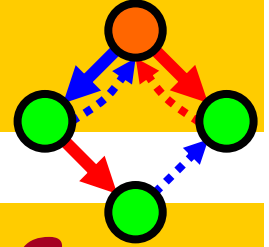


The problem of biology is not to stand aghast  
at the complexity but to conquer it. Sydney Brenner.

Artificial  
Biochemistry



# Abstract Machines of Systems Biology

Luca Cardelli

Microsoft Research

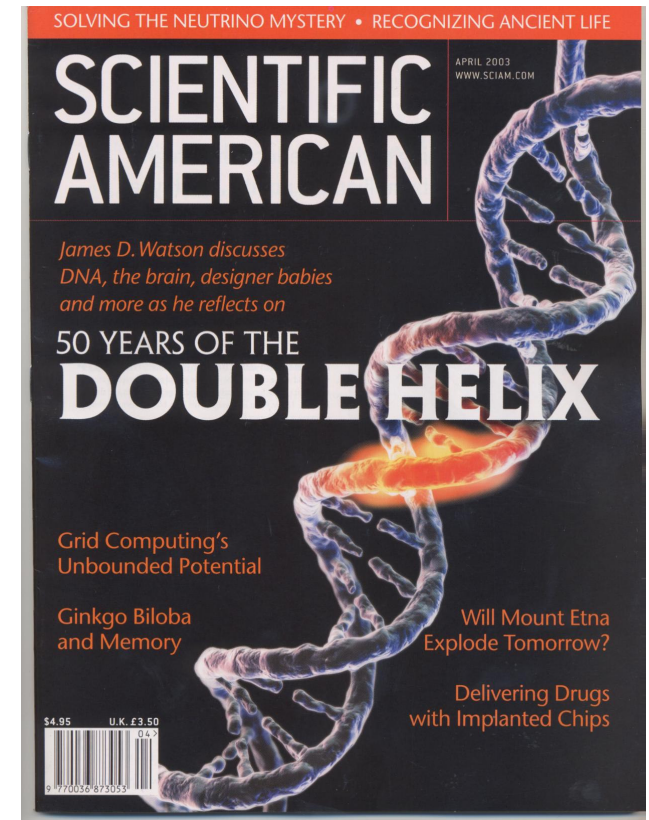
The Microsoft Research - University of Trento  
Centre for Computational and Systems Biology

Trento, 2006-05-22..26

[www.luca.demon.co.uk/ArtificialBiochemistry.htm](http://www.luca.demon.co.uk/ArtificialBiochemistry.htm)

# 50 Years of Molecular Cell Biology

- **Genes are made of DNA**
  - Store digital information as sequences of 4 different nucleotides
  - Direct protein assembly through RNA and the Genetic Code
- **Proteins (>10000) are made of amino acids**
  - Process signals
  - Activate genes
  - Move materials
  - Catalyze reactions to produce substances
  - Control energy production and consumption
- **Bootstrapping still a mystery**
  - DNA, RNA, proteins, membranes are today interdependent. Not clear who came first
  - Separation of tasks happened a long time ago
  - Not understood, not essential



