((\text{new } x)P)|Q) & \equiv & (\text{new } x)(P|Q) \quad \text{if } x \notin FN(Q) & \text{scope extrusion} \\ !P & \equiv & P|!P & \text{replication} \end{array}$$

Reaction rules

$$(\cdots + \overline{x}\langle z\rangle.Q)|(\cdots + x(y).P) \to Q|P\{z/y\} \qquad \text{communication (COMM)}$$

$$\frac{P \to P'}{P|Q \to P'|Q} \qquad \text{reaction under parallel composition (PAR)}$$

$$\frac{P \to P'}{(\text{new } x)P \to (\text{new } x)P'} \qquad \text{reaction under restriction (RES)}$$

$$\frac{Q \equiv P \quad P \to P' \quad P' \equiv Q'}{Q \to Q'} \qquad \text{structural congruence (STRUCT)}$$

Syntax

Chemical Mixing

Reactions

Stochastic π-calculus

- Stochastic extension of π-calculus. [C.Priami]

  Associate a single parameter r (**rate**) in (0, infinity] of an **exponential distribution** to each activity a; it describes the stochastic behavior of the activity

  a.P is replaced by (a, r).P

  Exponential distribution guarantees the **memoryless property**: the time at which a change of state occurs is independent of the time at which the last change of state occurred.

  Race condition is defined in a **probabilistic competitive** context: all the activities that are enabled in a state compete and the fastest one succeeds.

- New implementation: SPiM. [A.Phillips]. Paper at BioConcur.

Proteins

MAPK Cascade - Huang&Ferrell

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.

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.

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.

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

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

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

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

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

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

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

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

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

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

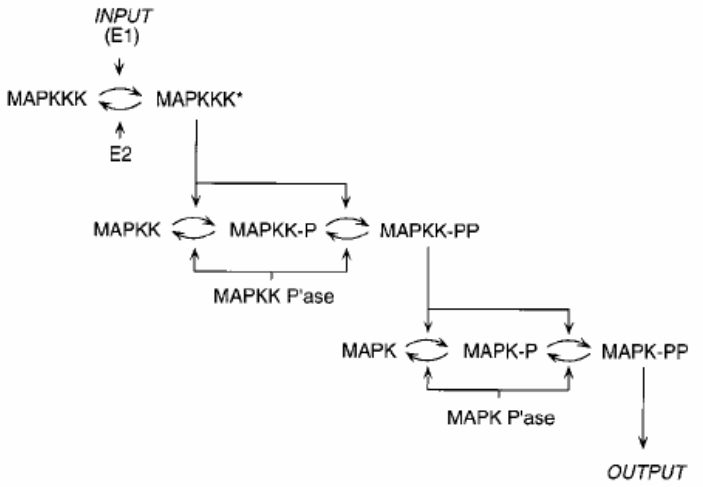


FIG. 1. Schematic view of the MAPK cascade. Activation of MAPK depends upon the phosphorylation of two conserved sites [Thr-183 and Tyr-185 in rat p42 MAPK/Erk2 (4, 5)]. Full activation of MAPKK also requires phosphorylation of two sites [Ser-218 and Ser-222 in mouse Mek-1/MKK1 (6–10)]. Detailed mechanisms 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.

As 18 Ordinary Differential Equations

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

… Plus 7 conservation equations

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

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

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

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

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

Each molecule in exactly one state

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

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

$$[\text{E1}_{tot}] = [\text{E1}] + [\text{KKK}\cdot\text{E1}] \quad [30]$$

$$[\text{E2}_{tot}] = [\text{E2}] + [\text{KKK}^*\cdot\text{E2}] \quad [31]$$

$$[\text{KK}_{tot}] = [\text{KK}] + [\text{KK-P}] + [\text{KK-PP}] + [\text{KK}\cdot\text{KKK}^*] + [KK\text{-}P\cdot KKK^*] + [KK\text{-}P\cdot KK\,P'ase] + [KK\text{-}PP\cdot KK\,P'ase] + [KK\text{-}PP\cdot K] + [KK\text{-}PP\cdot K\text{-}P] \quad [32]$$

$$[\text{KK P'ase}_{tot}] = [\text{KK P'ase}] + [\text{KK P'ase}\cdot\text{KK-P}] + [KK\,P'ase\cdot KK\text{-}PP] \quad [33]$$

$$[\text{K}_{tot}] = [\text{K}] + [\text{K-P}] + [\text{K-PP}] + [\text{KK-PP}\cdot\text{K}] + KK\text{-}PP\cdot K\text{-}P] + [K\text{-}P\cdot K\,P'ase] + [K\text{-}PP\cdot K\,P'ase] \quad [34]$$

$$[\text{K P'ase}_{tot}] = [\text{K P'ase}] + [\text{K-P}\cdot\text{K P'ase}] + [K\text{-}PP\cdot K\,P'ase] \quad [35]$$

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.

