

Languages & Notations for Systems Biology

Luca Cardelli

Microsoft Research
Cambridge UK

2004-05-17 University College London

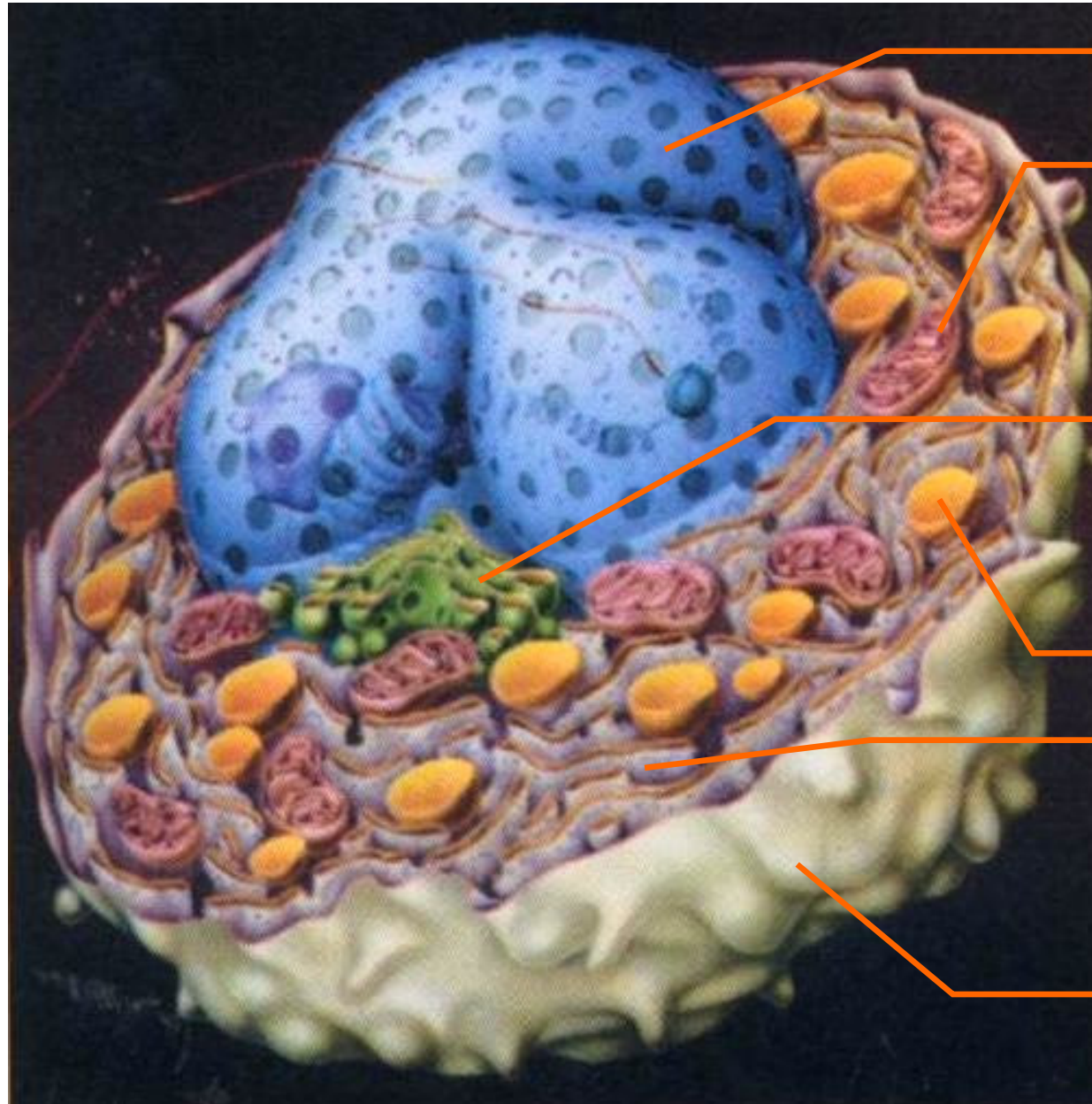
www.luca.demon.co.uk

Structural Architecture

Eukaryotic Cell

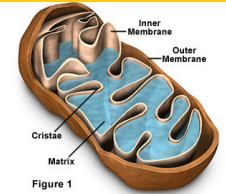
(10~100 trillion in human body)

Membranes everywhere

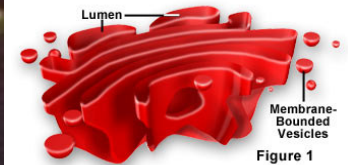


Nuclear membrane

Mitochondria

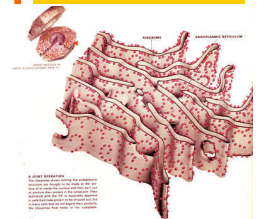


Golgi



Vesicles

E.R.



Plasma membrane (<10% of all membranes)

The Virtual Machines of Biochemistry

A survey of Molecular Biology Notations

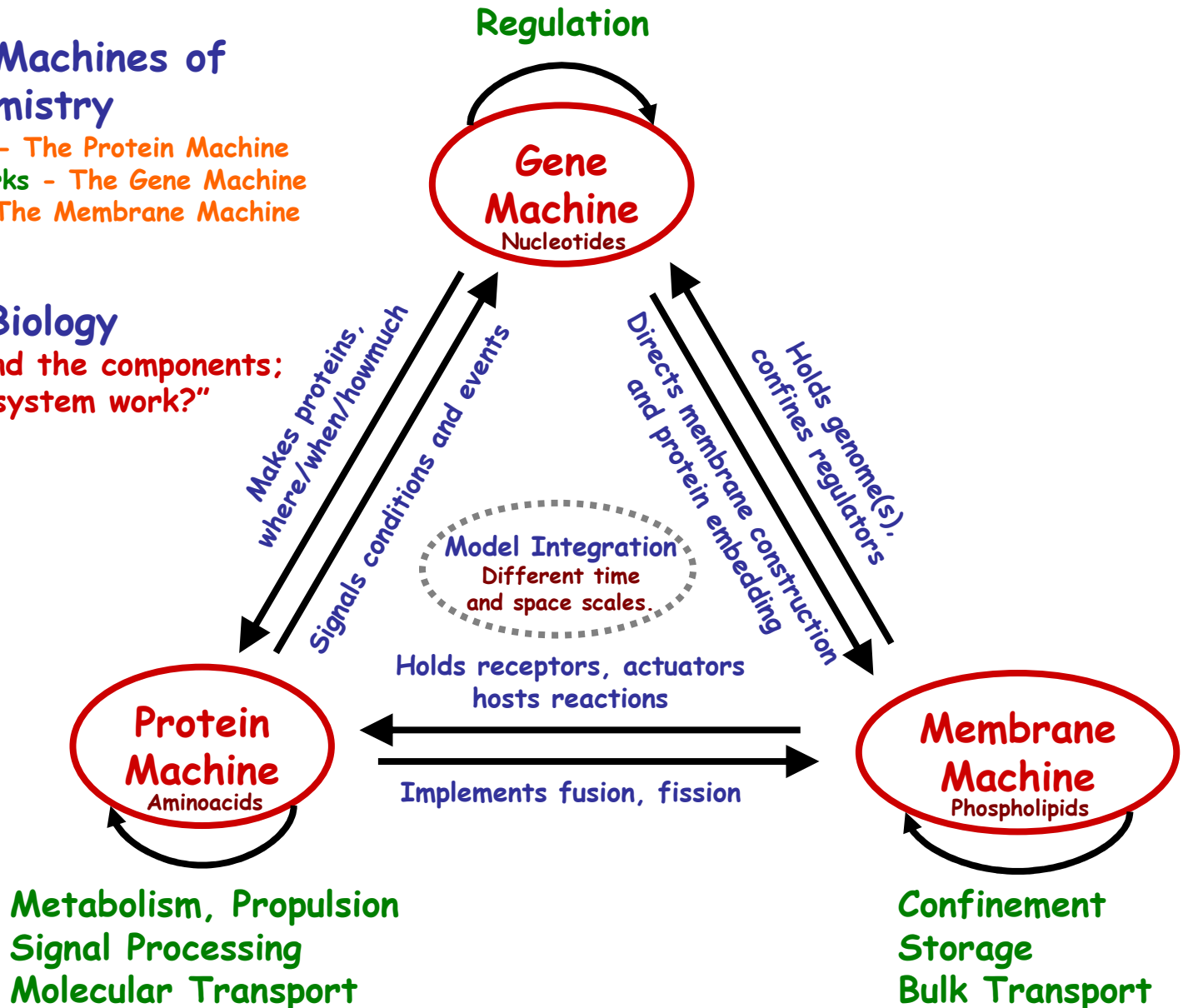
Functional Architecture

The Virtual Machines of Biochemistry

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

Systems Biology

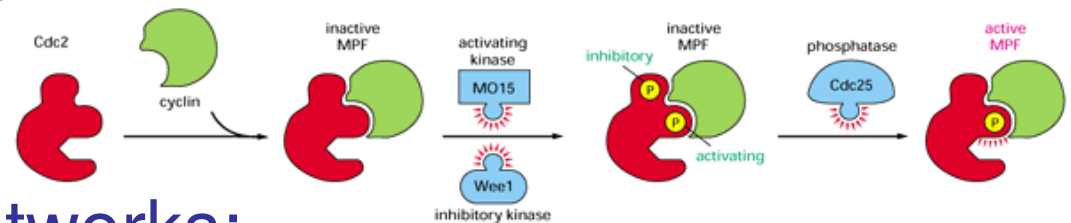
"We (kind of) understand the components; but how does the system work?"



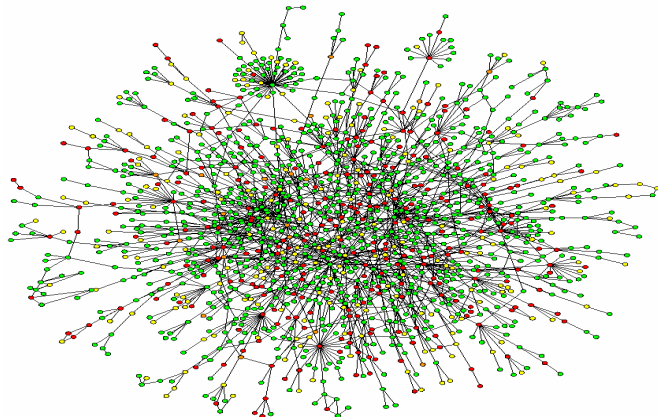
1. The Protein Machine

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.



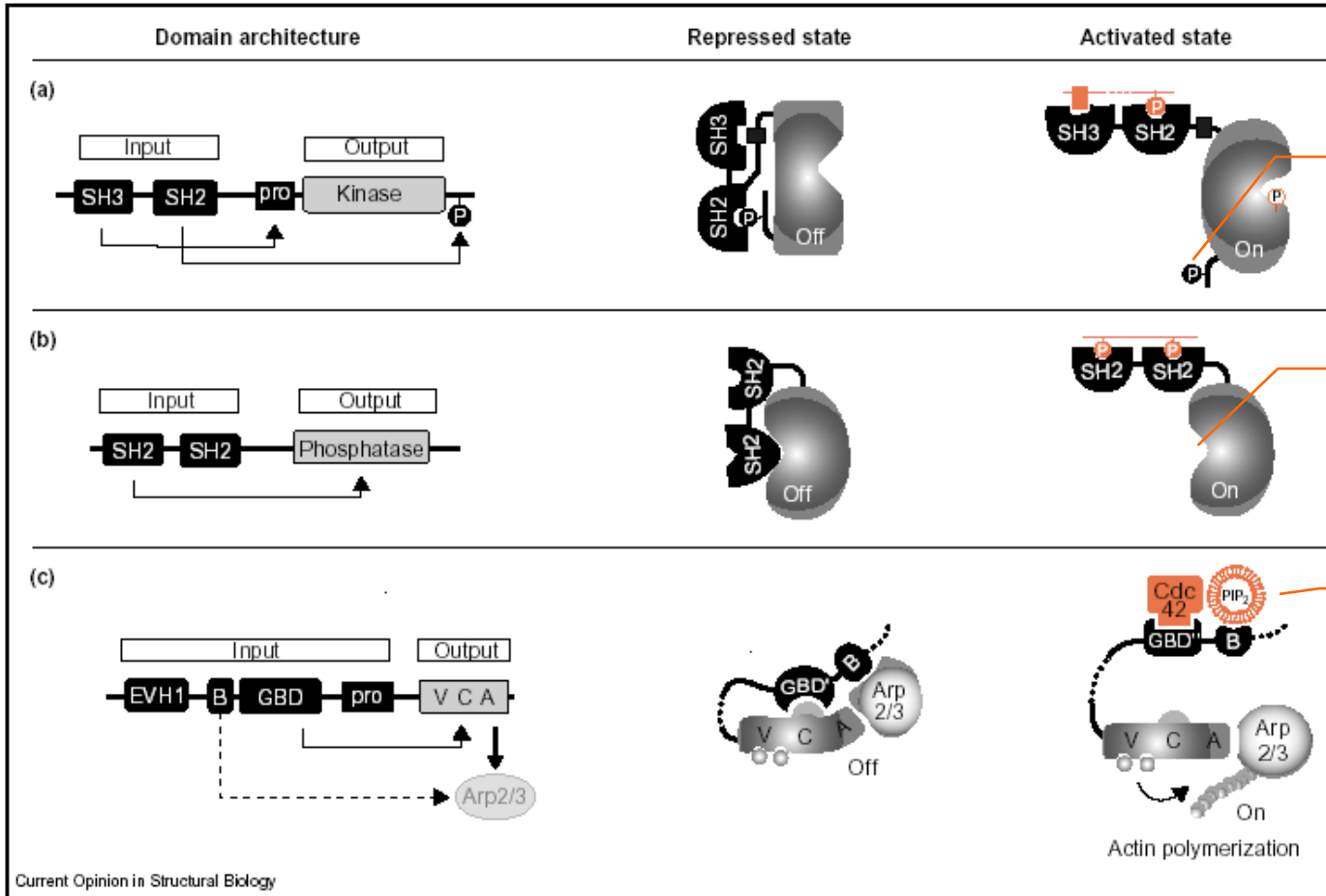
- Mapping out such networks:
 - In principle, it's “just” a very large set of chemical equations.
 - Notations have been developed to summarize and abstract.



An actual molecular interaction network.

(Nodes are distinct protein kinds,
arcs mean that two kinds of proteins interact.)

Some Allosteric Switches



Allosteric ("other shape") reactions modify accessibility.

Kinase

= donates phosphate P
= phosphorylates other proteins

Phosphatase

= accepts phosphate P
= dephosphorylates 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.

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 PIP₂ synergistically activate N-WASP.

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

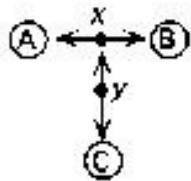
MIM: Molecular Interaction Maps (Kohn)



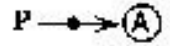
The double-headed line indicates that proteins **A** and **B** can bind to each other. The "node" placed on the line represents the **A:B** complex.



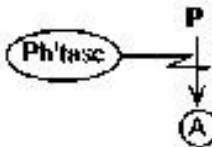
Asymmetric binding where protein **A** donates a peptide that binds to a receptor site or pocket on protein **B**.



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-headed line indicates that **A** can exist in a phosphorylated state. The node represents the phosphorylated species.



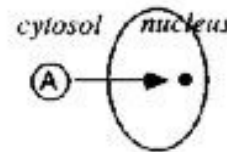
Cleavage of a covalent bond: dephosphorylation of **A** by a phosphatase.



Proteolytic cleavage at a specific site within a protein.



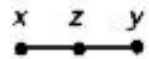
Stoichiometric conversion of **A** into **B**.



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.

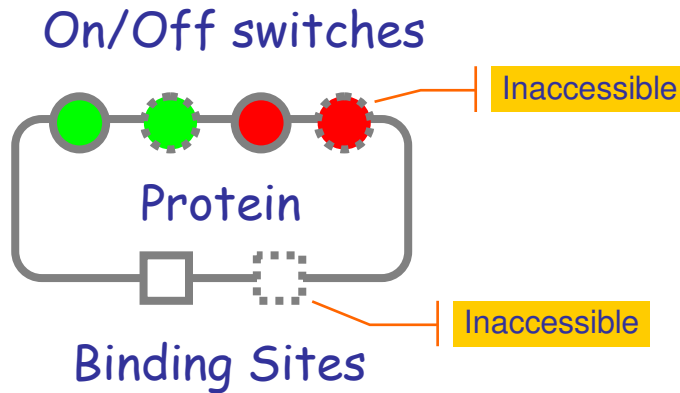


Degradation products

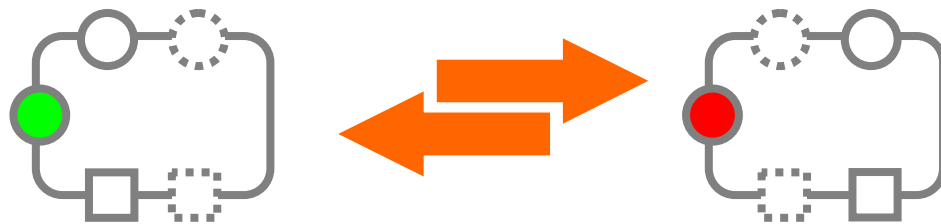
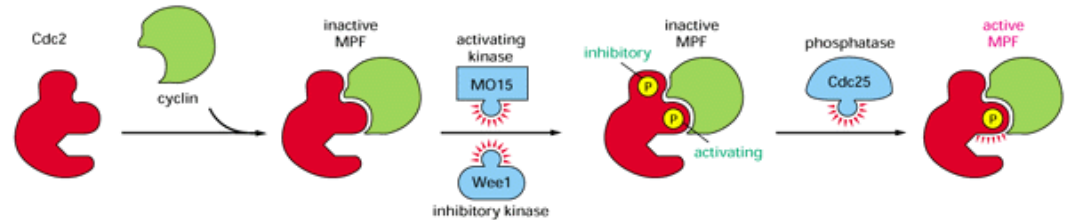
Taken from Kurt W. Kohn

The "Instruction Set" of the Protein Machine

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

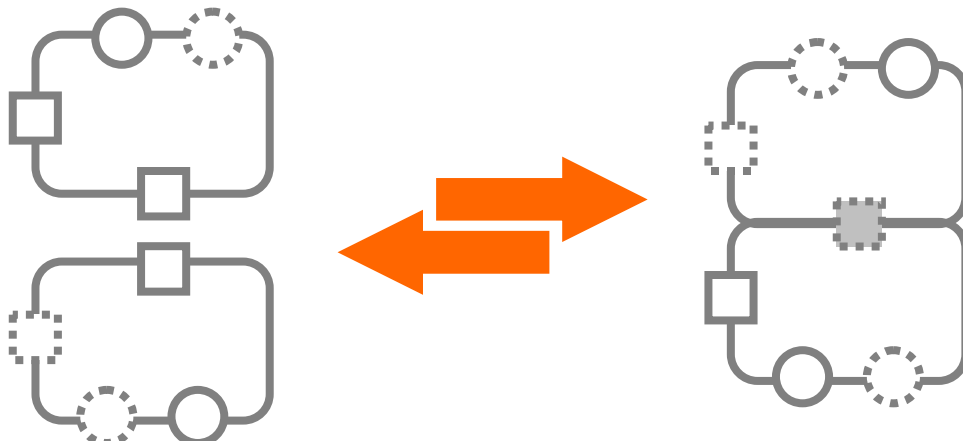


Each protein has a structure of binary switches and binding sites. But not all may be always *accessible*.