# Towards Systems Biology

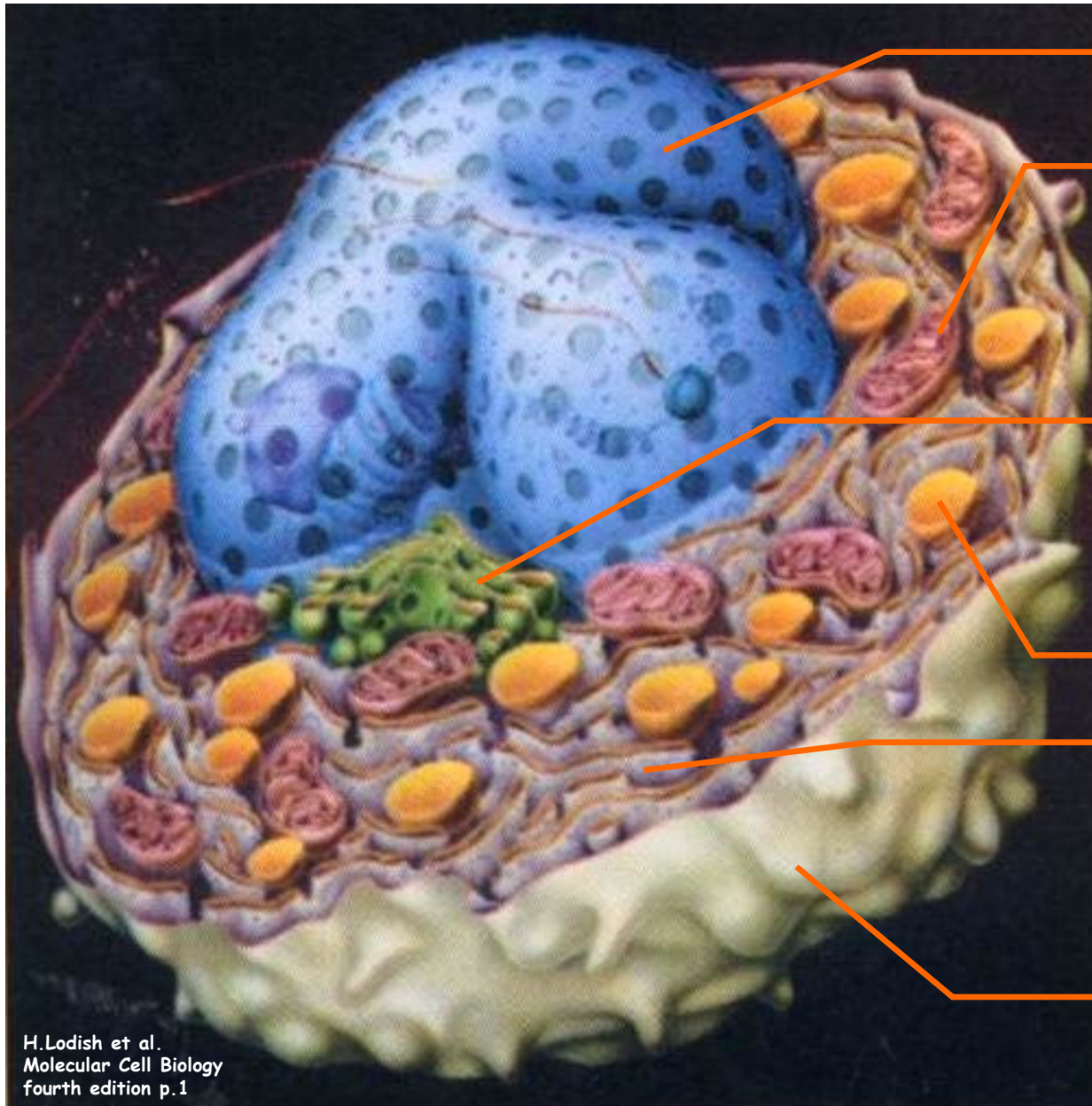
- Biologists now understand many of the cellular components
  - A whole team of biologists will typically study a single protein for years
  - **Reductionism: understand the components in order to understand the system**
- But this has not led to understand how "the system" works
  - Behavior comes from **complex patterns of interactions between components**
  - Predictive biology and pharmacology still rare
  - Synthetic biology still unreliable
- **New approach: try to understand "the system"**
  - Experimentally: massive data gathering and data mining (e.g. Genome projects)
  - Conceptually: modeling and analyzing networks (i.e. interactions) of components
- **What kind of a system?**
  - Just beyond the basic chemistry of energy and materials processing...
  - Built right out of digital information (DNA)
  - Based on information processing for both survival and evolution
  - *Highly* concurrent
- **Can we fix it when it breaks?**
  - Really becomes: How is information structured and processed?

# Structural Architecture

## Eukaryotic Cell

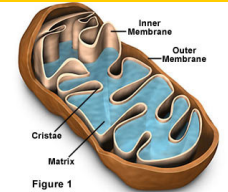
(10~100 trillion in human body)

Membranes everywhere

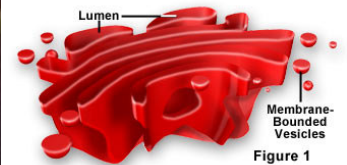


Nuclear membrane

Mitochondria

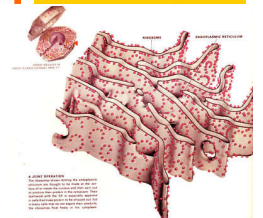


Golgi



Vesicles

E.R.