MAPK Cascade in SPiM

```
| !KKK();
   new d1:r1:<>
    (a1<d1>;(d1<>;KKK<> + k1<>;KKKst<>))        [1]substrate
| !KKKst();
   new d2:r2:<>
    (a2<d2>;(d2<>;KKKst<> + k2<>;KKK<>) +       [2]substrate
     a3(d3);(d3();KKKst<> + k3();KKKst<>) +     [3]kinase
     a5(d5);(d5();KKKst<> + k5();KKKst<>))      [5]kinase

| !E1();
   a1(d1);(d1();E1<> + k1();E1<>)               [1]enzyme
| !E2();
   a2(d2);(d2();E2<> + k2();E2<>)               [2]enzyme

| !KK();
   new d3:r3:<>
    (a3<d3>;(d3<>;KK<> + k3<>;KK_P<>))          [3]substrate

| !KK_P();
   new d4:r4:<>  new d5:r5:<>
    (a4<d4>;(d4<>;KK_P<> + k4<>;KK<>) +         [4]substrate
     a5<d5>;(d5<>;KK_P<> + k5<>;KK_PP<>))       [5]substrate
```

KKK:E1 complex

E1:KKK complex

```
| !KK_PP();
   new d6:r6:<>
    (a6<d6>;(d6<>;KK_PP<> + k6<>;KK_P<>) +      [6]substrate
     a7(d7);(d7();KK_PP<> + k7();KK_PP<>) +     [7]kinase
     a9(d9);(d9();KK_PP<> + k9();KK_PP<>))      [9]kinase

| !KKPse();
   a4(d4);(d4();KKPse<> + k4();KKPse<>) +       [4]phtase
   a6(d6);(d6();KKPse<> + k6();KKPse<>)         [6]phtase

| !K();
   new d7:r7:<>
    (a7<d7>;(d7<>;K<> + k7<>;K_P<>))            [7]substrate

| !K_P();
   new d8:r8:<>  new d9:r9:<>
    (a8<d8>;(d8<>;K_P<> + k8<>;K<>) +           [8]substrate
     a9<d9>;(d9<>;K_P<> + k9<>;K_PP<>))         [9]substrate

| !K_PP();
   new d10:r10:<>
    (a10<d10>;(d10<>;K_PP<> + k10<>;K_P<>))     [10]substrate

| !KPse();
   a8(d8);(d8();KPse<> + k8();KPse<>) +         [8]phtase
   a10(d10);(d10();KPse<> + k10();KPse<>)       [10]phtase
```

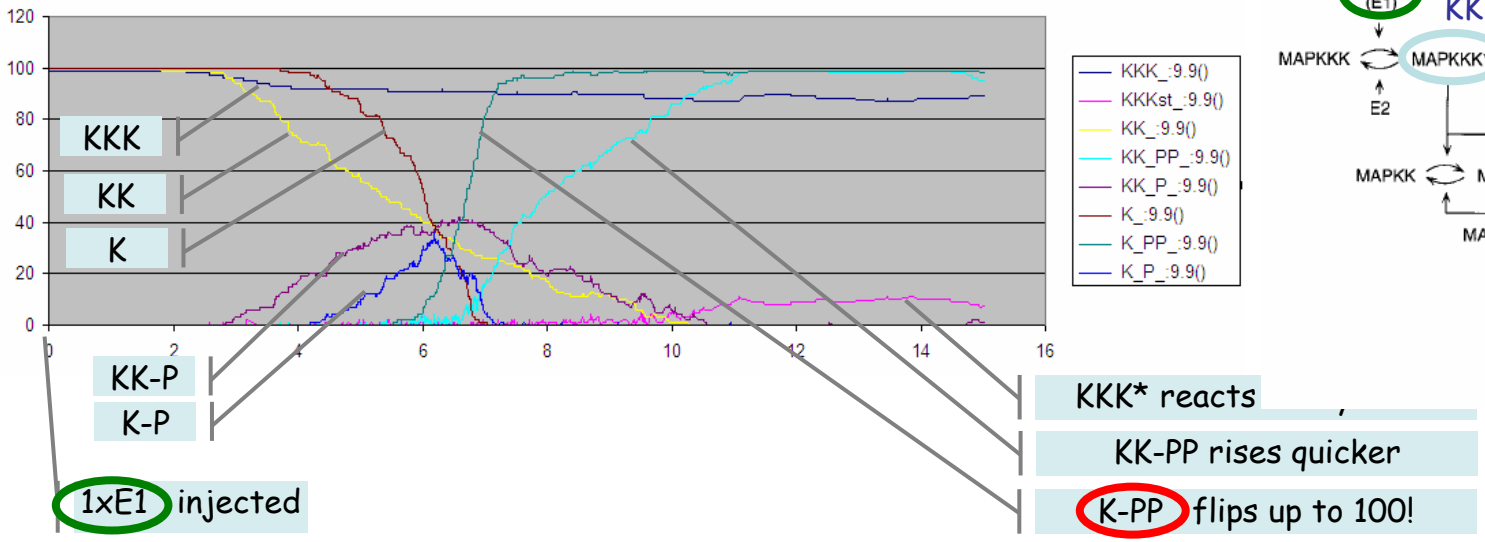
MAPK Cascade in SPiM (the complete program)