Switching of accessible switches.

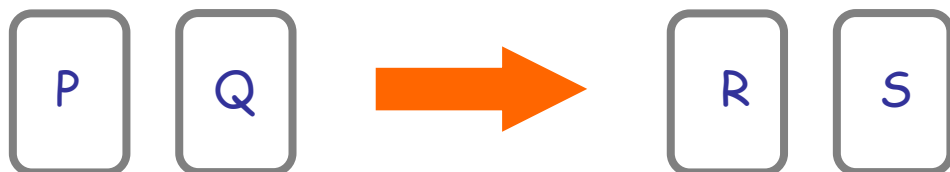
- May cause other switches and binding sites to become (in)accessible.
- May be triggered or inhibited by nearby specific proteins in specific states.



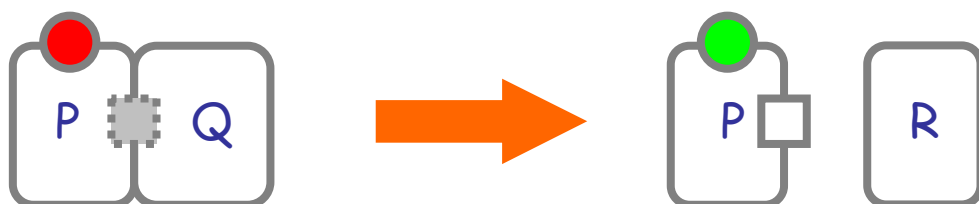
Binding on accessible sites.

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

...



Ordinary Chemical Reactions



Any combination of the above

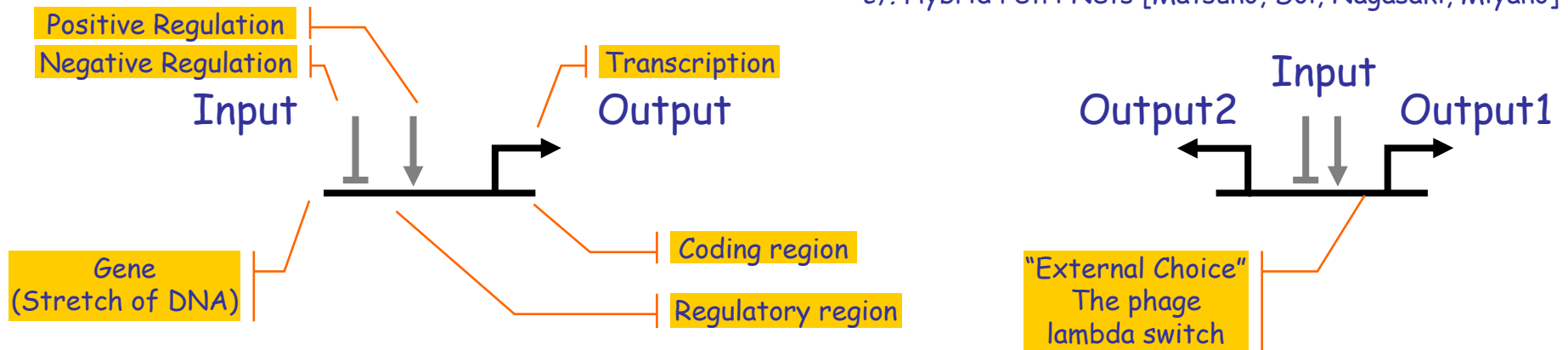
Notations for the Protein Machine

- 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.
- Stochastic π -Calculus
 - Priami formalizes a stochastic version of π -calculus where channels have communication *rates*.
- k-calculus
 - Danos and Laneve (following Kitano's BioCalculus) define a calculus where complex formation is primitive.
- 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.

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)

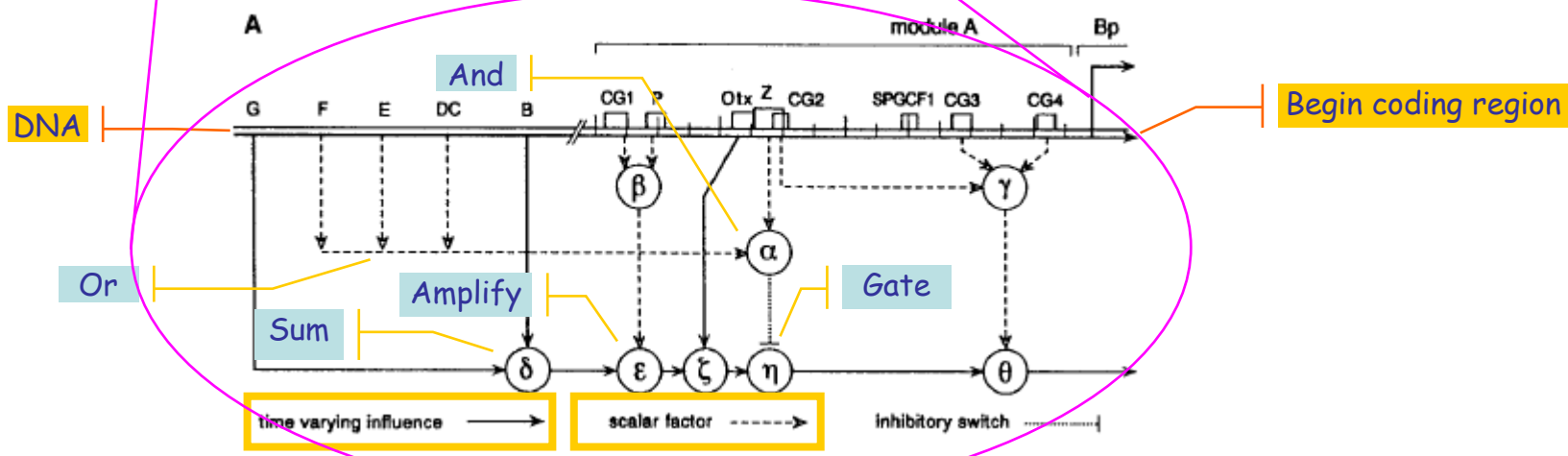
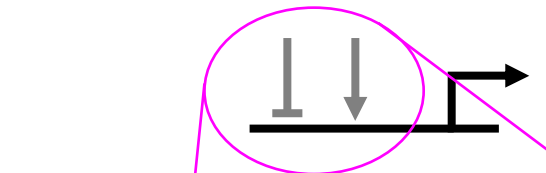
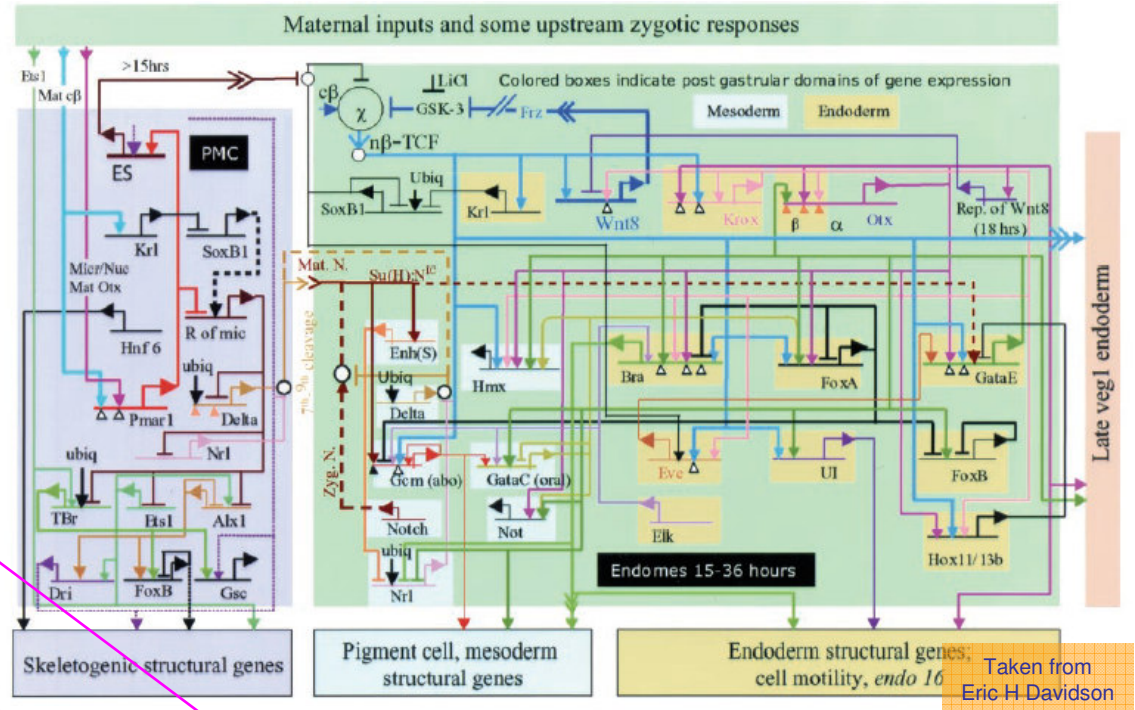
Yeast (eukarya): 12Mbp 3MB (MP3 song)

Wheat 17Gbp 4.25GB (DVD)

Gene Regulatory Networks

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

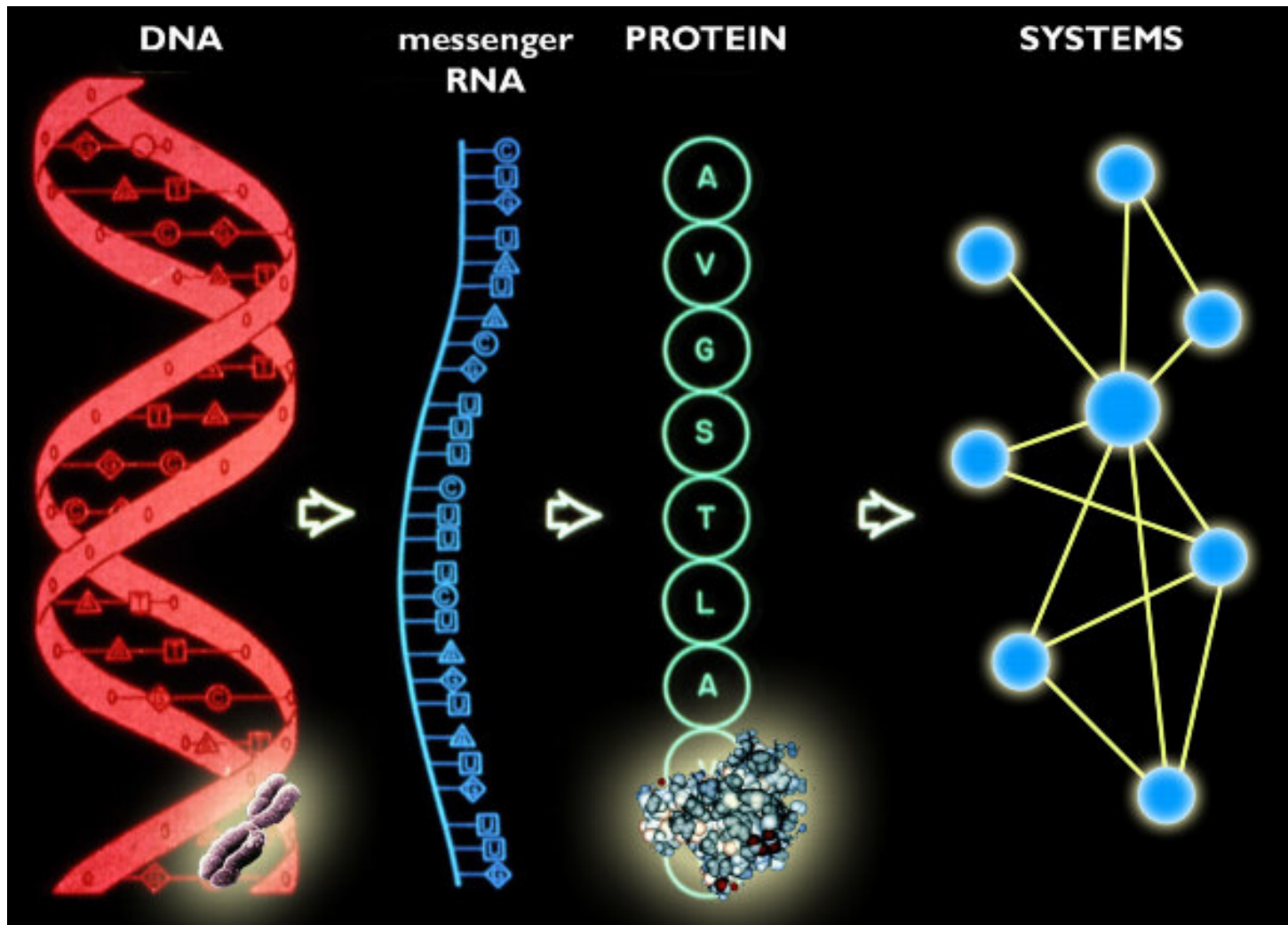
NetBuilder



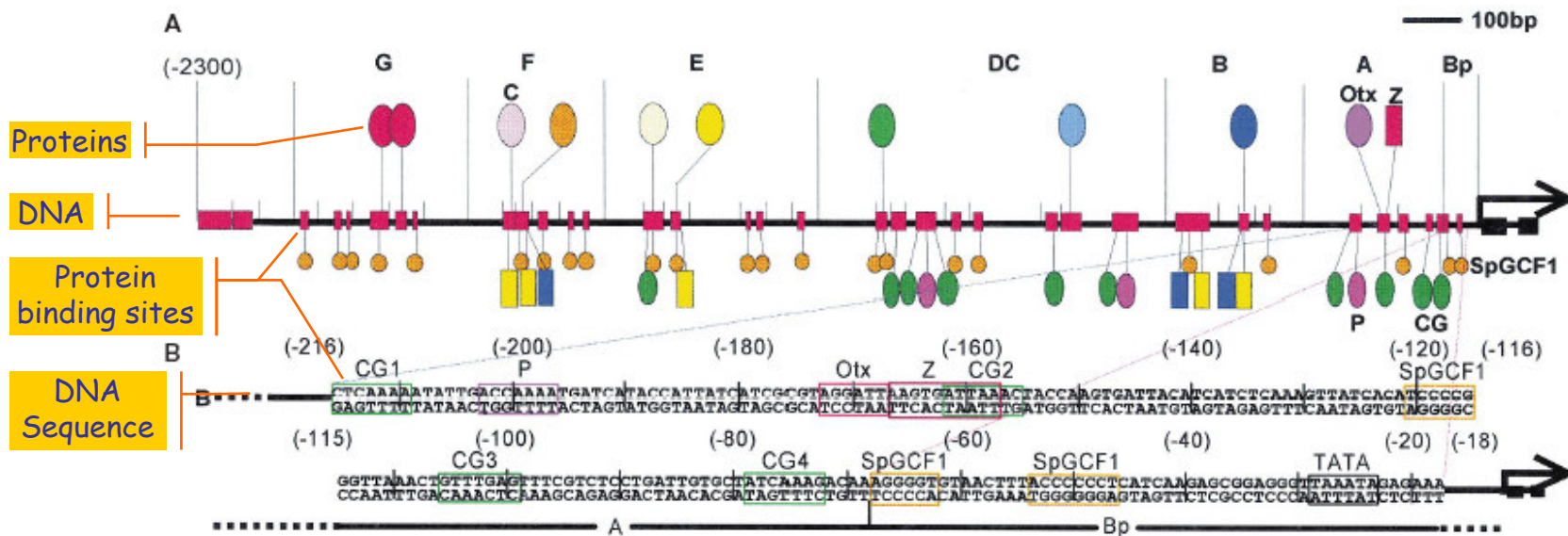
Taken from Eric H. Davidson

Structure of the Coding Region

- The Central Dogma of Molecular Biology:



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:

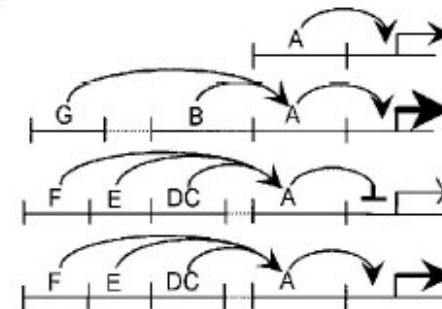
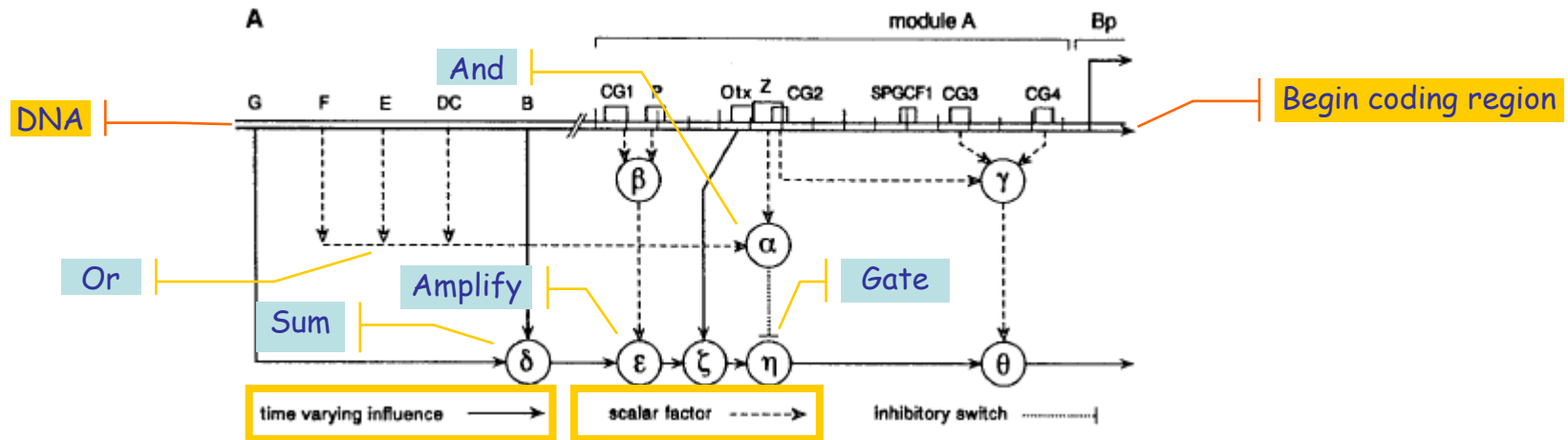