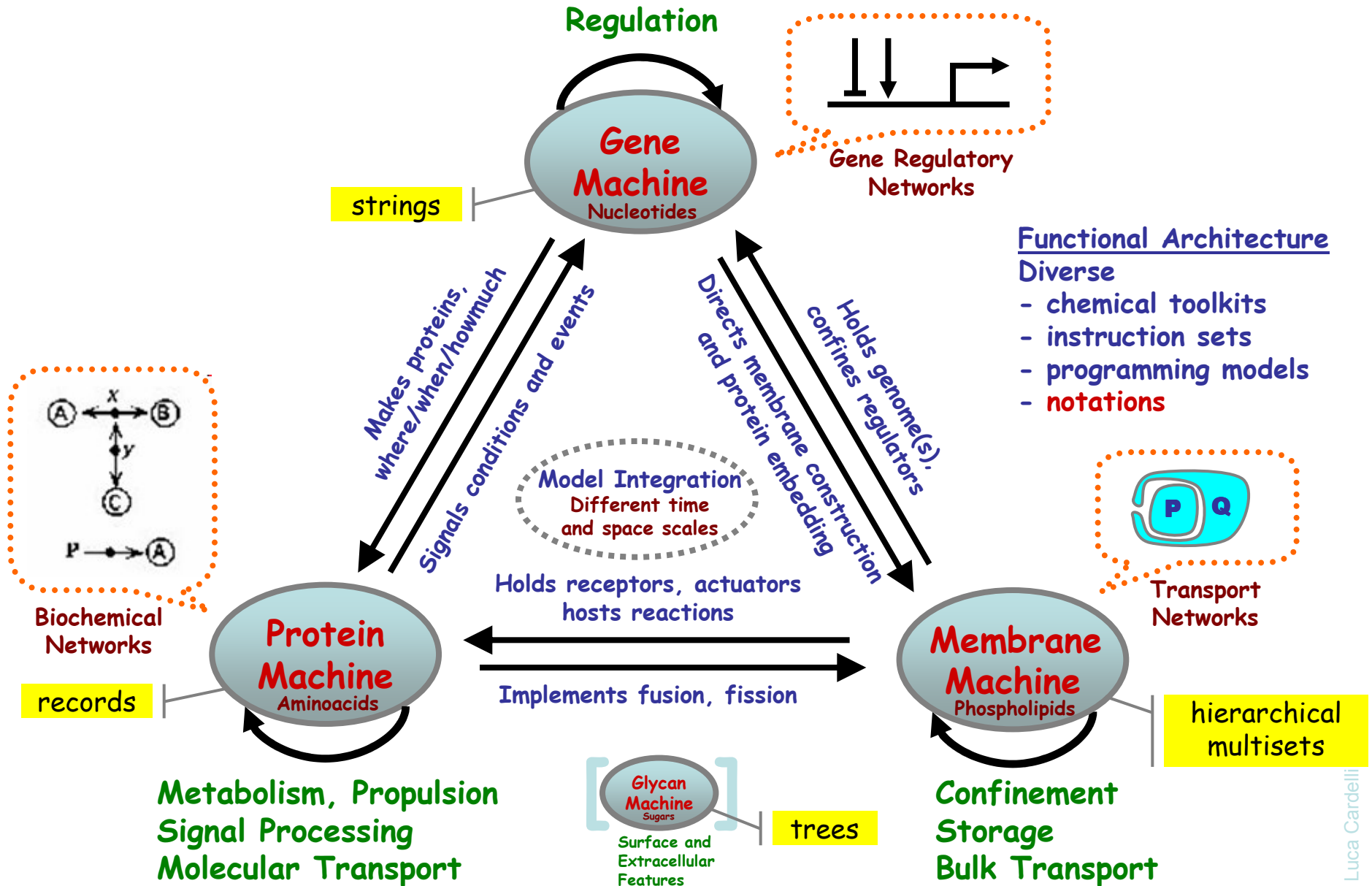
Plasma membrane (<10% of all membranes)