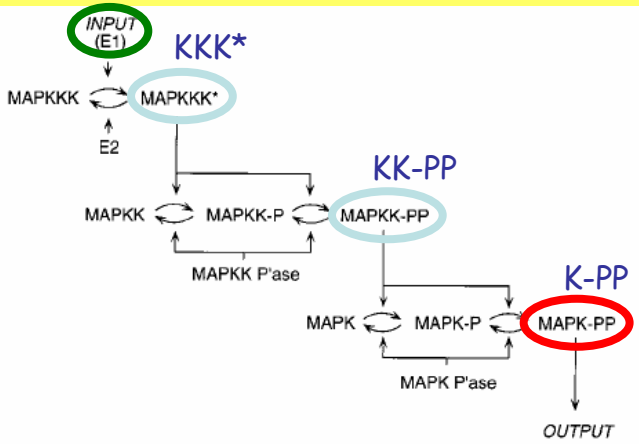
```
new KKK:<> new KKKst:<> new E1:<> new E2:<>
new KK:<> new KK_P:<> new KK_PP:<> new KKPse:<>
new K:<> new K_P:<> new K_PP:<> new KPse:<>

new a1:1.0:<<>> new k1:1.0:<> new a2:1.0:<<>> new k2:1.0:<>
new a3:1.0:<<>> new k3:1.0:<> new a4:1.0:<<>> new k4:1.0:<>
new a5:1.0:<<>> new k5:1.0:<> new a6:1.0:<<>> new k6:1.0:<>
new a7:1.0:<<>> new k7:1.0:<> new a8:1.0:<<>> new k8:1.0:<>
new a9:1.0:<<>> new k9:1.0:<> new a10:1.0:<<>> new k10:1.0:<>

new spike:<<>,int>      (* a spike #2 high of #1 molecules *)
( !spike(a,n); if n=0 then () else (a<> | spike<a,n-1>)

| !KKK();
   new d1:1.0:<>
    (a1<d1>;(d1<>;KKK<> + k1<>;KKKst<>))

| !KKKst();
   new d2:1.0:<>
     (a2<d2>;(d2<>;KKKst<> + k2<>;KKK<>) +
    a3(d3);(d3();KKKst<> + k3();KKKst<>) +
    a5(d5);(d5();KKKst<> + k5();KKKst<>))

| !E1();
   a1(d1);(d1();E1<> + k1();E1<>)

| !E2();
   a2(d2);(d2();E2<> + k2();E2<>)

| !KK();
   new d3:1.0:<>
    (a3<d3>;(d3<>;KK<> + k3<>;KK_P<>))
```

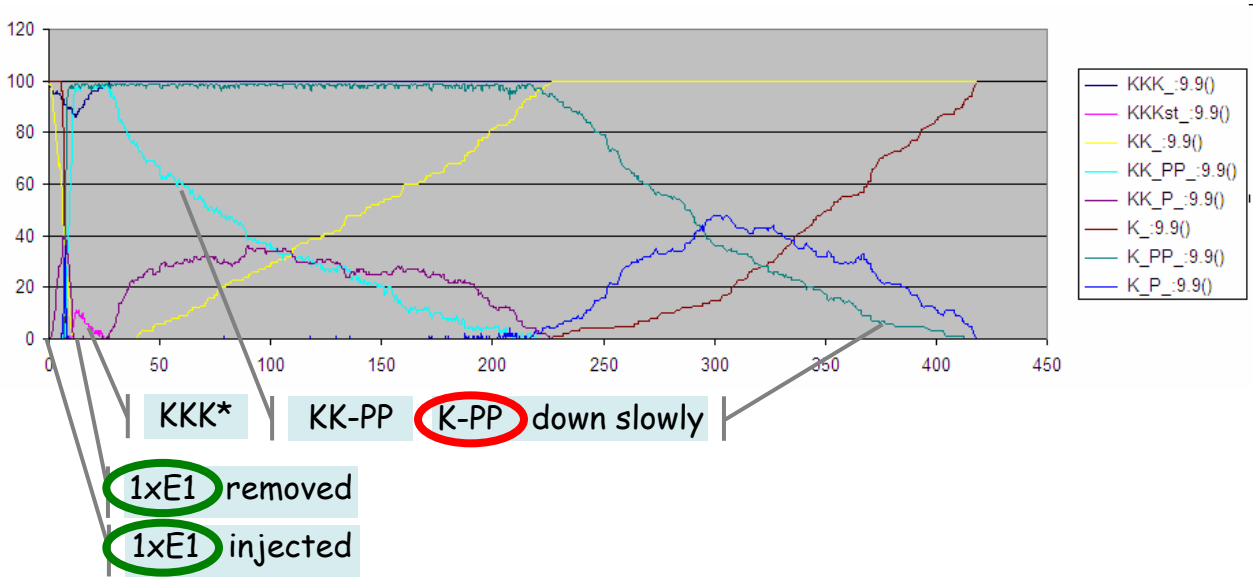
All rates 1.0 !?!

```
| !KK_P();
   new d4:1.0:<>
   new d5:1.0:<>
    (a4<d4>;(d4<>;KK_P<> + k4<>;KK<>) +
     a5<d5>;(d5<>;KK_P<> + k5<>;KK_PP<>))

| !KK_PP();
   new d6:1.0:<>
     (a6<d6>;(d6<>;KK_PP<> + k6<>;KK_P<>) +
    a7(d7);(d7();KK_PP<> + k7();KK_PP<>) +
    a9(d9);(d9();KK_PP<> + k9();KK_PP<>))

| !KKPse();
   a4(d4);(d4();KKPse<> + k4();KKPse<>) +
   a6(d6);(d6();KKPse<> + k6();KKPse<>)

| !K();
   new d7:1.0:<>
    (a7<d7>;(d7<>;K<> + k7<>;K_P<>))

| !K_P();
   new d8:1.0:<>
   new d9:1.0:<>
    (a8<d8>;(d8<>;K_P<> + k8<>;K<>) +
     a9<d9>;(d9<>;K_P<> + k9<>;K_PP<>))

| !K_PP();
   new d10:1.0:<>
    (a10<d10>;(d10<>;K_PP<> + k10<>;K_P<>))

| !KPse();
   a8(d8);(d8();KPse<> + k8();KPse<>) +
   a10(d10);(d10();KPse<> + k10();KPse<>)

| E1<>  (* input signal *)  | E2<> | KKPse<> | KPse<>
| spike<KKK,100> | spike<KK,100> | spike<K,100>   )
```