Fig. 1. *Endo16* cis-regulatory system and interactive roles of module A. (A) Diversity of protein binding sites and organization into modular subregions [modified from (7)]. Specific DNA binding sites are indicated as red blocks; modular subregions are denoted by letters G to A (Bp, basal promoter). 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 CG₃ and CG₄ 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



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 CG₁ = 1) Both P and CG₁, needed for synergistic link with module B
 $\beta = 2$
 else $\beta = 0$
 if (CG₂ = 1 and CG₃ = 1 and CG₄ = 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
 $\epsilon(t) = \beta * \delta(t)$ Synergistic amplification of module B output by CG₁-P subsystem

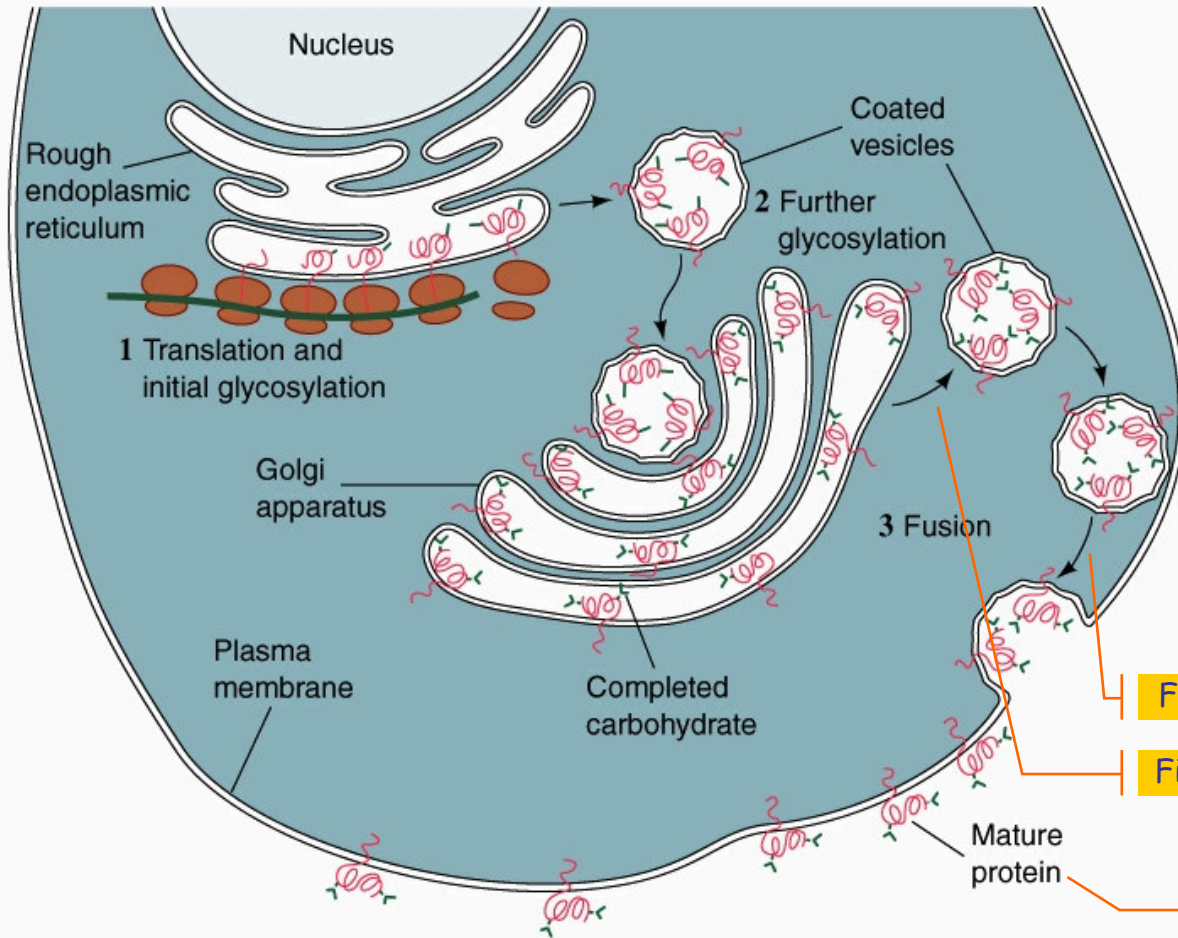
 if ($\epsilon(t) = 0$) Switch determining whether Otx site in module A, or upstream modules (i.e., mainly module B), will control level of activity
 $\xi(t) = Otx(t)$
 else $\xi(t) = \epsilon(t)$
 if ($\alpha = 1$) Repression function inoperative in endoderm but blocks activity elsewhere
 $\eta(t) = 0$
 else $\eta(t) = \xi(t)$
 $\theta(t) = \gamma * \eta(t)$ Final output communicated to BTA

Notations for the Gene Machine

- Hybrid Petri Nets
 - [Matsuno, Doi, Nagasaki, Miyano]
- ...
- Many of the same techniques as for the Protein Machine apply.
 - Process Calculi
 - Term-Rewriting Systems
 - ...
- Gene Regulation Diagrams

3. The Membrane Machine

Very far from the atoms.



Molecular transport and transformation through dynamic compartment fusion and fission.

These "Life of a Saint" diagrams (all temporal stages shown at once) are popular because this is what people actually see in microscopes.

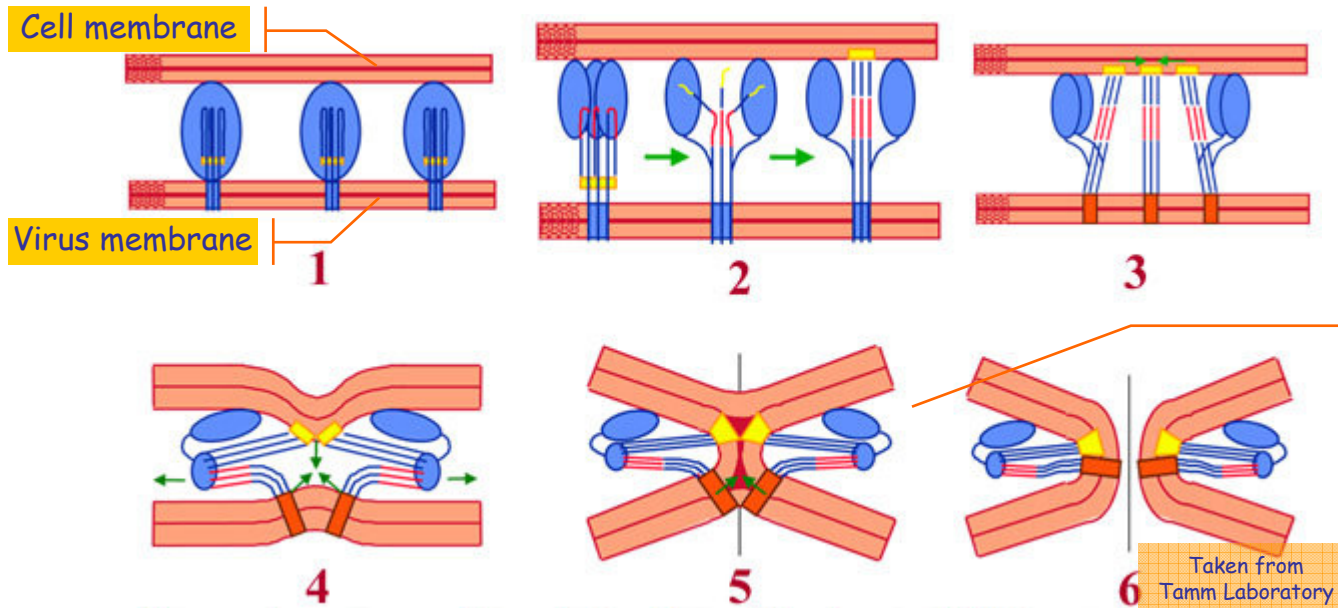
Fusion

Fission

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

Membrane Fusion

Positive curvature to
Negative curvature
transition in 3D

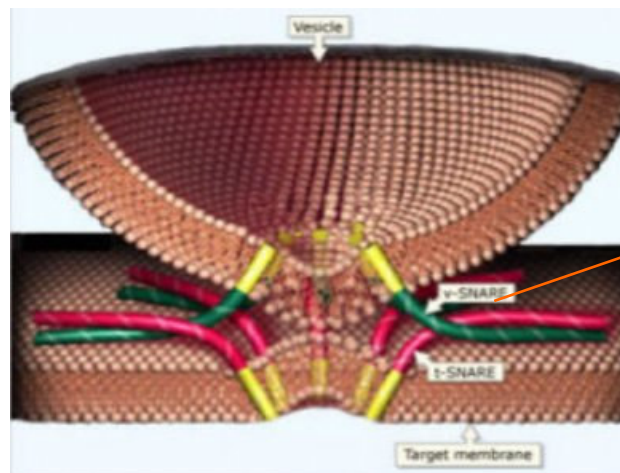


Proposed sequence of events in pH sensitive hemagglutinin membrane fusion

Aggressive fusion (virus)

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

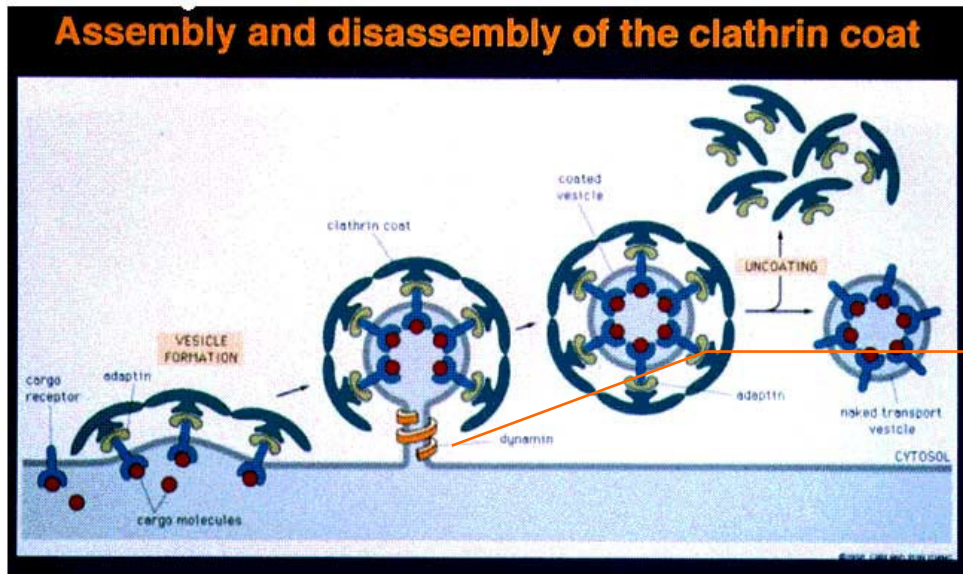
Cooperative fusion (vesicle)



"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

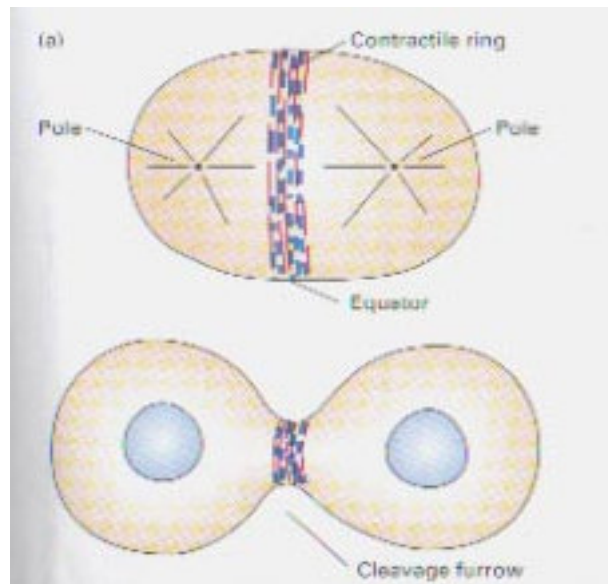


Vesicle Formation



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]

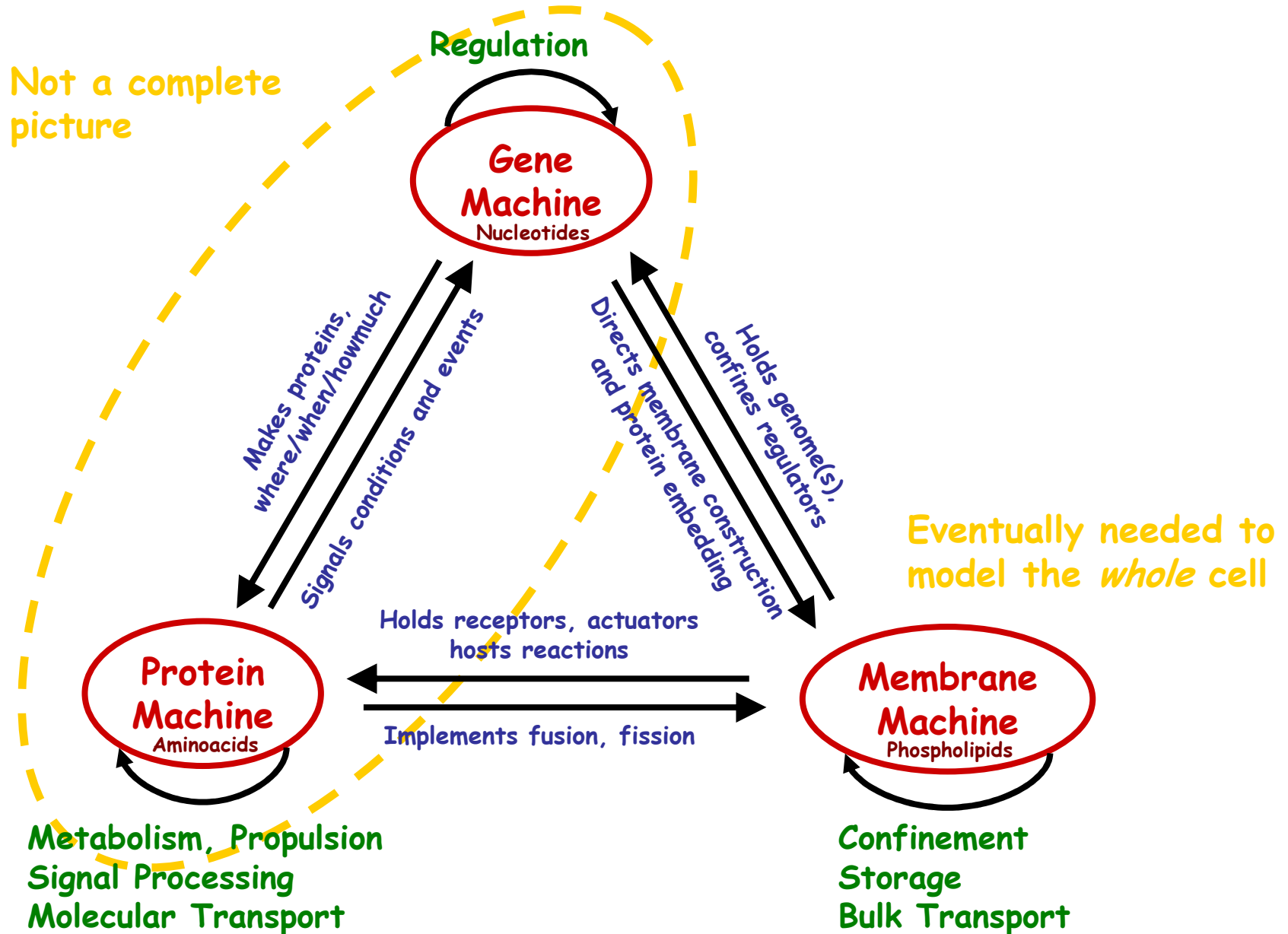


Cytokinesis (Mitosis)

Notations for the Membrane Machine

- 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...

Summary



Modeling with Process Calculi

Write Things Down!

Sydney Brenner: "*When you want to have a predictive science, you have to be able to calculate.*"

When you want to calculate, you have to be able to write things down.

- Write biological systems as **programs**, as if they were software systems
 - Software is a *precise* (yet not quite *predictable*) notation for systems of **high structural and combinatorial complexity**.
 - Small programs can express highly complex behavior. Especially true in nondeterministic concurrency (and in deterministic chaos).
 - We don't use differential equations to write operating systems.
- Write them as **text**, to better describe dynamic behavior
 - Not as cartoons or diagrams
 - Need to choose a **syntax**
 - Always a food fight.
 - But needed for tools to work on: simulation, analysis, storage, search.
 - In C++, Haskell, Prolog?
 - Not likely... We need highly concurrent analyzable formal languages.
 - Representing *processes*, not just data. Concurrency, stochasticity.

Process Calculi

- Write them down with:
 - **Process calculi.**
- Chemistry is ok
 - Yes, chemical reactions *are* a process calculus! In fact, chemical analogies inspired the early definitions of process calculi.
 - But a large biochemical system becomes a flat list of a huge number of reactions: modules and higher-level functional abstractions are lost in the soup.
- Process calculi are:
 - **The modular representation of concurrent processes.**
- They are language-oriented
 - In order to be compositional.
 - Combining separate modules or systems should be easy.
 - To fully represent dynamics.
 - Process evolution should be implicit in process syntax.
 - Graphs don't usually cut it.
 - Either property above can fail in graph-oriented descriptions of processes.
 - Process calculi do not usually make nice pictures.