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



# Abstract Machines of Systems Biology

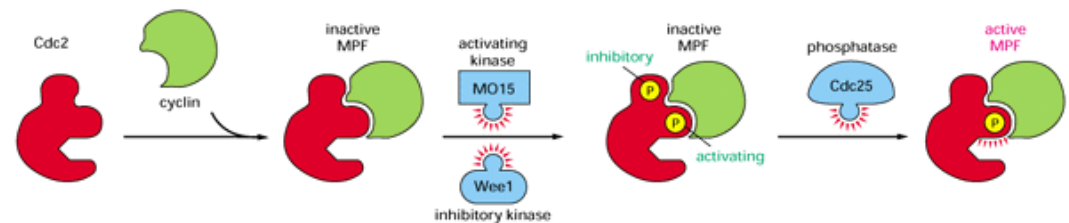


# The Protein Machine

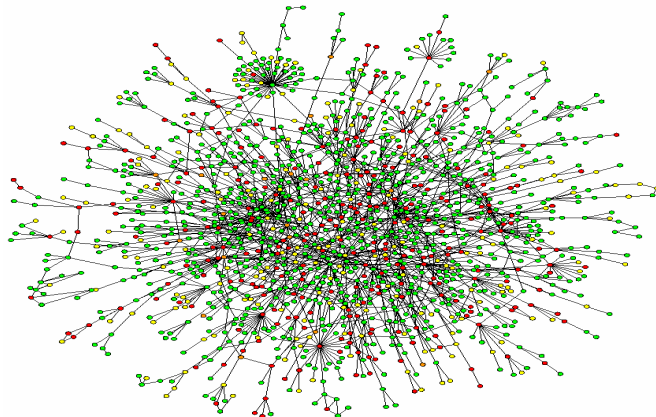
# 1. The Protein Machine

*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.
  - ... needed to summarize and abstract.

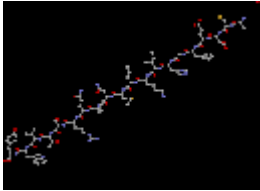


**An actual molecular interaction network.**

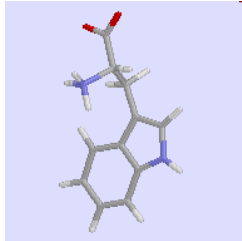
(Nodes are distinct protein kinds, arcs mean that two kinds of proteins interact.)