MAPK Cascade Simulation



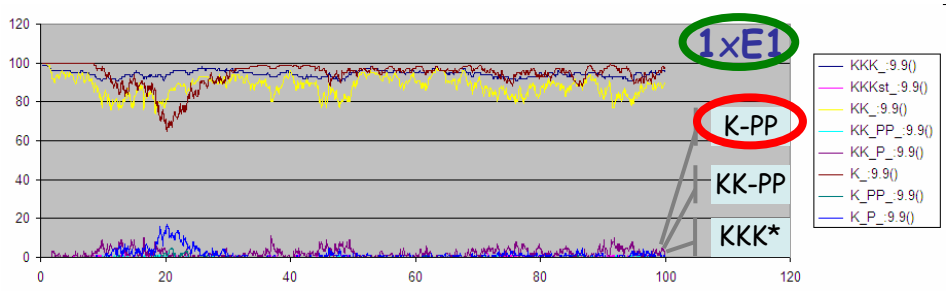


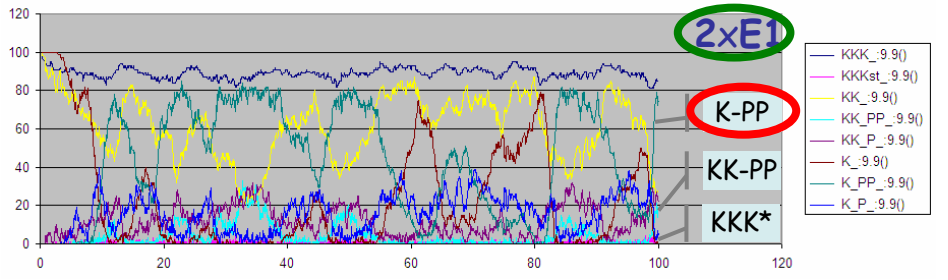


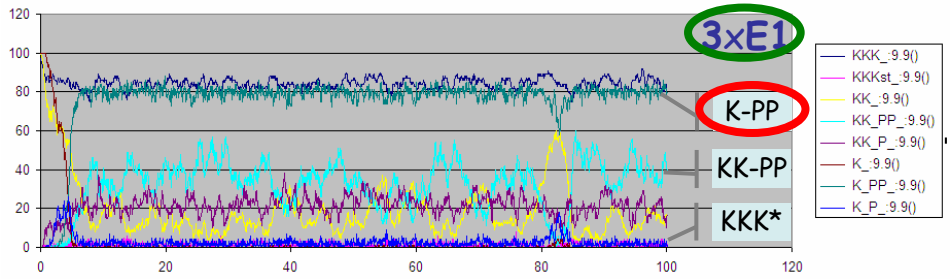
All coefficients 1.0 !!!
100xKKK, 100xKK, 100xK,
1xE2, 1xKKPse, 1xKPse.

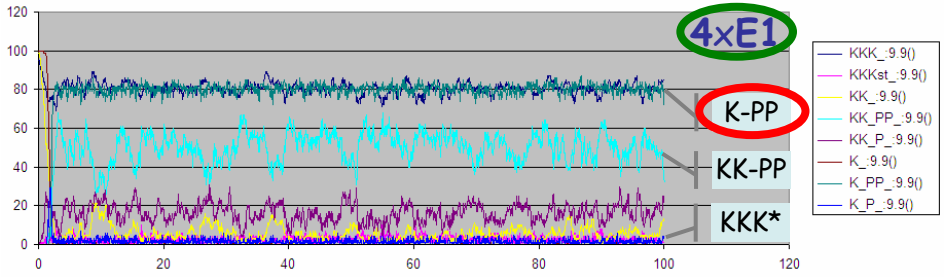
Input is **1xE1**.
Output is **100xK-PP**
(ultrasensitivity).