π -calculus

Syntax

$$\pi ::= x(y) \text{ receive } y \text{ along } x \\ \bar{x}(y) \text{ send } y \text{ along } x$$

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

Structural congruence

Renaming of bound variables

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

Structural congruence laws

$P \mid Q \equiv Q \mid P$	commutativity of parallel composition
$(P \mid Q) \mid R \equiv P \mid (Q \mid R)$	associativity of parallel composition
$P + Q \equiv Q + P$	commutativity of summation
$(P + Q) + R \equiv P + (Q + R)$	associativity of summation
$(\text{new } x)0 \equiv 0$	restriction of inert processes
$(\text{new } x)(\text{new } y)P \equiv (\text{new } y)(\text{new } x)P$	polyadic restriction
$((\text{new } x)P) \mid Q \equiv (\text{new } x)(P \mid Q)$	scope extrusion
$!P \equiv P \mid P$	replication

Reaction rules

$$(\dots + \bar{x}(z).Q) \mid (\dots + x(y).P) \rightarrow Q \mid P \{z/y\} \quad \text{communication (COMM)}$$

$$\frac{P \rightarrow P'}{P \mid Q \rightarrow P' \mid Q} \quad \text{reaction under parallel composition (PAR)}$$

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

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

Syntax

Chemical
Mixing

Reactions

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

Chemistry



System =

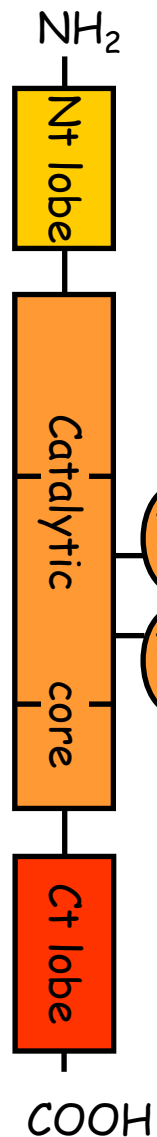


Na	=	e!	NaPlus	ready to give electron
NaPlus	=	e?	Na	ready to take electron
Cl	=	e?	ClMinus	ready to take electron
ClMinus	=	e!	Cl	ready to give electron



Biochemistry

ERK1



Binding MP1 molecules

Regulatory T-loop:
Change conformation

Kinase site:
Phosphorylate Ser/Thr residues (PXT/SP motifs)

ATP binding site:
Bind ATP, and use it for phosphorylation

Binding to substrates

SYSTEM ::= ... | ERK1 | ERK1 | ... | MEK1 | MEK1 | ...

ERK1 ::= (new internal_channels)
(Nt_LOBE | CATALYTIC_CORE | Ct_LOBE)

Domains, molecules, systems ~ Processes

Compartments, membranes ~ Restricted channels

Interaction capability ~ Global channels

Change of future interactions ~ Mobility (channel-passing)

N.B.: Compartments as restriction:

- This is a modeling *strength* of π -calculus: general mechanisms used in new ways
- But is also a modeling *weakness*: indirect representation of basic concepts
- A reason to consider other calculi (later amended by ambient calculus)

Stochastic π -calculus

- Priami: stochastic extension of π -calculus.

Associate a single parameter r (**rate**) in $(0, \text{infinity}]$ 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.

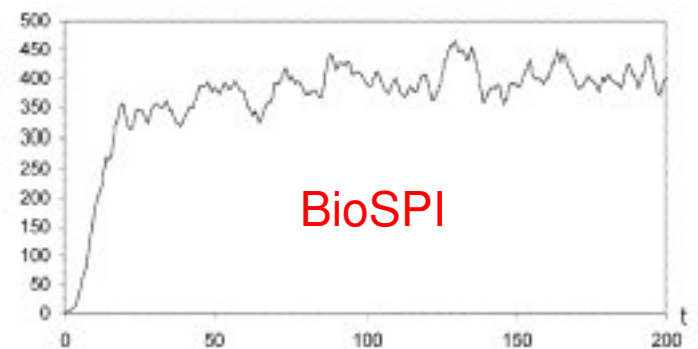
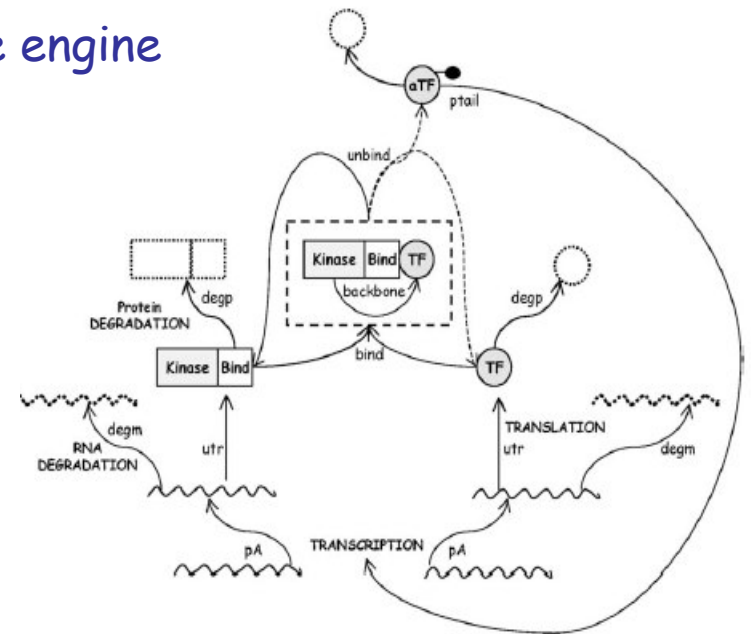
Biochemical Stochastic π -calculus - BioSPI

- Compiles (full) π -calculus to FCP/Logix
- Incorporates Gillespie's algorithm in the runtime engine

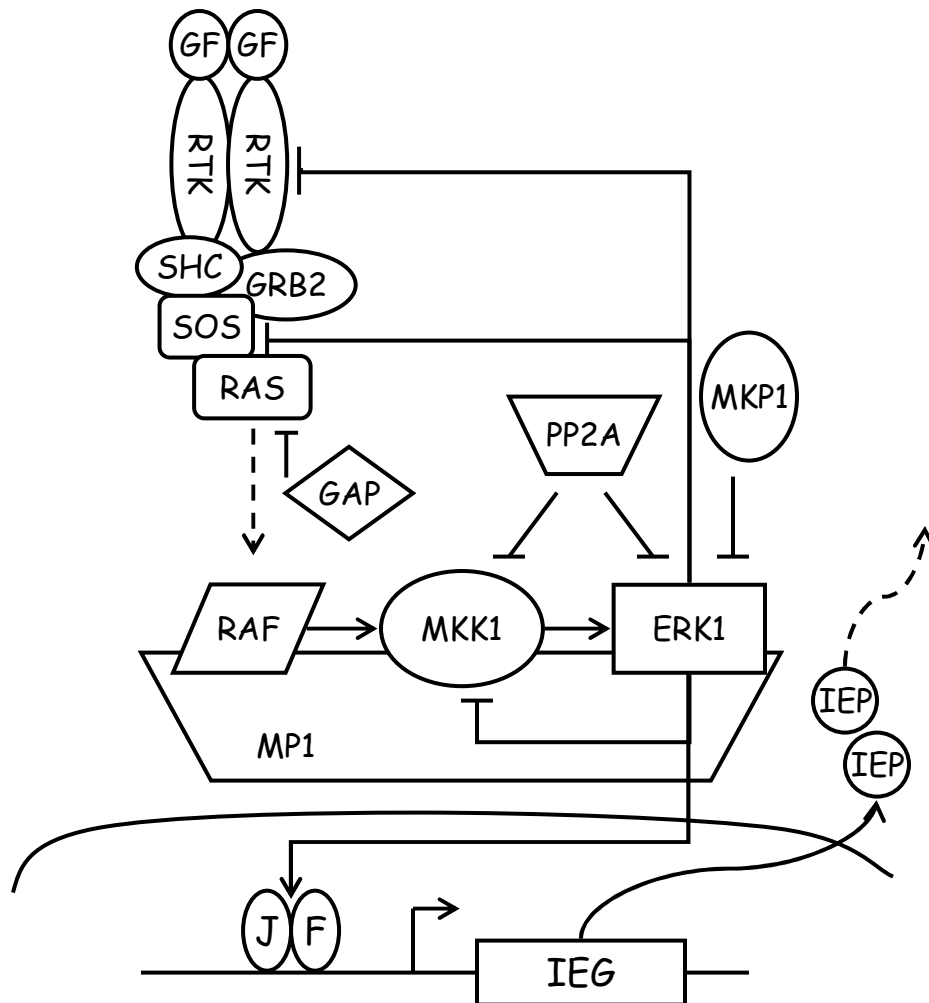
```

Sys = Gene_A|Gene_TF|Transcr|Transl|RNA_Deg|Protein_Deg
Gene_A = (basal(), 4).(Gene_A|RNA_A) + (pA(), 40).(Gene_A|RNA_A)
RNA_A = (utr(), 1).(RNA_A|Protein_A) + (degm(), 1)
Protein_A = ( $\nu$ bb1, bb2, bb3)(Binding_Site|Kinase)
Binding_Site = ( $\overline{\text{bind}}$ (bb1, bb2, bb3), 0.1).Bound_Site + (degp(), 0.1).( $\overline{\text{bb3}}$ ,  $\infty$ )
Bound_Site = ( $\overline{\text{bb1}}$ , 10).Binding_Site + (degp(), 0.1).( $\overline{\text{bb3}}$ ,  $\infty$ ).( $\overline{\text{bb3}}$ ,  $\infty$ )
Kinase = ( $\overline{\text{bb2}}$ (ptail), 10).Kinase + (bb3(),  $\infty$ )
Gene_TF = (basal(), 4).(Gene_TF|RNA_TF) + (pA(), 40).(Gene_TF|RNA_TF)
RNA_TF = (utr(), 1).(RNA_TF|Protein_TF) + (degm(), 1)
Protein_TF = (bind(c_bb1, c_bb2, c_bb3), 0.1).Bound_TF + (degp(), 0.1)
Bound_TF = (c_bb1(), 10).Protein_TF + (c_bb3(),  $\infty$ ) +
(c_bb2(tail), 10).((c_bb1(), 10).Active_TF(tail) + (c_bb3(),  $\infty$ ))
Active_TF(tail) = ( $\overline{\text{tail}}$ , 100).Active_TF(tail) + (degp(), 0.1)
Transcr = ( $\overline{\text{basal}}$ , 4).Transcr + (ptail(), 100).( $\overline{\text{pA}}$ , 40).Transcr
Transl = ( $\overline{\text{utr}}$ , 1).Transl
RNA_Deg = ( $\overline{\text{degm}}$ , 1).RNA_Deg
Protein_Deg = ( $\overline{\text{degp}}$ , 0.1).Protein_Deg
    
```

Transcriptional Regulation By Positive Feedback



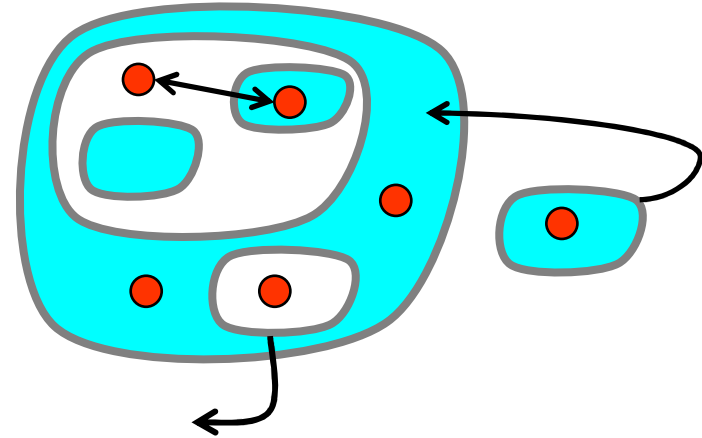
The RTK-MAPK pathway



- 16 molecular species
- 24 domains; 15 sub-domains
- Four cellular compartments
- Binding, dimerization, phosphorylation, de-phosphorylation, conformational changes, translocation
- ~100 literature articles
- 250 lines of code

BioAmbients

- An extension of BioSpi
 - Dynamic membranes: operations for merging, splitting, interacting through membrane channels.
 - Good abstraction:
 - partitions subsystems
 - models membranes as a whole
 - Implemented by Aviv Regev.
- An adaptation of Ambient Calculus
 - A process language for dynamic containers (mobile agents, distributed locations, etc.)
 - This is where I came in.



Processes can communicate across membranes

Membranes are processes; they can move in and out of other membranes

Multi-Level Organization

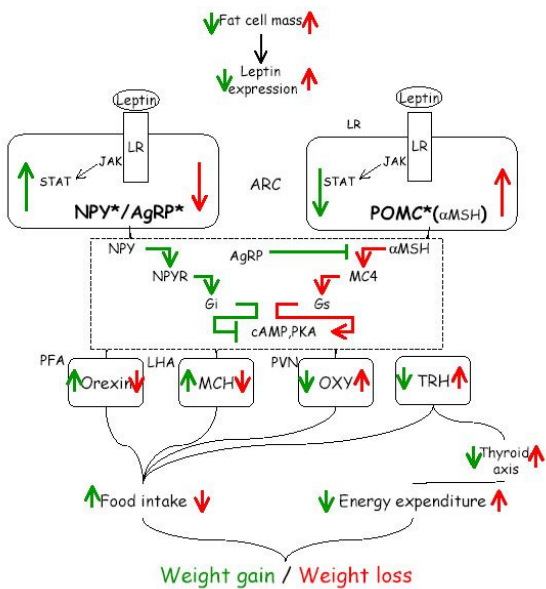


Fig. 16. Hypothalamic pathways for weight regulation. A partial view of molecular pathways, neurons and nuclei involved in weight control. Orexinergic (weight gaining) signals are in green, anorexigenic (weight loss) ones are in red. For further details see the main text. Adapted from [17].

Weight regulation system from the literature.

Example chosen because it involves several levels of biological organization: molecular, cellular, and anatomical.

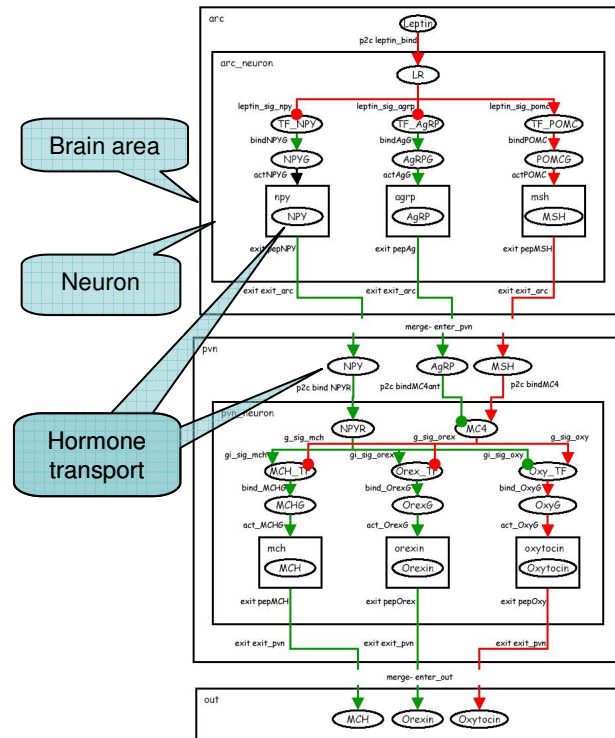


Fig. 17. A scheme for an ambient calculus model for hypothalamic weight regulation. The ambient model is depicted graphically, with ambients as rectangles and molecular processes as ovals. Channel names used in communications or capabilities are shown as labeled arrows, green and red for orexinergic and anorexigenic signals, respectively. Pointed arrowheads represent activatory events, round heads for inhibitory events.

BioAmbients representation of the same system.

(Schematic representation of a BioAmbients script, hand-drawn)

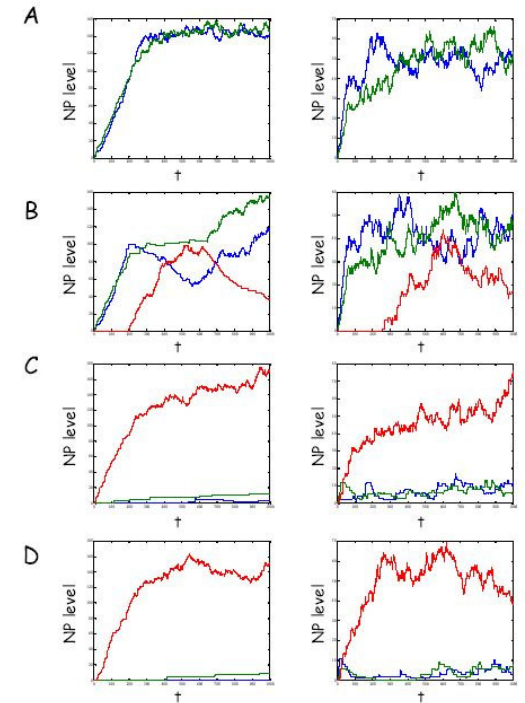


Fig. 18. Neuro peptide profiles under different levels of Leptin. Simulation results of neuro peptide levels under various Leptin creation rates (A) 0.0001 (B) 0.01 (C) 1 (D) 100. In each panel first order hormones, AgRP (blue), NPY (green), and MSH (red) are shown on the right, and second order hormones, MCH (blue), Orexin (green) and Oxytocin (red) are shown on the left. The anorexigenic hormones (red) are high when leptin levels are high, while the orexinergic ones (green,blue) are high when Leptin is low, as expected.

Stochastic simulation (1 neuron per functional area, ~100 receptors per neuron)

N.B.: discrete processes: thousands of components are enough, not billions.

Model Validation: Stochastic Simulation

- Basic algorithm: Gillespie
 - Exact (i.e. based on physics) stochastic simulation of chemical kinetics.
 - Can compute concentrations and reaction times for biochemical networks.
- Stochastic Process Calculi
 - BioSPi [Shapiro, Regev, Priami, et. al.]
 - Stochastic process calculus based on Gillespie.
 - BioAmbients [Regev, Panina, Silverma, Cardelli, Shapiro]
 - Extension of BioSpi for membranes.
 - Case study: Lymphocytes in Inflamed Blood Vessels [Lecaa, Priami, Quaglia]
 - Original analysis of lymphocyte rolling in blood vessels of different diameters.
 - Case study: Lambda Switch [Celine Kuttler, IRI Lille]
 - Model of phage lambda genome (well-studied system).
 - Case study: VICE [U. Pisa]
 - Minimal prokaryote genome (180 genes) and metabolism of *whole* VIRTUAL CELL, in stochastic π -calculus, simulated under stable conditions for 40K transitions.
- More traditional approaches
 - Charon language [UPenn]
 - Hybrid systems: continuous differential equations + discrete/stochastic mode switching.
 - Etc.