# Protein Structure

Primary

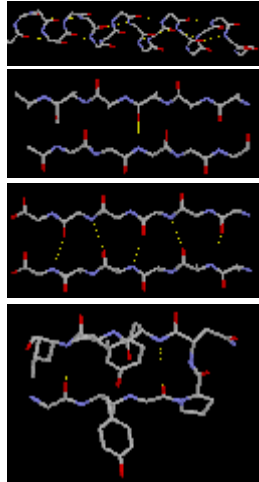


The 20 Aminoacids



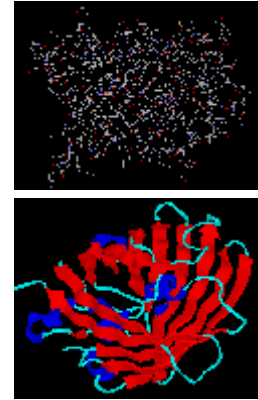
Tryptophan

Secondary



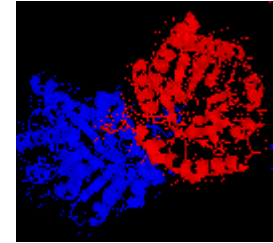
Alpha Helix, Beta Sheet

Tertiary



Green Fluorescent Protein

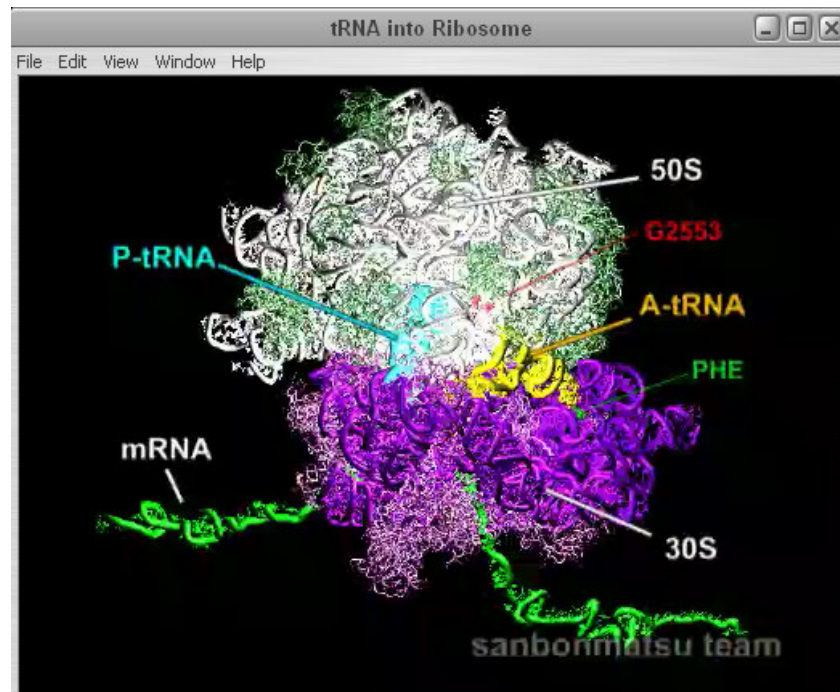
Quaternary



Triose Phosphate Isomerase



# Ribosome: mRNA to Protein



# Protein Function

Regulation

Degradation

Metabolism

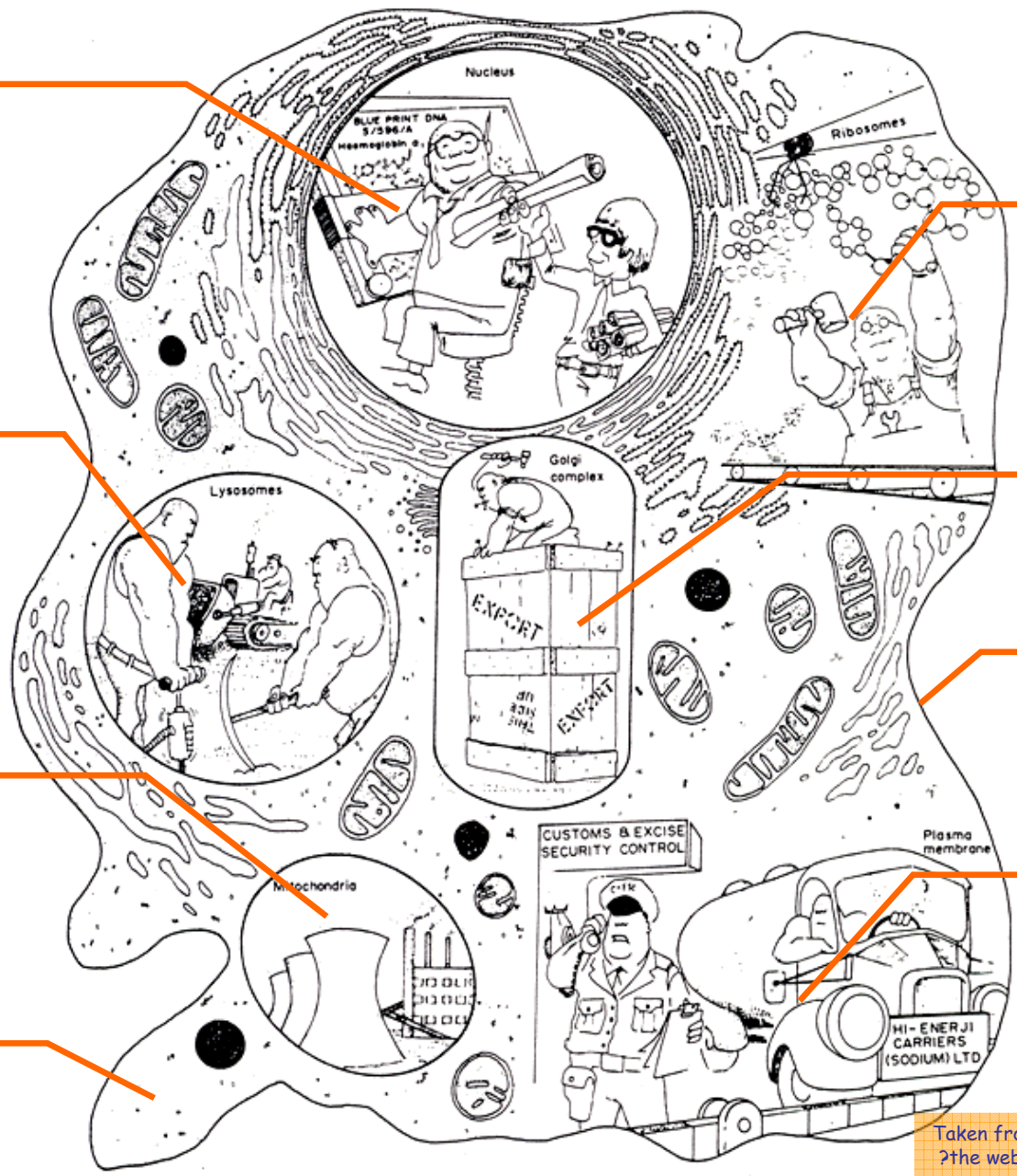
Movement

Assembly