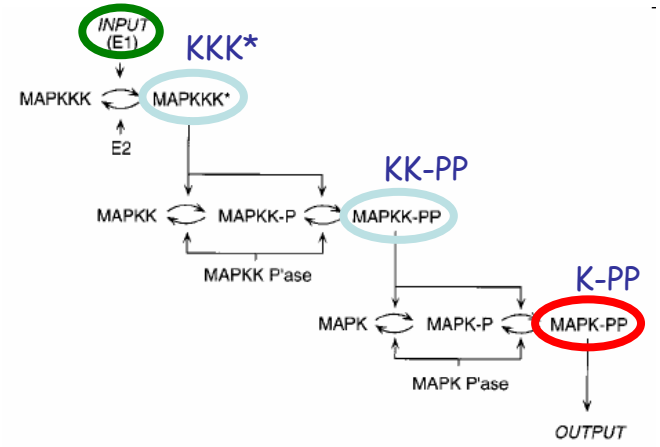
MAPK Cascade Simulation



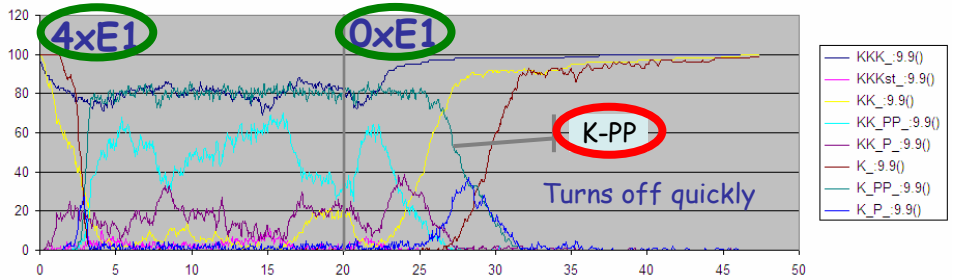








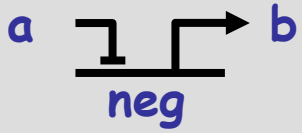
All coefficients 1.0 !!!
100xKKK, 100xKK, 100xK,
10xE2, 10xKKPse, 10xKPse.
(so 1xE1 is no longer sufficient to produce an output)



Genes

Gene Gates and Circuits

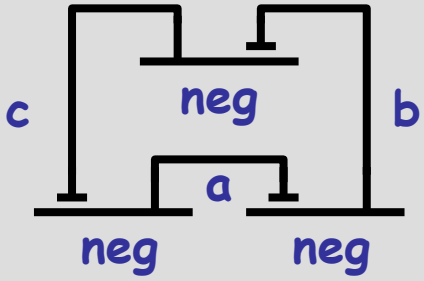
A gene gate



$neg[a,b] \triangleq$
$a_r().\ \tau_h.\ neg[a,b] +$
$\tau_\varepsilon.\ (ptn[b] \mid neg[a,b])$

$ptn[p] \triangleq p_r\langle\rangle.ptn[p] + \tau_\delta$

A genetic circuit (engineered in E.Coli)

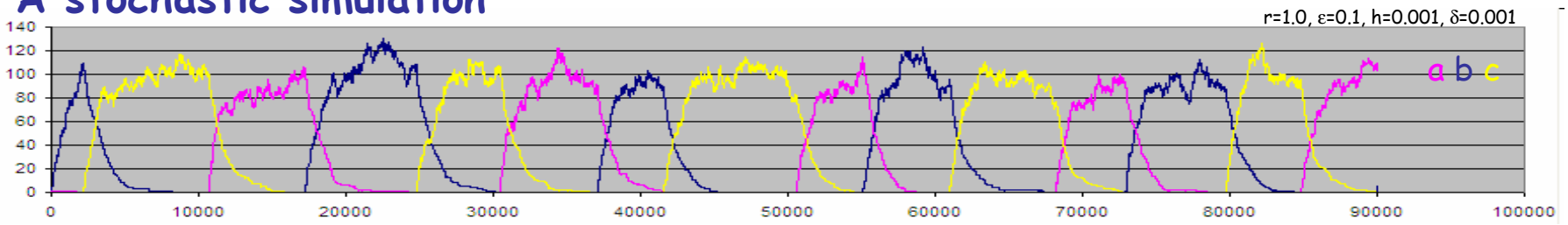


neg[a,b] |
neg[b,c] |
neg[c,a]

The SPiM program

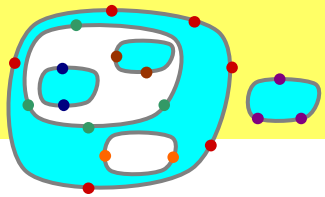
```
new ptn:<<>>          (* Protein *)
new dk:0.001:<>       (* Decay rate *)

new neg:<<>,<>>       (* Neg Gate *)
new tInh:0.001:<>     (* Inhibition rate *)
new tCst:0.1:<>       (* Constitutive rate *)

(* Protein-Gene interactions *)
new a:1.0:<> new b:1.0:<> new c:1.0:<>

( !ptn(p); (p<>;ptn<p>+dk<>;())
| !dk()

| !neg(a,b);
    (a(); (tInh(); neg<a,b>) +
     tCst(); (ptn<b> | neg<a,b>))
| !tCst<> | !tInh<>

(* The circuit *)
| neg<a,b> | neg<b,c> | neg<c,a>
)
```

A stochastic simulation



Membranes

Brane Calculi

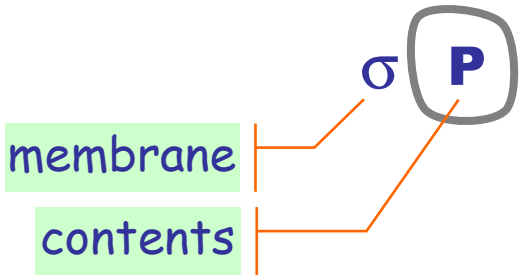


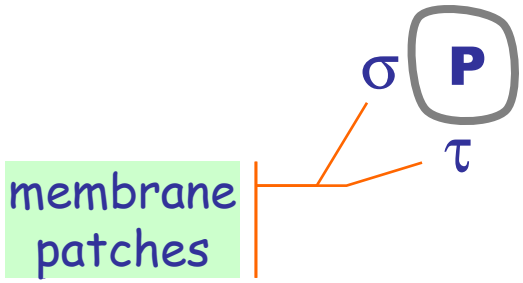
systems	$P,Q ::= \diamond \mid P \circ Q \mid !P \mid \sigma(P)$	nests of membranes
branes	$\sigma,\tau ::= 0 \mid \sigma \mid \tau \mid !\sigma \mid a.\sigma$	combinations of actions
actions	$a ::= 1 \mid \ldots$	(fill in as needed)