Model Validation: "Program" Analysis

- Causality Analysis

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

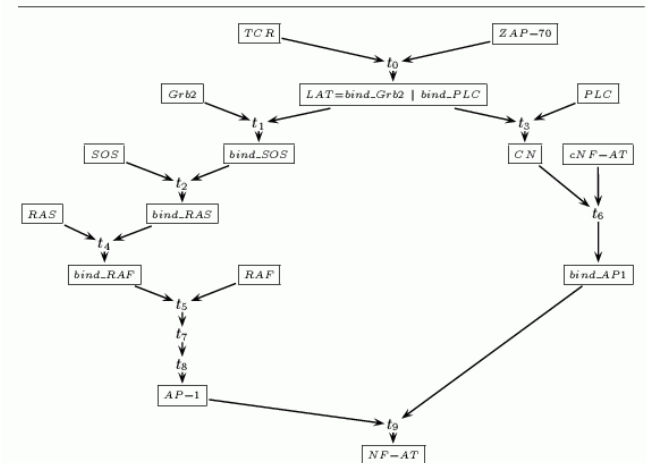


Fig.2. A computation of *Sys*. For readability, the processes, enclosed in boxes, have no address. Causality (both on transitions and processes) is represented by the (Hasse diagram resulting from the) arrows; their absence makes it explicit concurrent activities.

- Control Flow Analysis

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

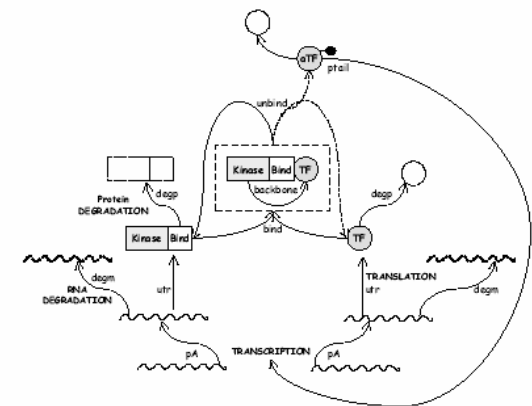


Fig. 1. Graphical presentation of Transcriptional Regulation by Positive Feedback [25].

- Probabilistic Abstract Interpretation

- [DiPierro Wicklicky].

Model Validation: Modelchecking

- **Temporal: NuSMV**

[Chabrier-Rivier Chiaverini Danos Fages Schachter]

- Analysis of mammalian cell cycle (after Kohn) in CTL.
 - E.g. is state S_1 a necessary checkpoint for reaching state S_2 ?

- **Quantitative: Simpathica/xssys**

[Antioniotti Park Policriti Ugel Mishra]

- Quantitative temporal logic queries of human Purine metabolism model.

Eventually(Always (PRPP = 1.7 * PRPP1))

implies

steady_state()

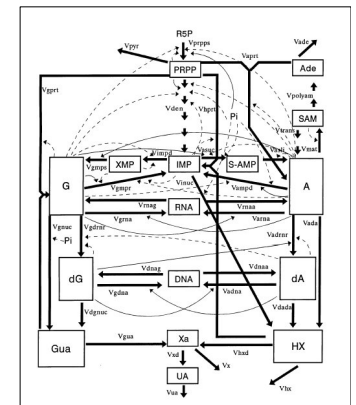
and Eventually(Always(IMP < 2 * IMP1))

and Eventually(Always(hx_pool < 10*hx_pool1))

- **Stochastic: Spring**

[Parker Normal Kwiatkowska]

- Designed for stochastic (computer) network analysis
 - Discrete and Continuous Markov Processes.
 - Process input language.
 - Modelchecking of probabilistic queries.

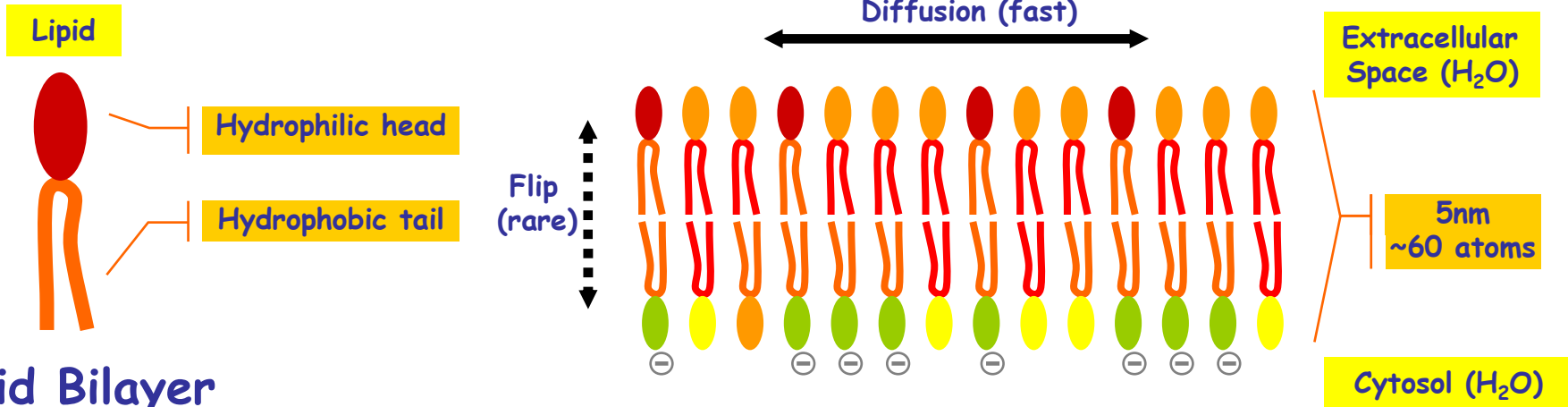


What Process Calculi Do For Us

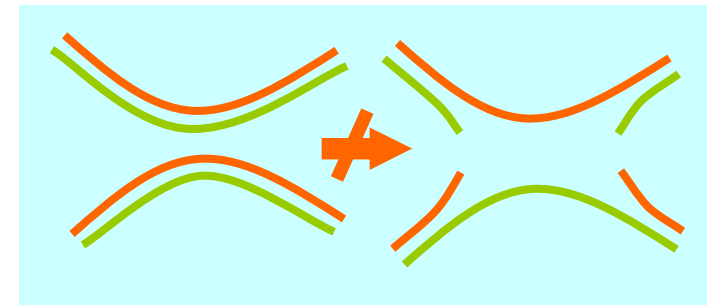
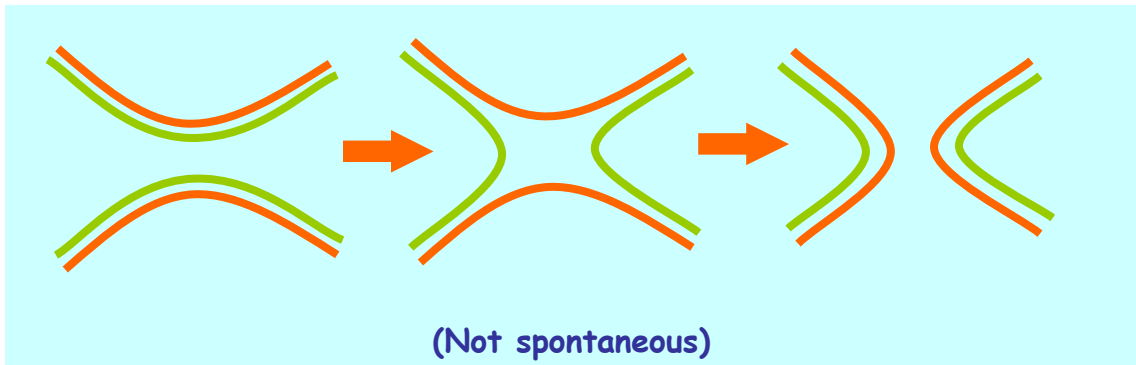
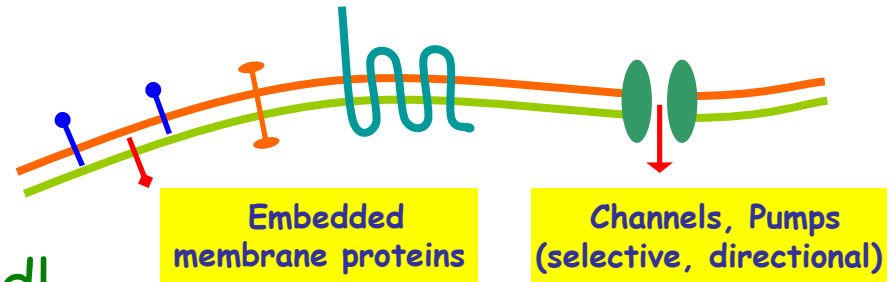
- **We can write things down**
 - We can modularly describe high structural and combinatorial complexity ("do programming").
 - Software teaches us that large and deep systems, even well engineered ones where each component is rigidly defined, eventually exhibit "emergent behavior" (damn!).
- **We can calculate and analyze**
 - Directly support simulation.
 - Support analysis (e.g. control flow, causality, nondeterminism).
 - Support state exploration (modelchecking).
 - This was invented to discover "emergent behavior" (=bugs) in software and hardware systems.
 - Should have interesting large-scale applications in biology.
- **We can reason**
 - Suitable equivalences on processes induce algebraic laws.
 - We can relate different systems (e.g. equivalent behaviors).
 - We can relate different abstraction levels.
 - We can use equivalences for state minimization (symmetries).
- **Disclaimers**
 - Some of these technologies are basically ready (medium-scale stochastic simulation and analysis, medium-scale nondeterministic and stochastic modelchecking).
 - Others need to scale up significantly to be really useful. This is (has been) the challenge for computer scientists.

Membrane Systems

Membranes are Oriented 2D Surfaces



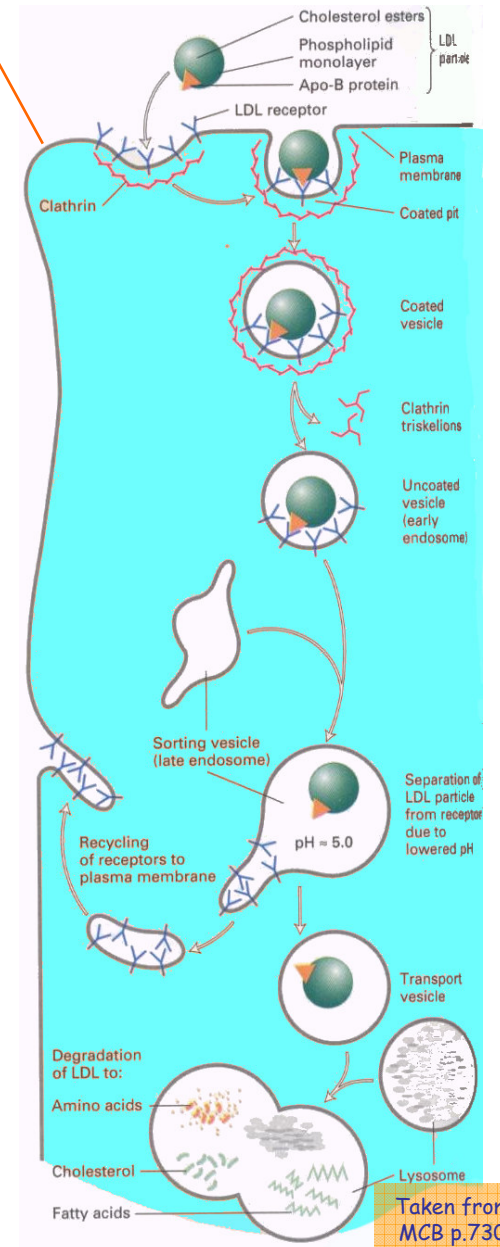
Lipid Bilayer
 Self-assembling
 Largely impermeable
 Asymmetrical (in real cells)
 With embedded proteins
A 2D fluid inside a 3D fluid!



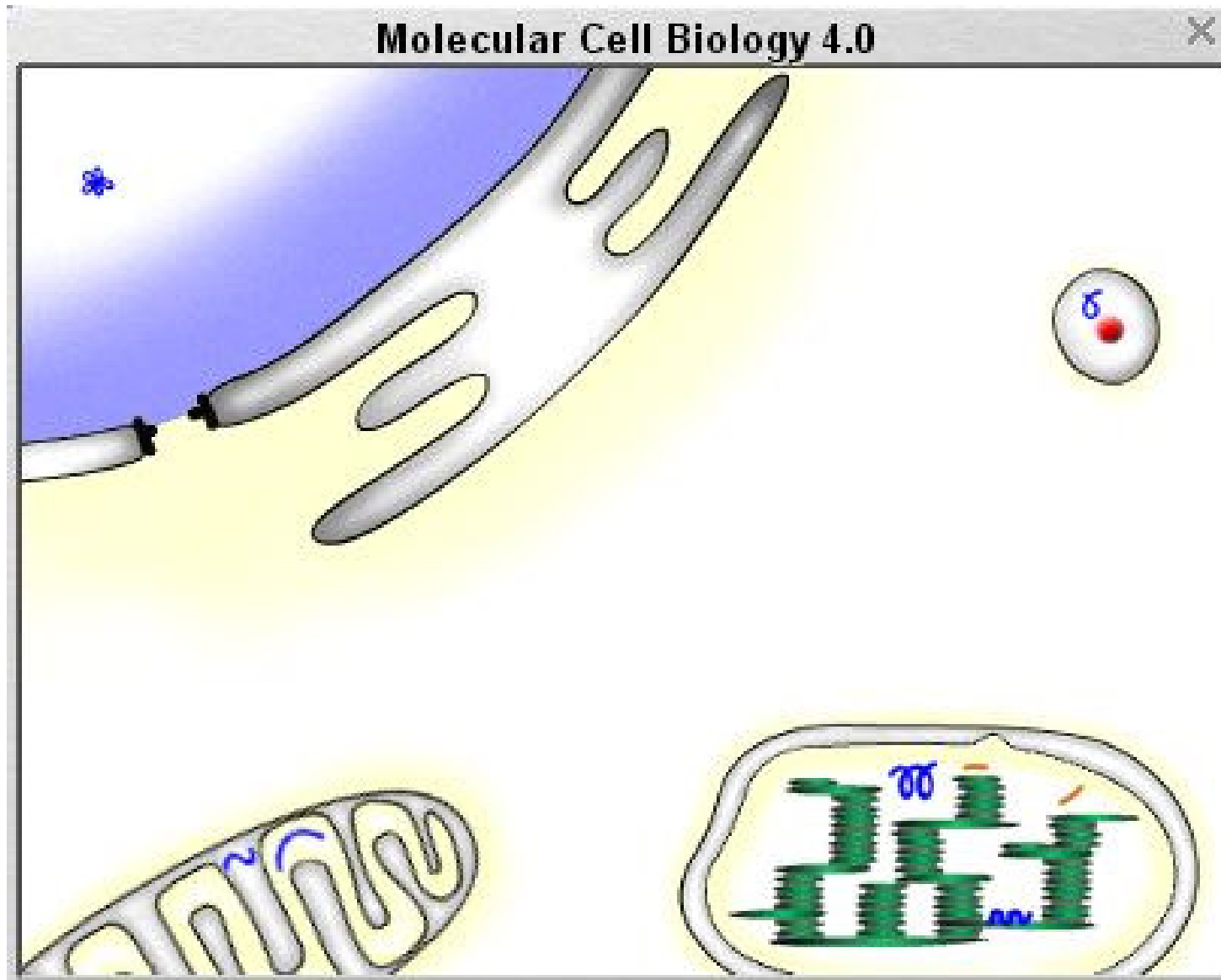
A Biological Algorithm

- LDL-Cholesterol Degradation
 - A cast of many thousands (molecules) just to get one molecule from A to B.
 - Membranes are key to the algorithm, we want to model *them*, not their individual millions of molecules.
- Some very fancy chemistry
 - But its "purpose" is to reliably implement a specific sequence of discrete steps.

Lipid bilayer



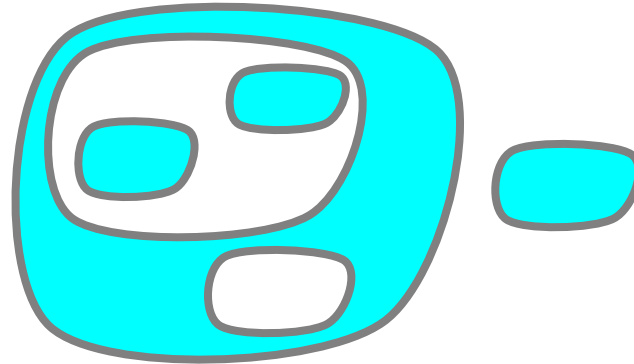
Dynamic Compartments



Brane Calculi

Local Membrane Reactions

Membrane System

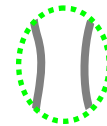


What reactions "make sense"?

Reactions that "make sense" from a local, molecular viewpoint



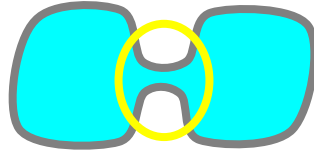
Switch 



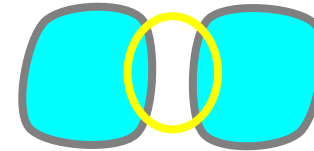
(Symmetric by 90° rotation.)