Transport

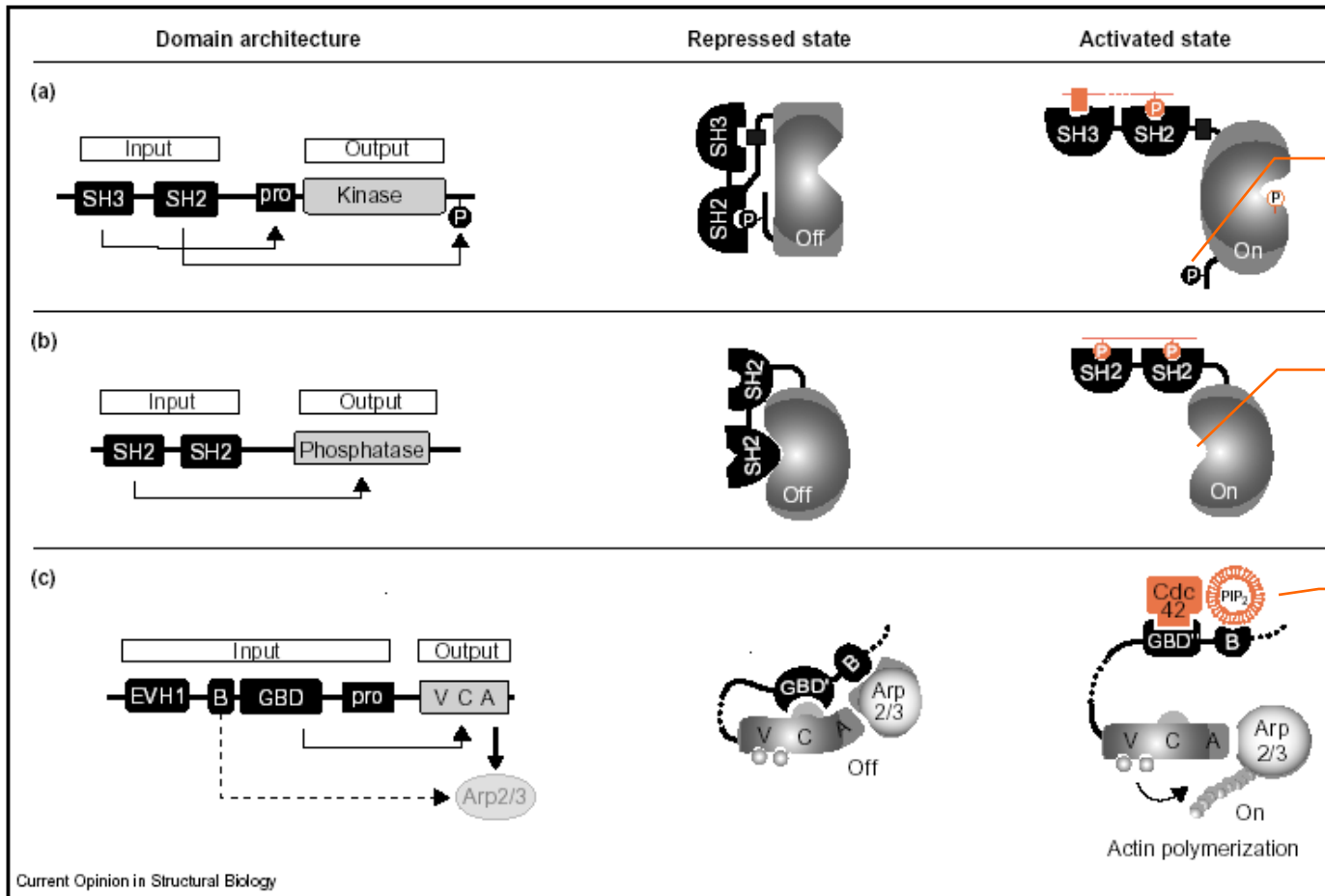
Structure

Signaling



Taken from  
?the web?

# 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 PIP<sub>2</sub> 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.

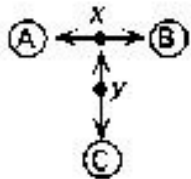
# MIM: Molecular Interaction Maps (Kohn)



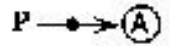
The double-headed line indicates that proteins **A** and **B** can bind to each other. The "node" placed on the line represents the **A:B** complex.