1D fluids (σ) inside a 2D fluid (P)

TWO commutative monoids instead of ONE of normal process calculi

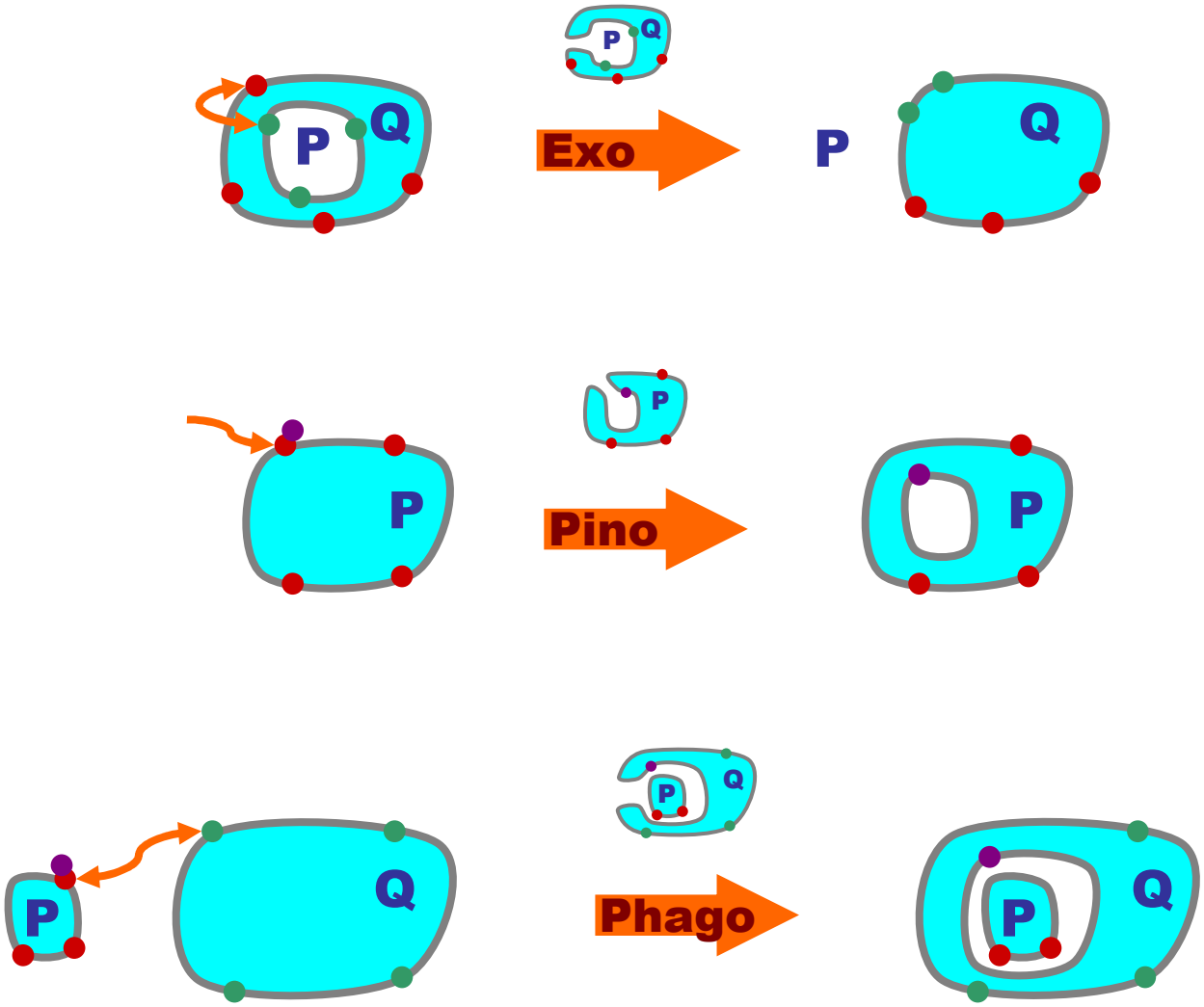
$\sigma(P)$ $\qquad$ $\sigma \mid \tau(P)$ $\qquad$ $a.\sigma \mid \tau = (a.\sigma) \mid \tau$





N.B. Restriction (νn) could be added to both systems and branes. It usually would originate in branes, but would extrude to whole systems.

Brane Reactions (Cartoons)



Brane-Molecule Reactions (Cartoons)

With *molecule multisets* **p**,**q**:

...