Global Membrane Reactions

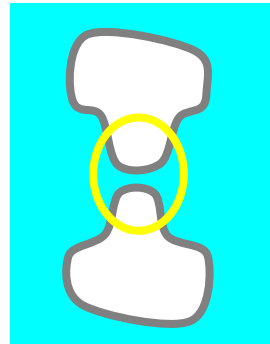
Reactions that "make sense" from a descriptive, global viewpoint



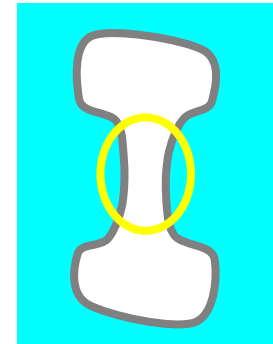
Mito →



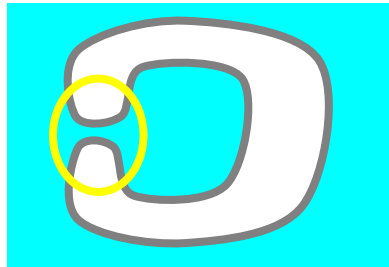
(Fission)



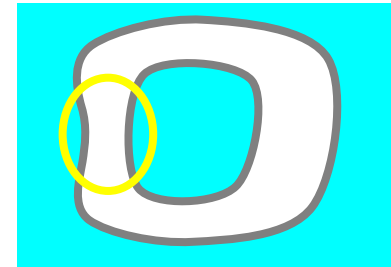
Mate →



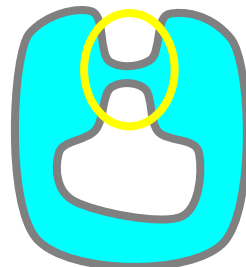
(Fusion)



Endo →



(Fission)



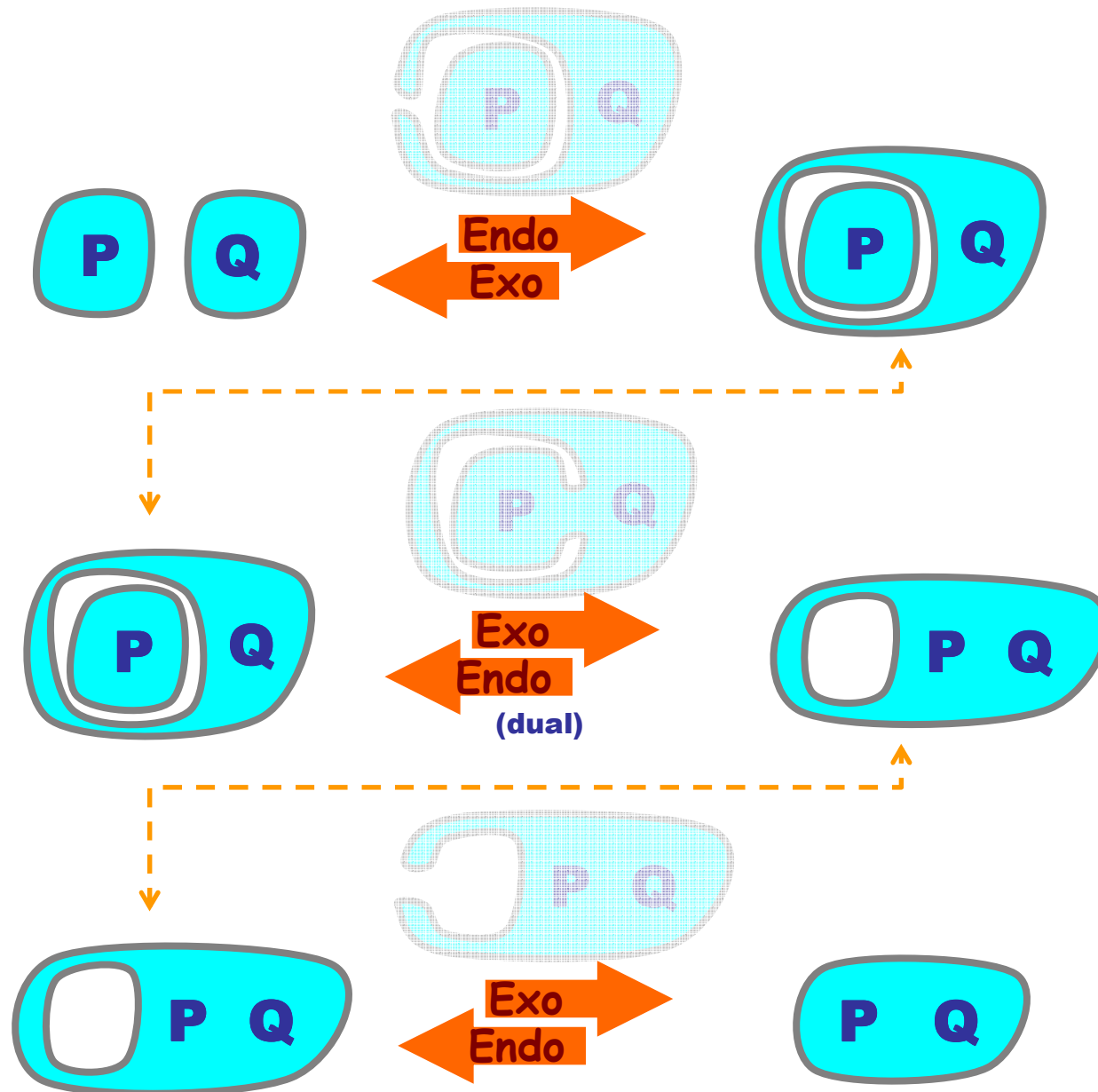
Exo →



(Fusion)

Same Local View!

Mito/Mate by 3 Endo/Exo



What makes Endo happen?

- Membrane transformations are usually "meant":
 - They do not happen spontaneously. They are regulated by membrane-embedded proteins.
 - We need to explain how/when certain membrane reactions happen.
- Formalization
 - Action/coaction interactions in process calculi.
 - Actions "on" the membranes, not "inside" them!
 - Leads to smoother modeling than previous attempts (e.g. BioAmbients).

Brane Calculi

systems $P, Q ::= \diamond \mid P \circ Q \mid !P \mid \sigma(P)$

nests of membranes

branes $\sigma, \tau ::= 0 \mid \sigma \mid \tau \mid !\sigma \mid a.\sigma$

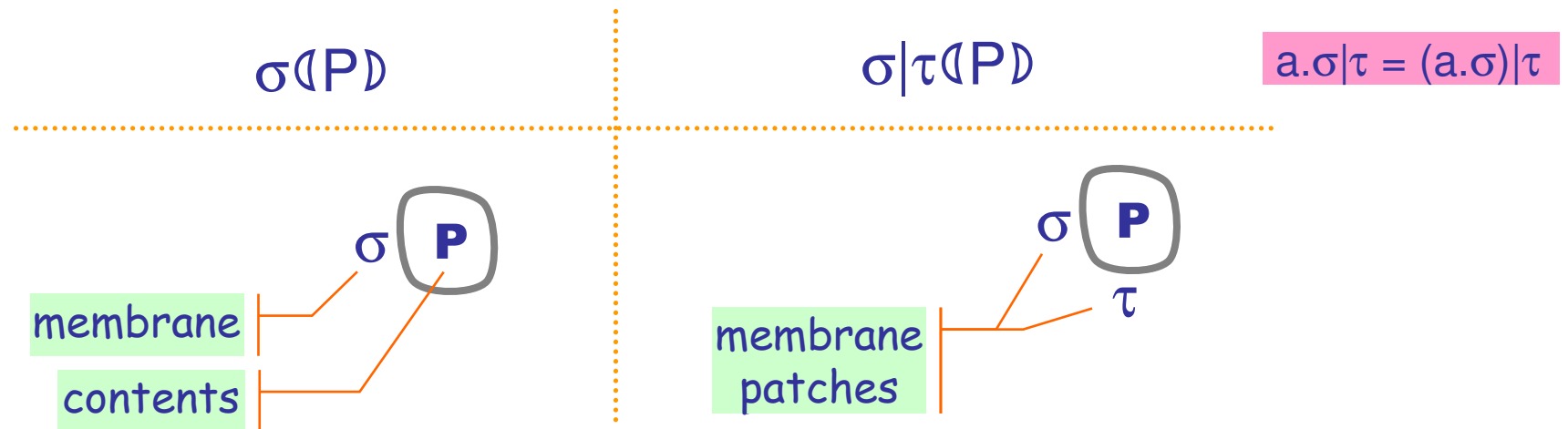
combinations of actions

actions $a ::= 1 \mid \dots$

(fill in as needed)

1D fluids (σ) inside a 2D fluid (P)

TWO commutative monoids instead of ONE of normal process calculi



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

Congruence \equiv and Reaction \rightarrow

	System	Brane
Fluidity	$P \circ Q \equiv Q \circ P$ $P \circ (Q \circ R) \equiv (P \circ Q) \circ R$ $P \circ \diamond \equiv P$	$\sigma \tau \equiv \tau \sigma$ $\sigma (\tau \rho) \equiv (\sigma \tau) \rho$ $\sigma 0 \equiv \sigma$
Plentitude	$!P \equiv P \circ !P$ etc.	$!\sigma \equiv \sigma !\sigma$ etc.
Units	$0(\diamond) \equiv \diamond$ Froth/Fizz	$1.\sigma \equiv \sigma$ Inaction
Congruence	$P \equiv Q \Rightarrow P \circ R \equiv Q \circ R$ $P \equiv Q \Rightarrow !P \equiv !Q$ $P \equiv Q \wedge \sigma \equiv \tau \Rightarrow \sigma(P) \equiv \tau(Q)$	$\sigma \equiv \tau \Rightarrow \sigma \rho \equiv \tau \rho$ $\sigma \equiv \tau \Rightarrow !\sigma \equiv !\tau$ $\sigma \equiv \tau \Rightarrow a.\sigma \equiv a.\tau$

Reaction is up to congruence

$$P \equiv P' \wedge P' \rightarrow Q' \wedge Q' \equiv Q \Rightarrow P \rightarrow Q$$

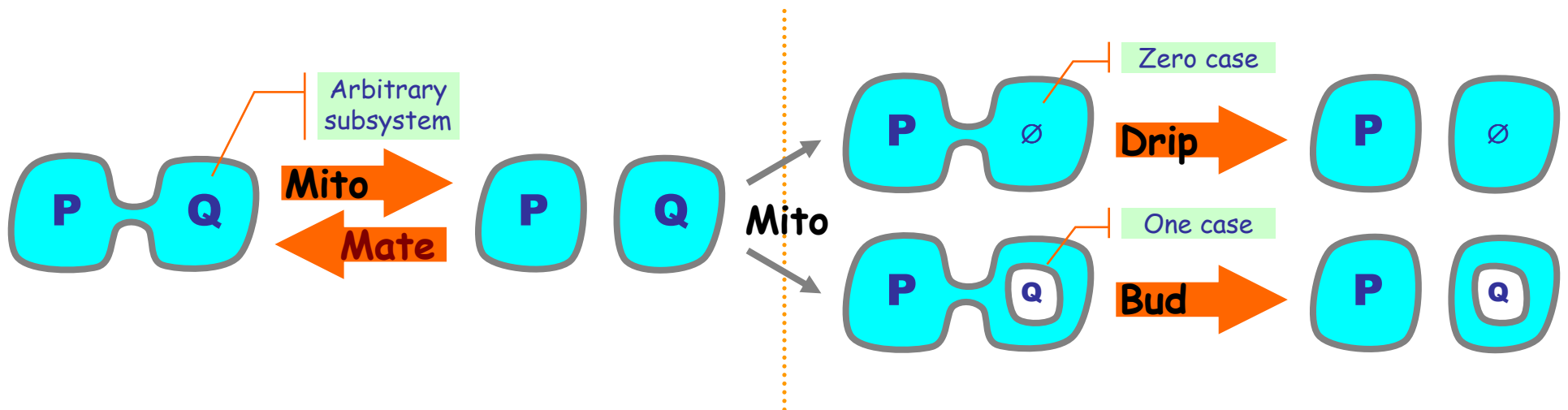
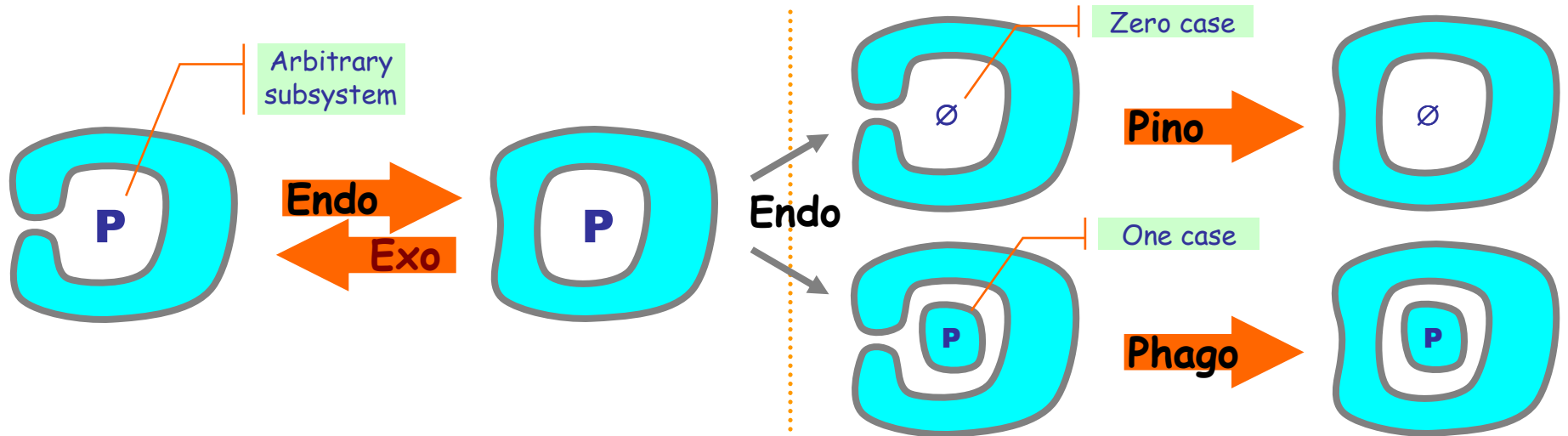
Reactions in solution

$$P \rightarrow Q \Rightarrow P \circ R \rightarrow Q \circ R$$

$$P \rightarrow Q \Rightarrow \sigma(P) \rightarrow \sigma(Q)$$

This is the whole semantics, except for the effects of individual actions.

"Determinization"



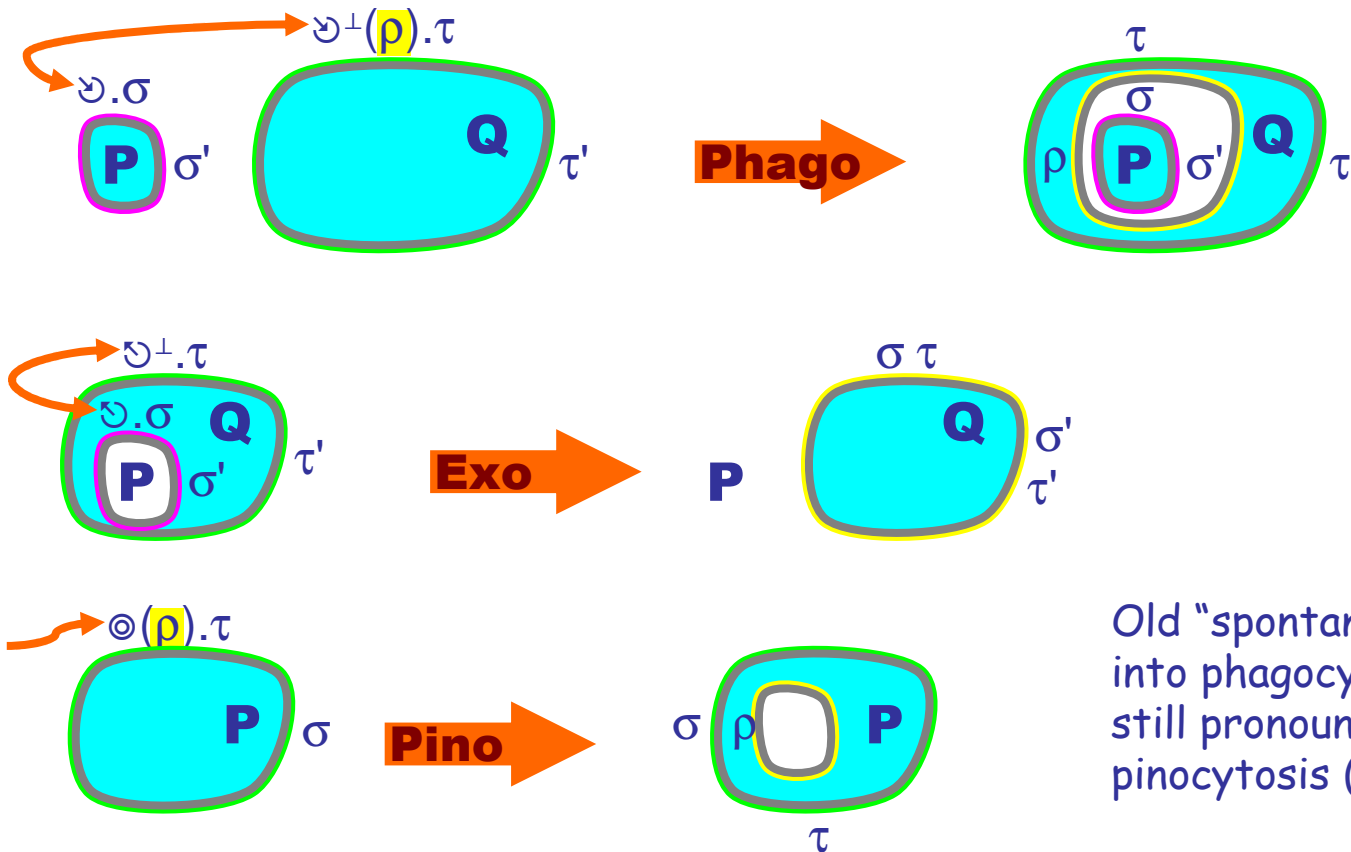
Brane Reactions

actions

$a ::= \dots \mid \vartheta_n \mid \vartheta_n^\perp(\rho) \mid \vartheta_n \mid \vartheta_n^\perp \mid \odot(\rho)$

phago ϑ , exo ϑ^\perp , pino \odot

coordination tags
sometimes omitted



Old "spontaneous" **endo** splits into phagocytosis (**phago**, often still pronounced **endo**) and pinocytosis (**pino**).

...