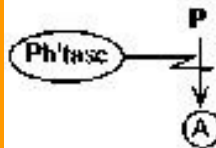
Asymmetric binding where protein **A** donates a peptide that binds to a receptor site or pocket on protein **B**.



Representation of multimeric 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.



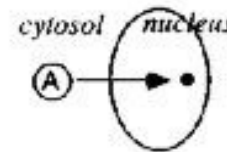
Cleavage of a covalent bond: dephosphorylation of **A** by a phosphatase.



Proteolytic cleavage at a specific site within a protein.



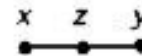
Stoichiometric conversion of **A** into **B**.



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



# Kohn Diagrams

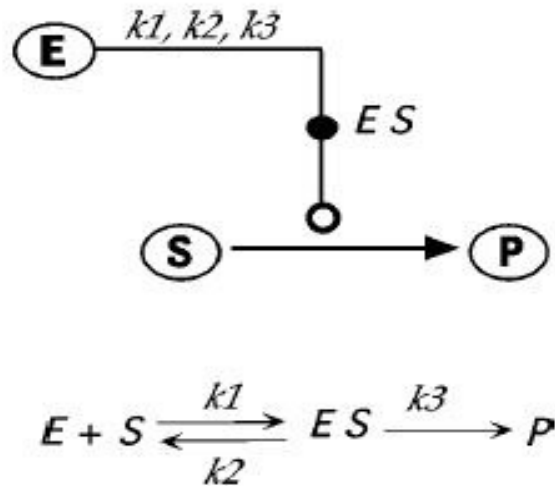


FIG. 3. Simple one-way enzymatic reaction. (If there is an energy source, such as ATP hydrolysis, it can be omitted when ATP concentration is not an important factor.) In explicit formulations, the reaction identifiers or rate constant designations can be placed on the enzyme reaction line, and the node ES can identify the enzyme-substrate species.

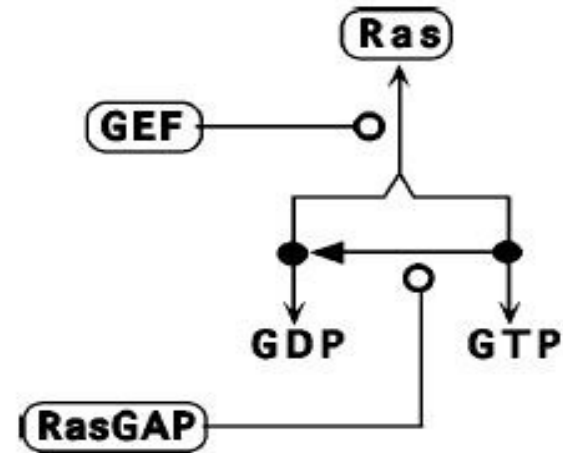


FIG. 4. Interconversions between the GTP- and GDP-bound states of Ras. (1) GDP and GTP compete with each other for binding to a site on Ras (this binding is only slowly reversible). (2) GEF (guanine nucleotide exchange factor) facilitates the binding or dissociation of GDP or GTP (the concentration of GTP normally far exceeds that of GDP). (Implicit is the reversible binding between GEF and Ras which opens the binding site for GDP/GTP exchange.) (3) Ras has an intrinsic GTPase activity that slowly converts bound GTP to bound GDP (stoichiometric conversion arrow points from the node representing Ras.GTP to the node representing Ras.GDP). (4) RasGAP (a GTPase activating protein) enhances the GTPase activity of Ras. (Implicit is the reversible enzyme-substrate binding between RasGAP and Ras.)