Phago $\circlearrowright_n.\sigma|\sigma'\langle P\rangle \circ \circlearrowright^{\perp}_n(\rho).\tau|\tau'\langle Q\rangle \rightarrow \tau|\tau'\langle\rho\langle\sigma|\sigma'\langle P\rangle\rangle\circ Q\rangle$

Exo $\circlearrowright^{\perp}_n.\tau|\tau'\langle\circlearrowright_n.\sigma|\sigma'\langle P\rangle\circ Q\rangle \rightarrow P \circ \sigma|\sigma'|\tau|\tau'\langle Q\rangle$

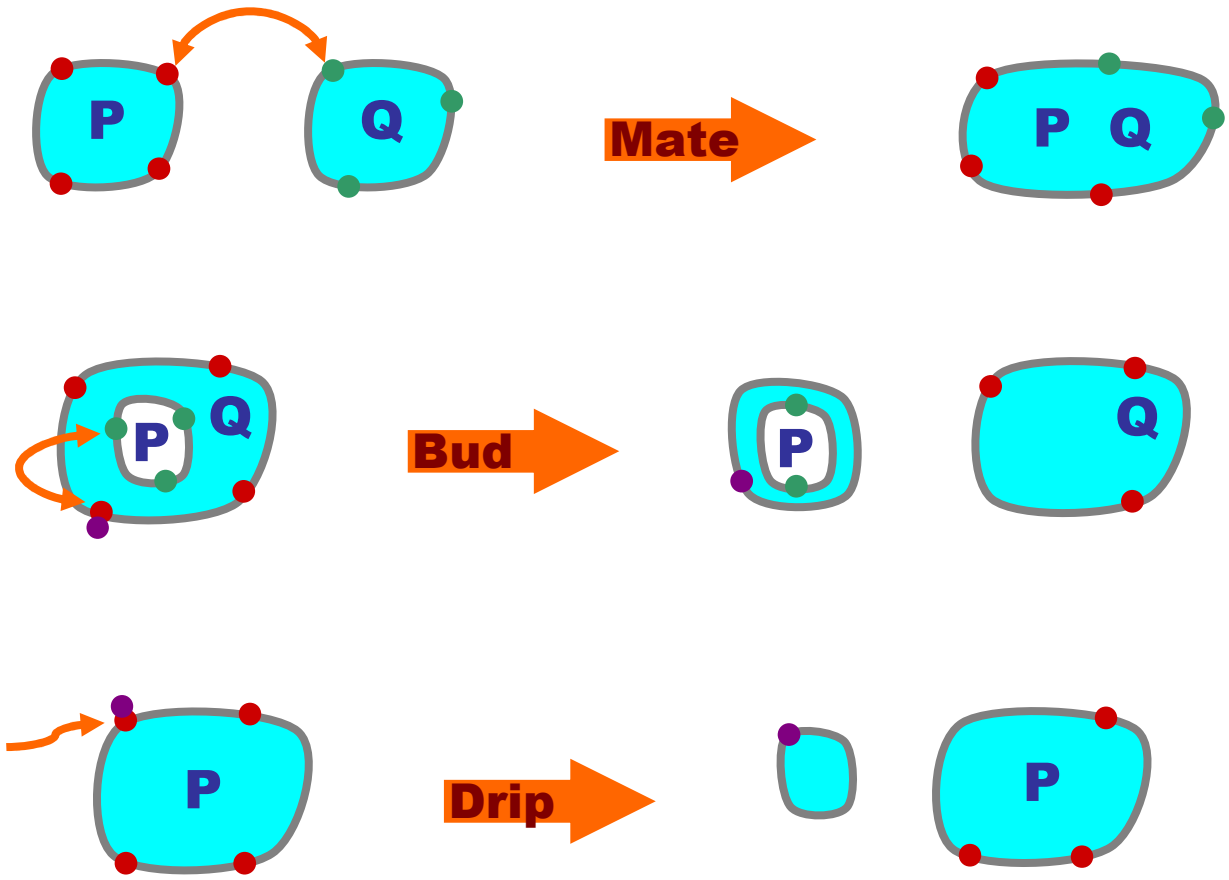
Pino $\circledcirc(\rho).\sigma|\sigma'\langle P\rangle \rightarrow \sigma|\sigma'\langle\rho\langle\diamond\rangle\circ P\rangle$