Phago $\vartheta_n.\sigma|\sigma'(P) \circ \vartheta_n^\perp(\rho).\tau|\tau'(Q) \longrightarrow \tau|\tau'(\rho(\sigma|\sigma'(P)) \circ Q)$

Exo $\vartheta_n^\perp.\tau|\tau'(\vartheta_n.\sigma|\sigma'(P) \circ Q) \longrightarrow P \circ \sigma|\sigma'|\tau|\tau'(Q)$

Pino $\curvearrowright(\rho).\sigma|\sigma'(P) \longrightarrow \sigma|\sigma'(\rho(\diamond) \circ P)$

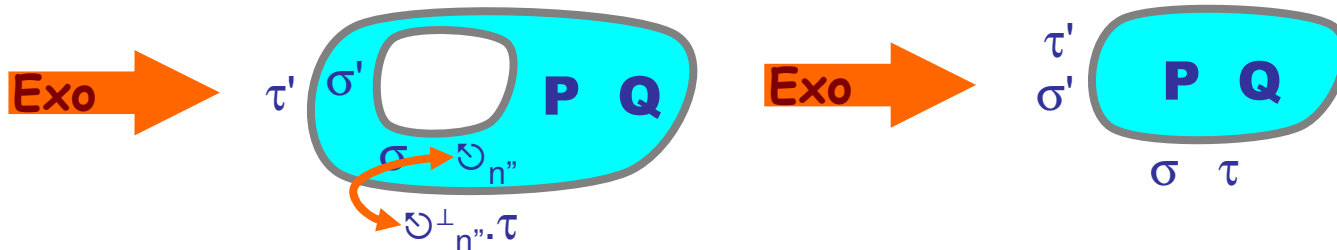
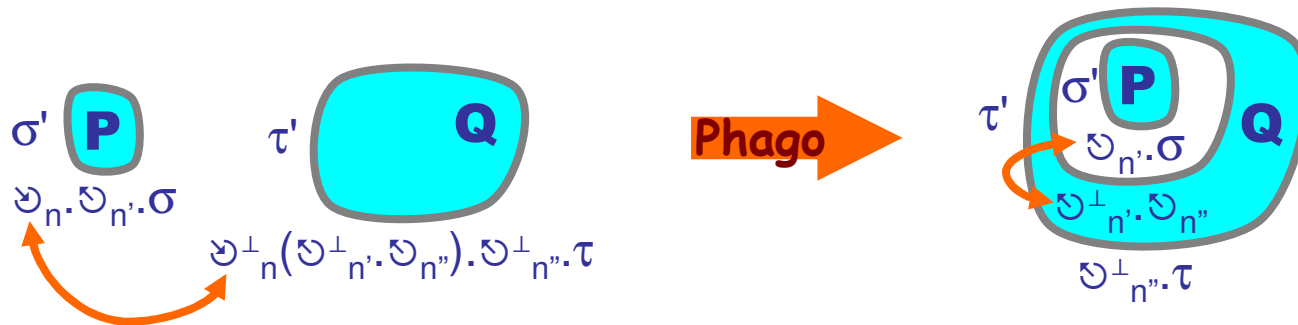
N.B.: the parity of nesting of P and Q is preserved;
this makes the reactions preserve bitonality.

N.B.: in Phago (and Pino), one could perhaps require r to be, conservatively, a piece of t , by a non-linear rewrite:

CPhago $\vartheta_n.\sigma|\sigma'(P) \circ \vartheta_n^\perp(\rho).\tau|\tau'|\rho(Q) \longrightarrow \tau|\tau'(\rho(\sigma|\sigma'(P)) \circ Q)$

Abbreviations: Mate

Mate $\text{mate}_n.\sigma = \vartheta_n.\vartheta_{n'}.\sigma$
 $\text{mate}^\perp_n.\tau = \vartheta^\perp_n(\vartheta^\perp_{n'}.\vartheta_{n''}).\vartheta^\perp_{n''}.\tau$



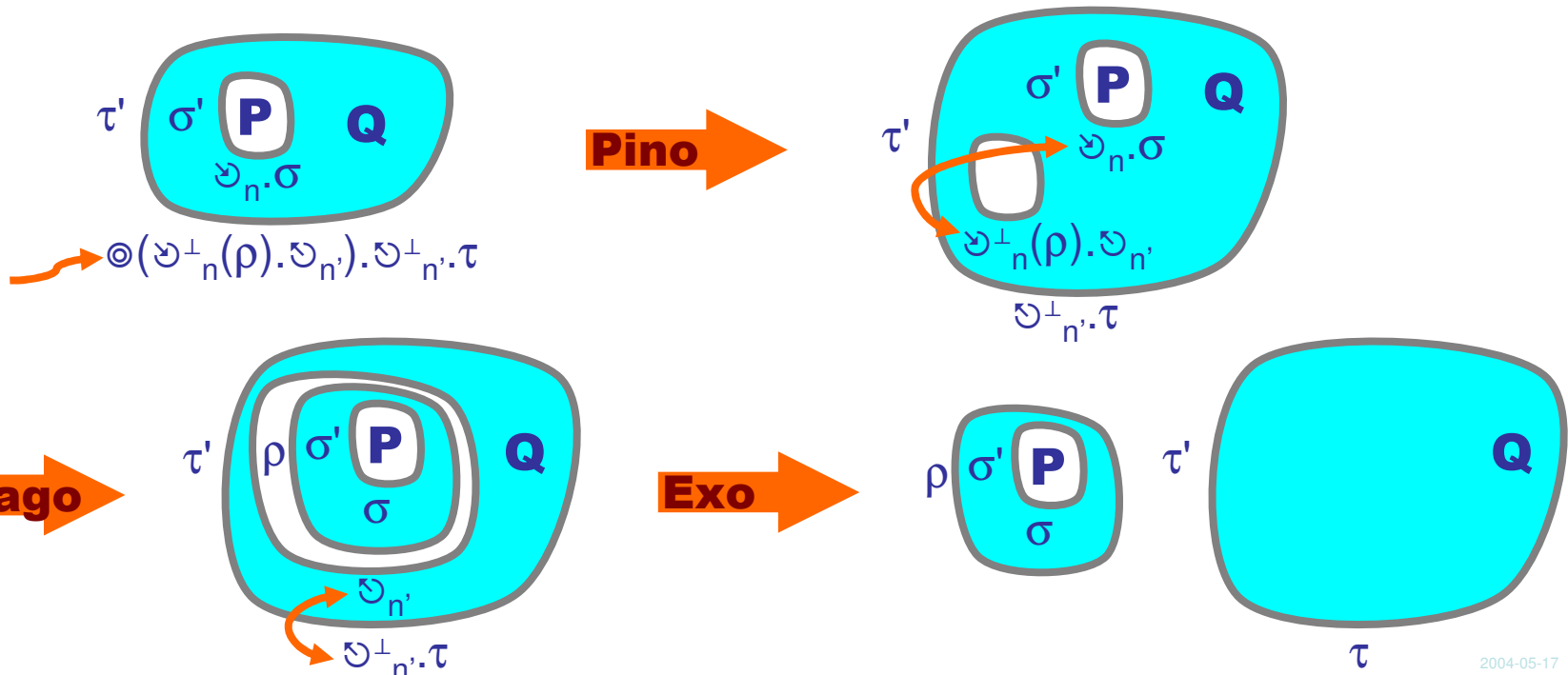
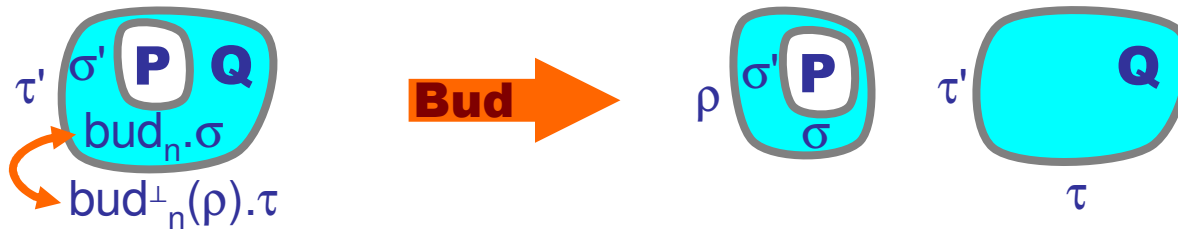
Abbreviations: Bud

Bud

$$\text{bud}_n.\sigma = \vartheta_n.\sigma$$

$$\text{bud}^\perp_n(\rho).\tau = \odot(\vartheta^\perp_n(\rho).\vartheta_{n'}) . \vartheta^\perp_{n'}.\tau$$

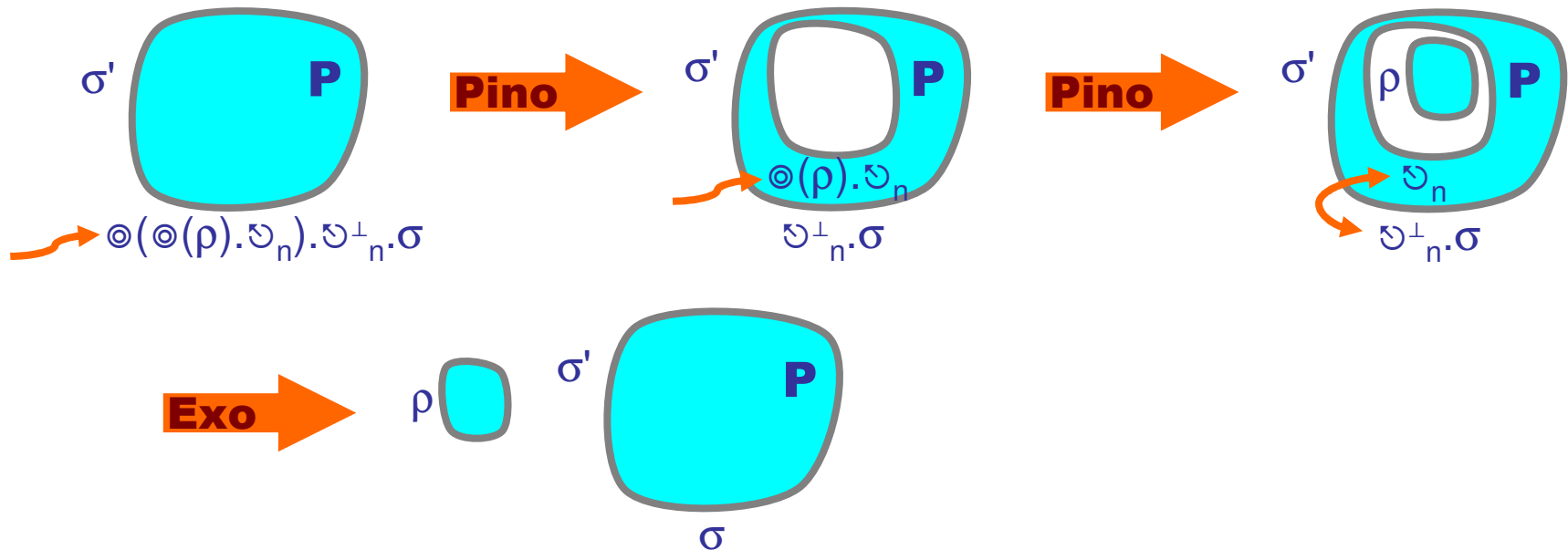
A budding version of old "spontaneous" mito, to avoid arbitrary splits. Follows the pattern of inverse-mate.



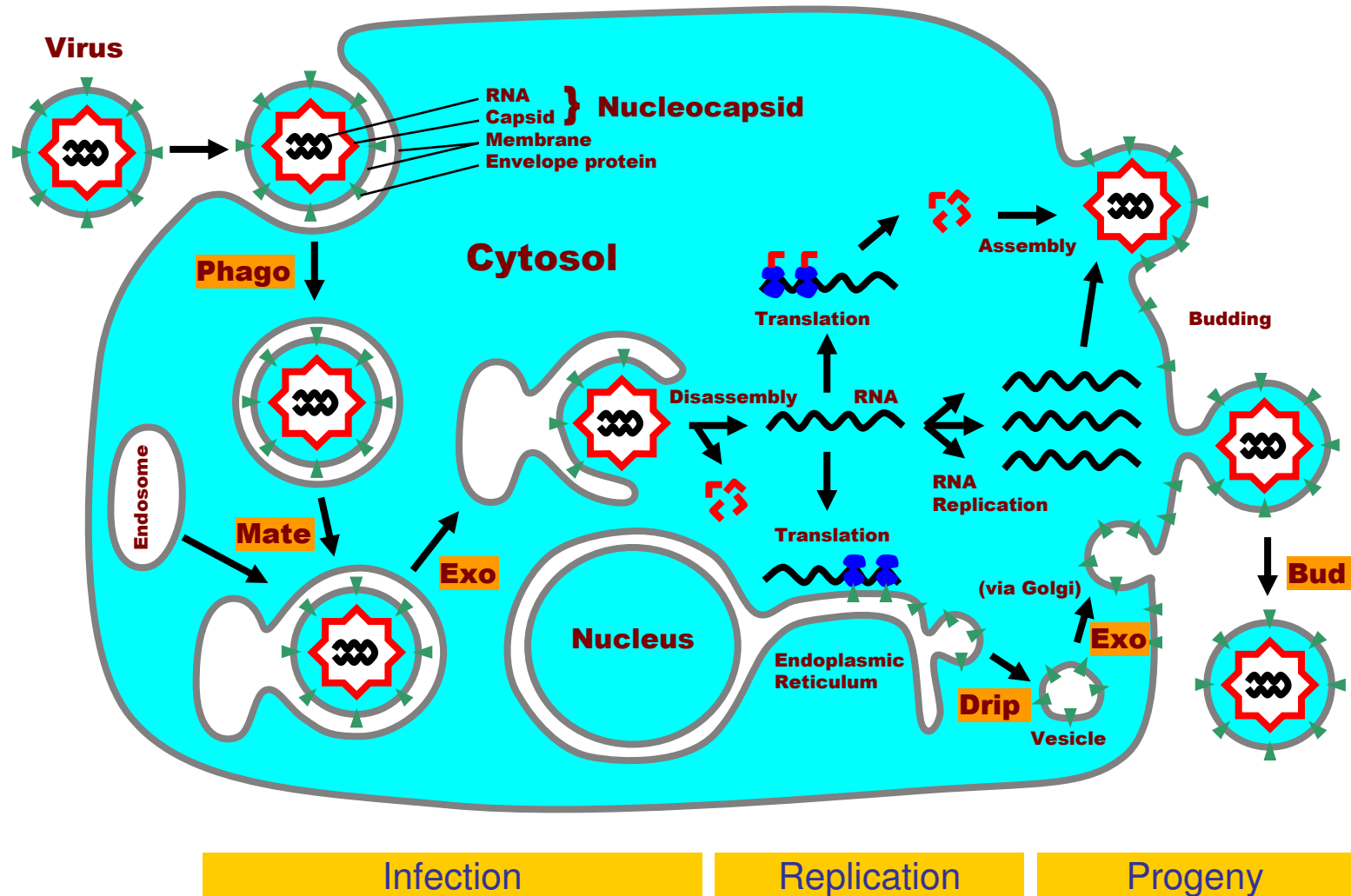
Abbreviations: Drip

Drip $\text{drip}_n(\rho).\sigma = \ominus(\ominus(\rho).\mathfrak{U}_n).\mathfrak{U}_n^\perp.\sigma$

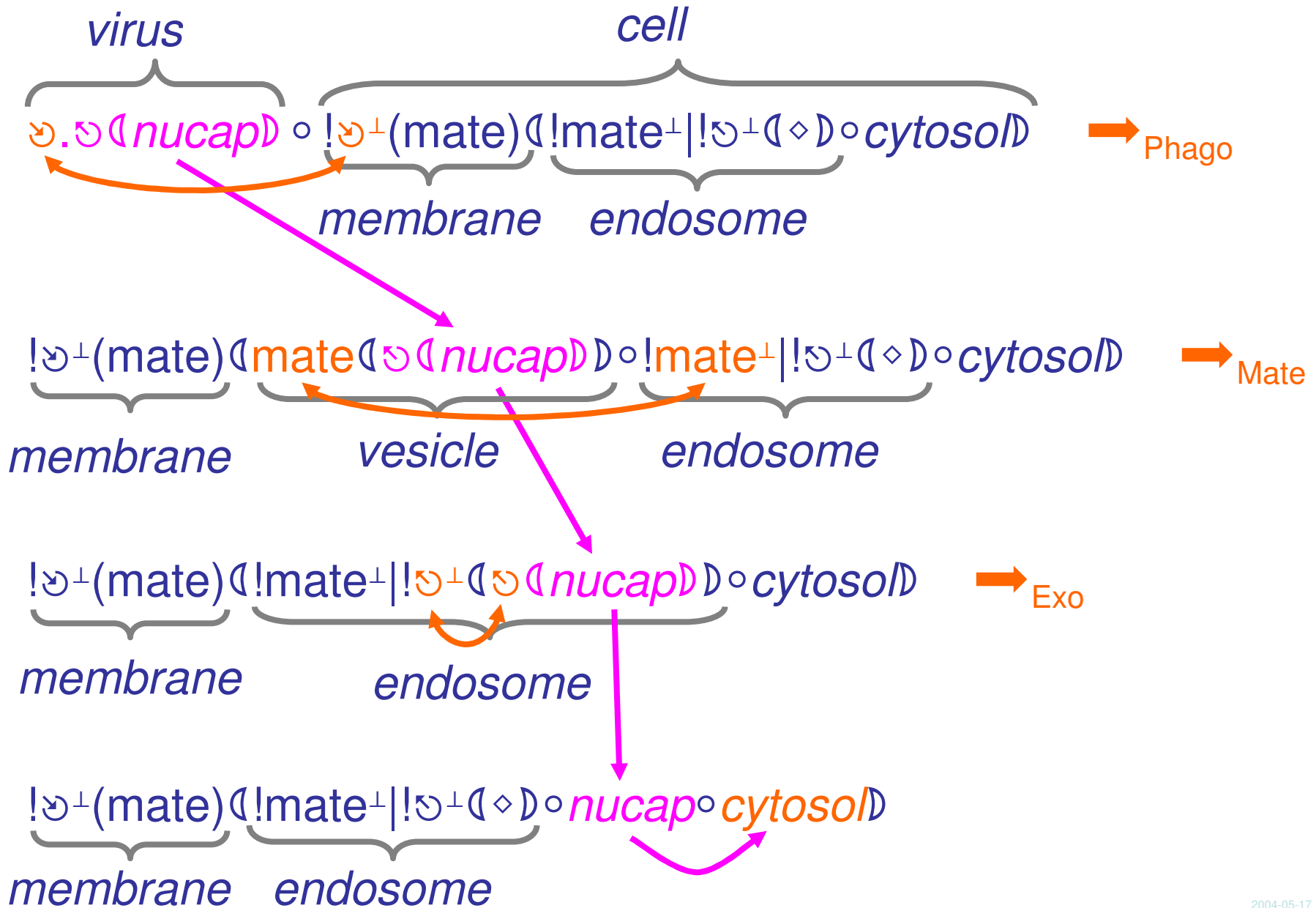
A zero-expelled-membranes version of old "spontaneous" mito, to avoid arbitrary splits. Follows the pattern of inverse-mate.