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

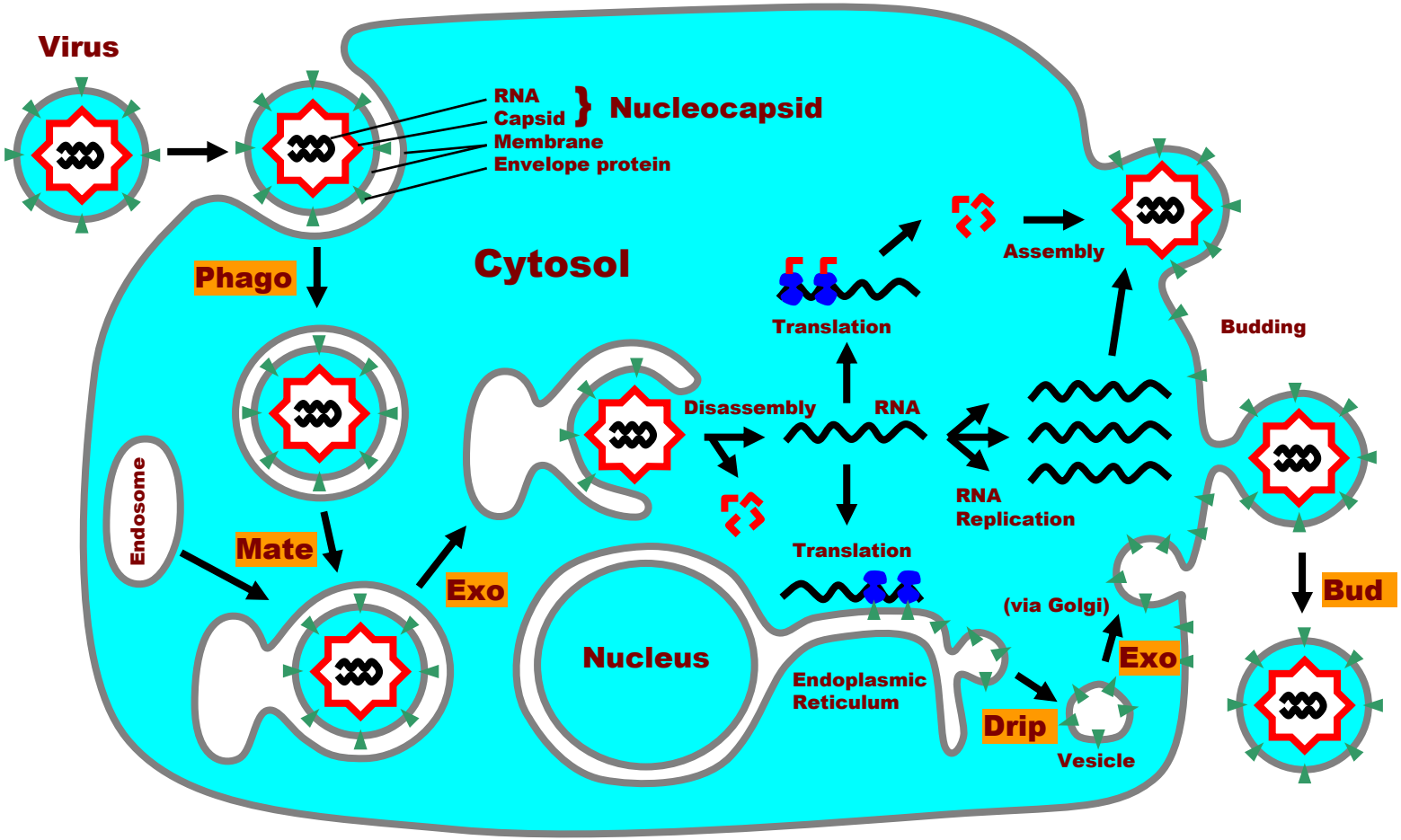
B&R $p_1 \circ p_1(p_2) \rightrightarrows q_1(q_2).\alpha|\sigma\langle p_2 \circ P\rangle \rightarrow q_1 \circ \alpha|\sigma\langle q_2 \circ P\rangle$

(multiset rewriting, inside and outside membranes)

Derivable Reactions (Cartoons)

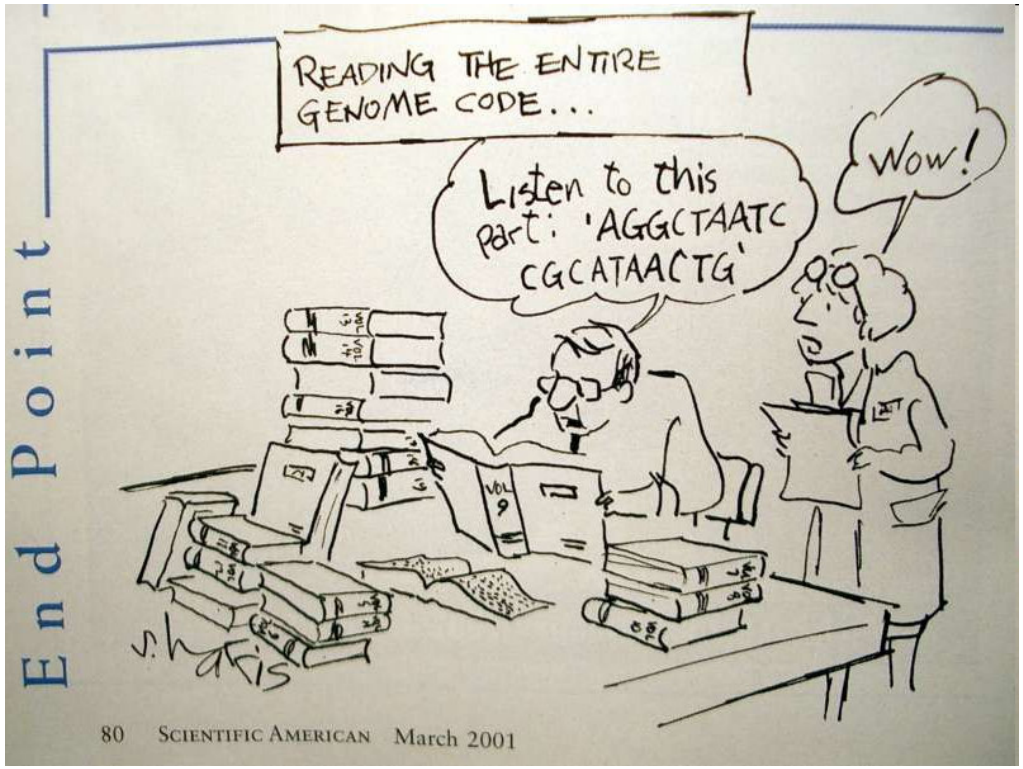


Viral Reproduction



[MBC p.279]
annotated

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.

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