Ex: Viral Reproduction



Ex: Viral Infection

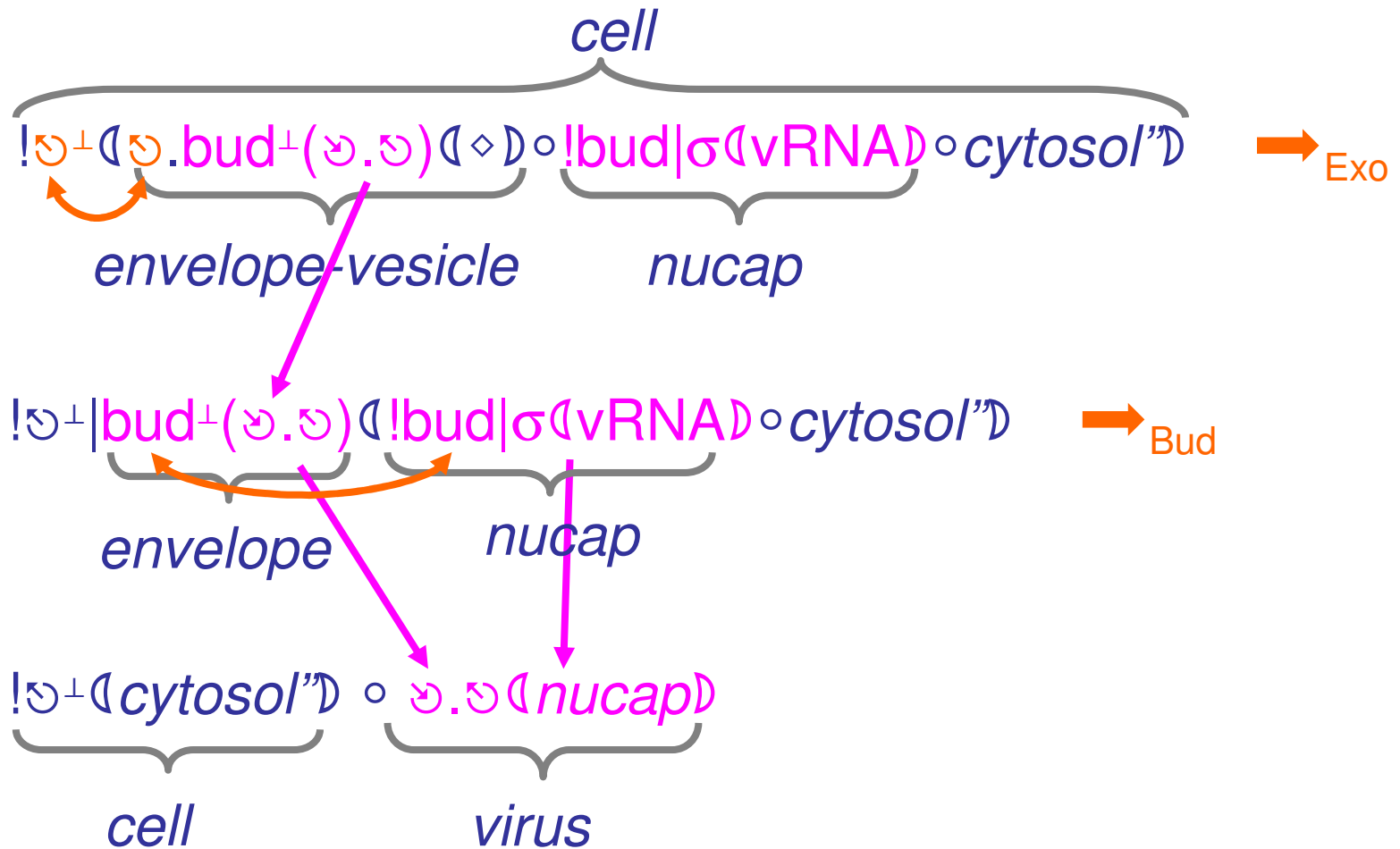


Ex: Viral Progeny

Assume:

$nucap \circ cytosol \rightarrow \rightarrow nucap^n \circ envelope-vesicle^m \circ cytosol'$
by available cellular machinery

Then:



Handling Molecules

systems

$P, Q ::= \dots \mid m$

$m \in M$ molecules

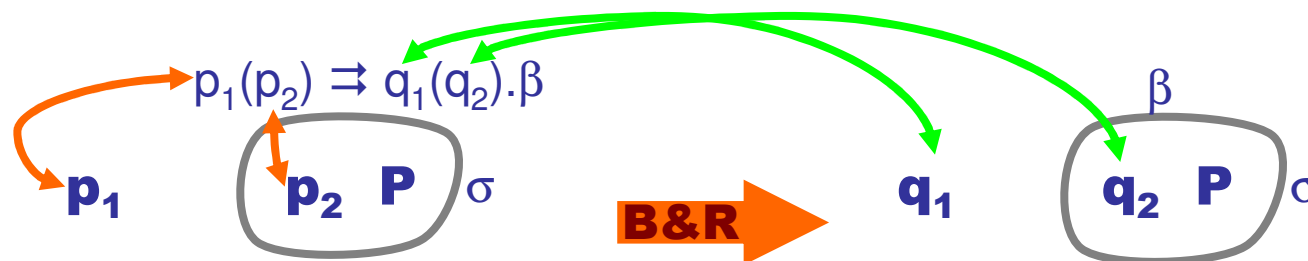
$p, q ::= m_1 \circ \dots \circ m_k$

molecule multisets

actions

$a ::= \dots \mid p_1(p_2) \Rightarrow q_1(q_2)$

bind&release



This single operation can essentially account for the whole Protein Machine, including its interactions with membranes. Except that, one must add some form of protein complexation, either as in BioSPi by adding restriction, or as in κ -calculus by adding complex molecules.

...

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

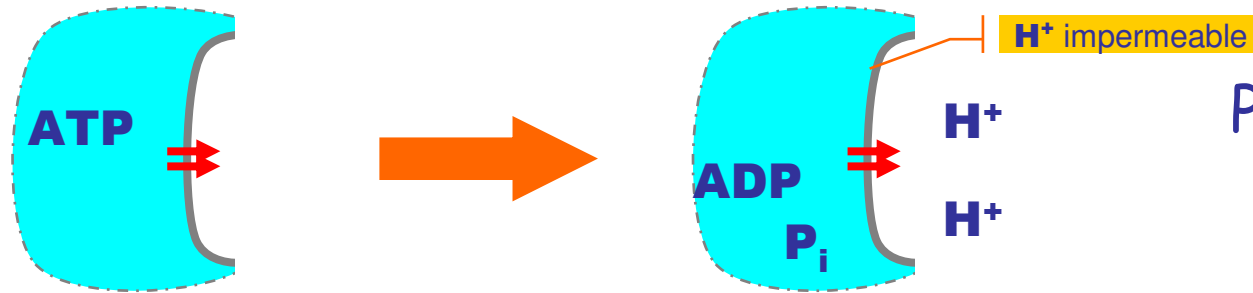
(multiset rewriting, inside and outside membranes)

Simple bindings and releases - “ $\diamond(\diamond)$ ” is omitted:

$m(\diamond) \Rightarrow$	bind out	$\Rightarrow m(\diamond)$	release out
$\diamond(m) \Rightarrow$	bind in	$\Rightarrow \diamond(m)$	release in

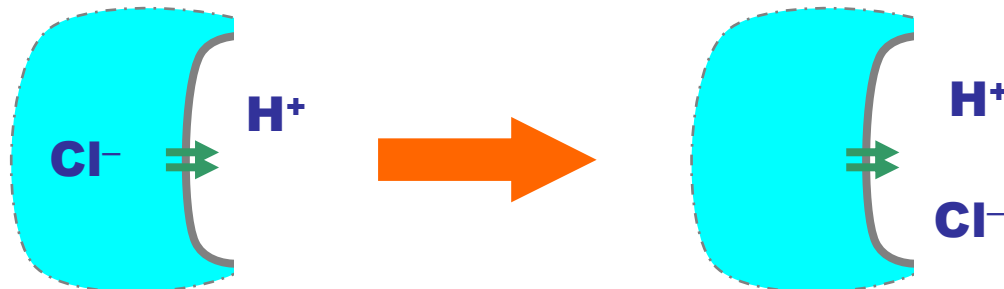
Ex: Molecular Pumps and Channels

E.g. plant vacuole (white).

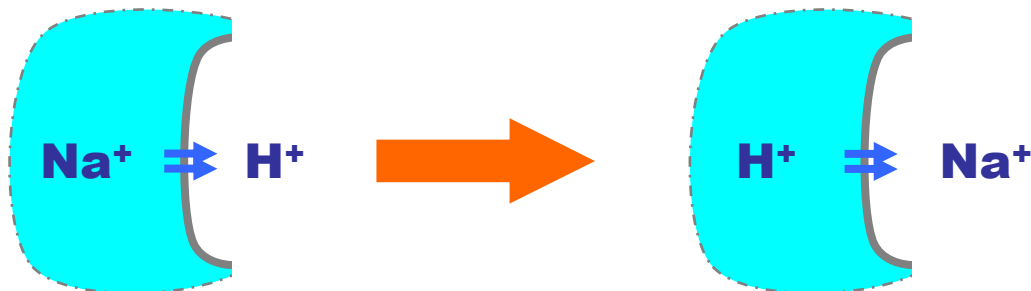


Proton Pump

ATP charges up the vacuole with H^+ . Several other pumps work off that charge.



Ion Channel



Proton Antiporter

A plant vacuole membrane has all those things on it.

...

ProtonPump = ! ATP(\diamond) \rightleftharpoons ADP \circ P_i(H⁺ \circ H⁺)

IonChannel = ! Cl⁻(H⁺) \rightleftharpoons \diamond (H⁺ \circ Cl⁻)

ProtonAntiporter = ! Na⁺(H⁺) \rightleftharpoons H⁺(Na⁺)

PlantVacuole =

ProtonPump | IonChannel | ProtonAntiporter (\diamond)

Hence this reaction notation, \rightleftharpoons , is "like" chemical reaction notation, \rightarrow , but talking about both sides on a membrane at once.

(N.B. no built-in conservation of mass in either case.)

Special Cases of B&R

Chemical reaction catalysis (inside a compartment)

$$p \longrightarrow q \triangleq ! p(\diamond) \Rightarrow q(\diamond)(\mathcal{D})$$

$$p \rightleftharpoons q \triangleq p \longrightarrow q \circ q \longrightarrow p$$

E.g. peptide bond between two aminoacids $R^1 R^2$:
 $R^1\text{-COOH} \circ H_2N\text{-}R^2 \longrightarrow R^1\text{-CO-HN-}R^2 \circ H_2O$

Compartment conditions (on the membrane of a compartment)

$$p \rightarrow q \triangleq ! \diamond(p) \Rightarrow \diamond(q)$$

$$p \rightarrow q | \sigma(\mathcal{P})$$

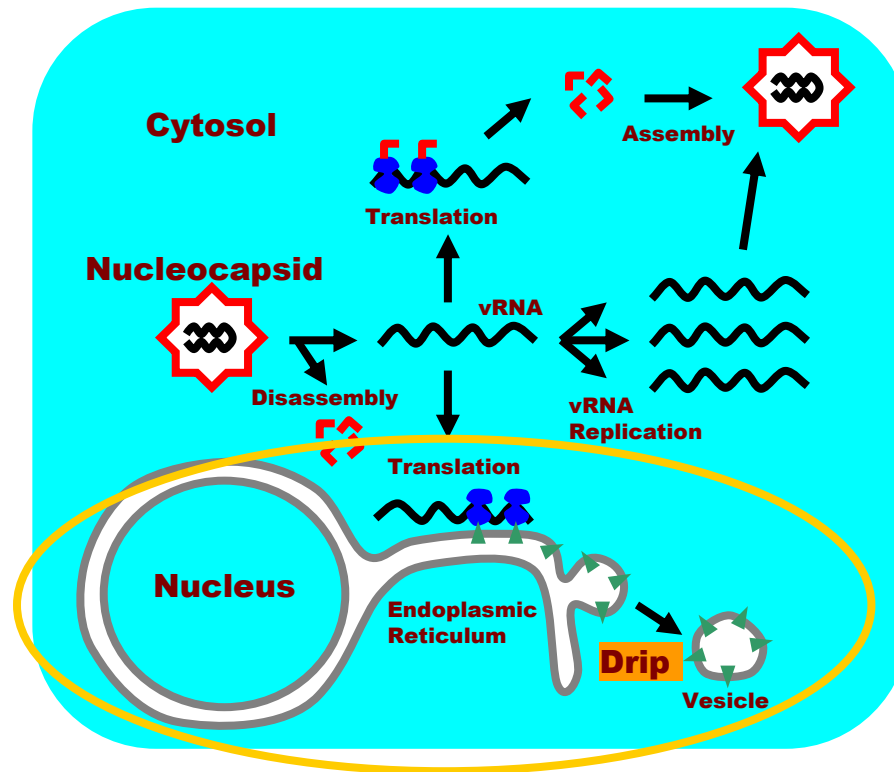
Condition affecting P

E.g. a condition-driven reaction:

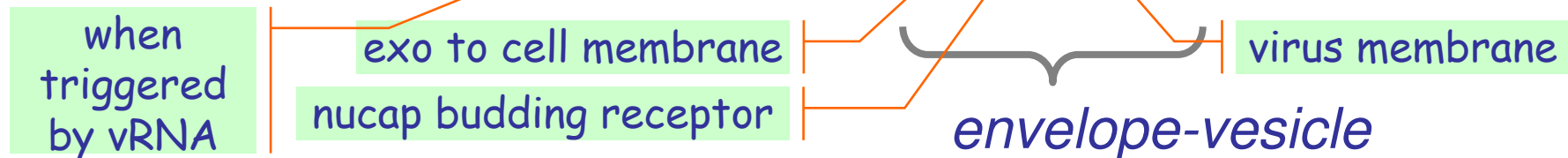
$$p \rightarrow q | \sigma(p) \longrightarrow p \rightarrow q | \sigma(q)$$

Ex: Virus Replication

$nucap \circ cytosol \rightarrow \rightarrow nucap^n \circ envelope-vesicle^m \circ cytosol'$



$ER \triangleq !vRNA(\diamond) \Rightarrow vRNA(\diamond). \text{drip}(\ominus.bud^+(\ominus.\ominus)) \langle Nucleus \rangle$



(See paper for the other two vRNA pathways)

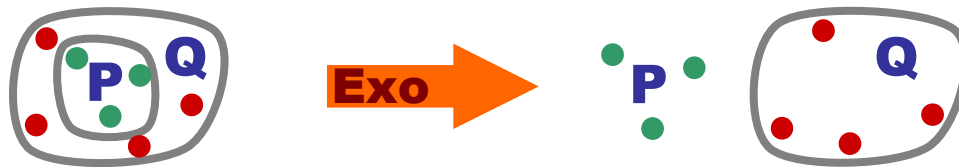
Summary of Instruction Set So Far

- Phago-exo-pino for the Membrane Machine
 - Plus mate-bud-drip, in principle definable.
- Bind&Release for the Protein Machine
 - Still could add κ -calculus complexation
 - Helps remove another need for π -restriction, which makes almost any analysis easier.
 - Helps avoid unrealistic uses of membranes for complexation.
- What about the Gene Machine?
 - Need some extra detailed mechanism for the regulatory regions?
 - Or conversely some simplifying abstraction for gene cascades?

Why do we need Brane Calculi, again?



Original "on brane"
Exo of Brane Calculus



"In brane" encoding
(e.g. in BioAmbients
or SMBL) goes wrong



"Ball bearing"
encoding; best we can
do "in brane"

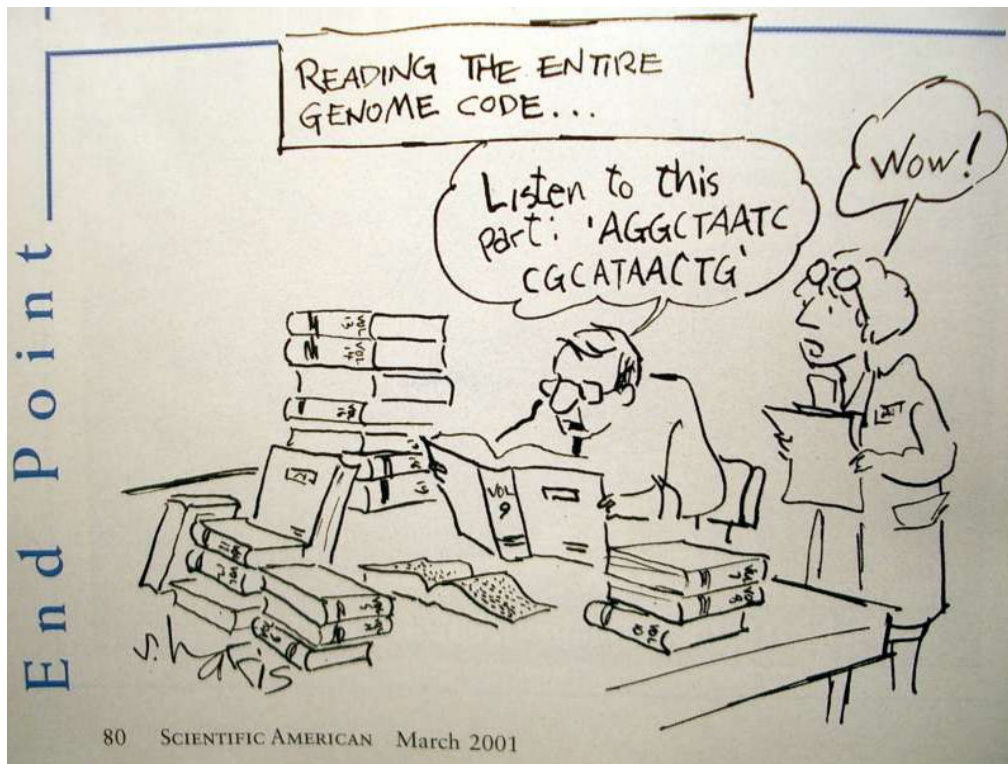
Awkward encoding. And all kinds of things
can go wrong in the intermediate state.

- One cannot easily represent the Exo reaction in BioAmbients or any such compartment-based calculus, nor can one easily add it as a new primitive!
- But we can add BioAmbients-like In/Out out to Brane Calculi if we want to.

Adding Frills to the Framework

- So far, purely combinatorial:
 - No name binding, channel creation, communication...
 - Closer to combinatorial flavor of protein interactions
 - Goes a long way.
- But one can easily add all that, and more:
 - CCS-style communication
 - Diffusion of molecules on cellular membrane
 - BioAmbients-style communication
 - Diffusion of molecules across cellular membrane
 - BioAmbients-like mobility
 - Non-bitonal
 - π -style restriction
- We have a framework where we can plug&play a rich set of interactions, while supporting compartments.

Conclusions



"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)

"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.

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