# Molecular Interaction Maps (Kohn)

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

JDesigner

## The p53-Mdm2 and DNA Repair Regulatory Network

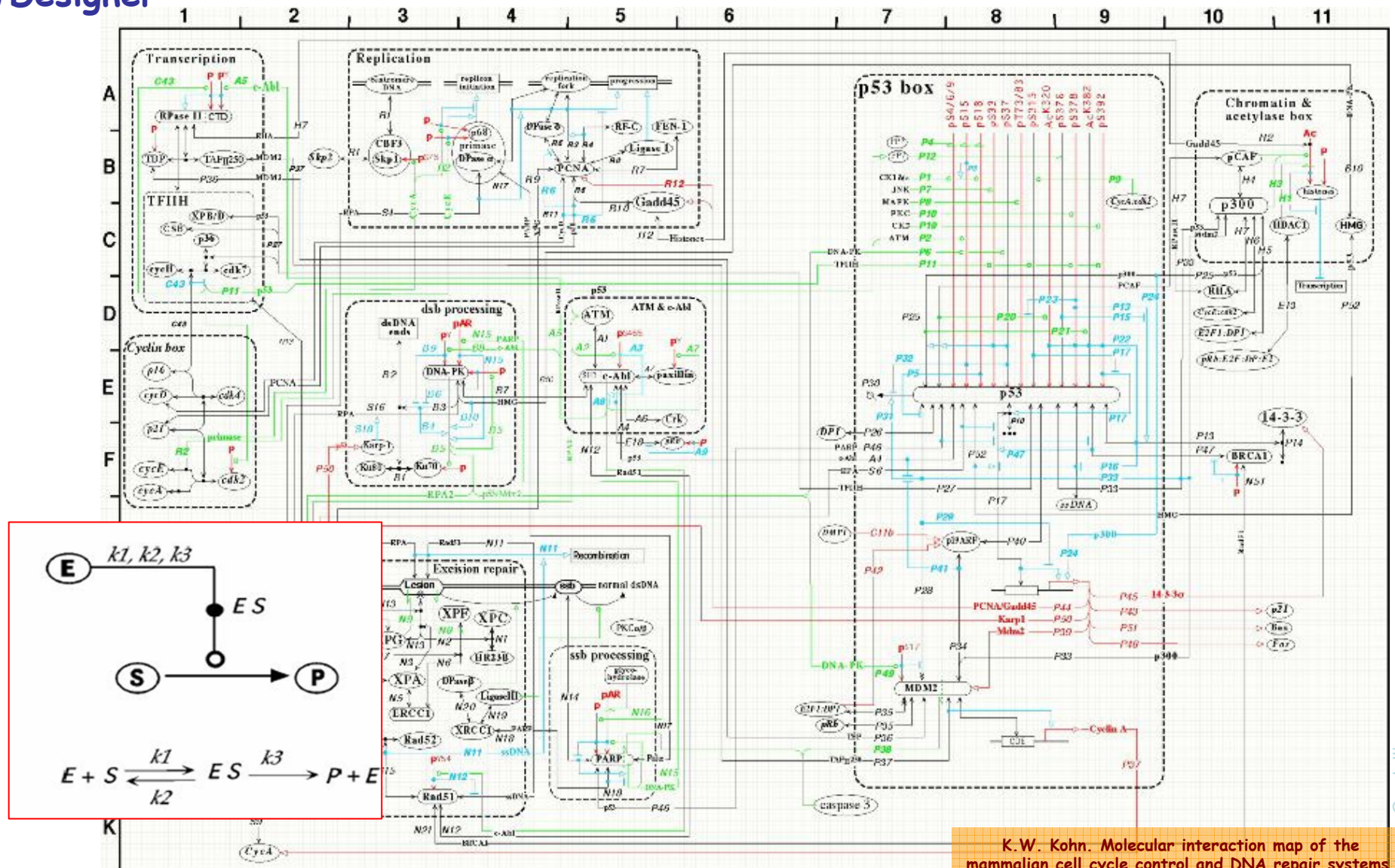


Figure 6B: The p53-Mdm2 and DNA repair regulatory network (version 2p - May 19, 1999)

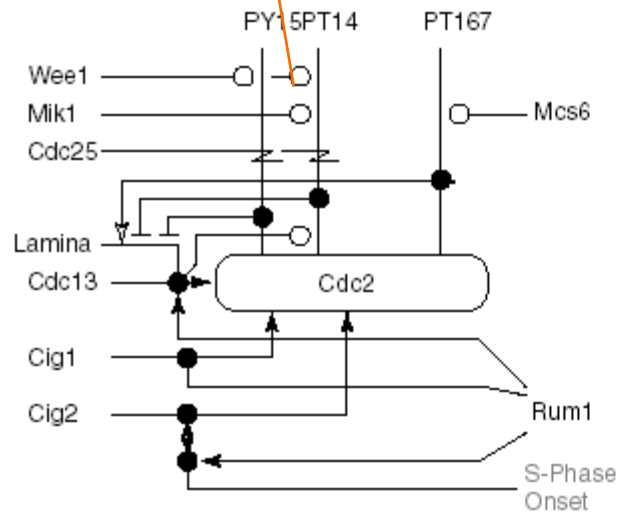
K.W. Kohn. Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Molecular Biology of the Cell* 10(8):2703-34, 1999.

Luca Cardelli



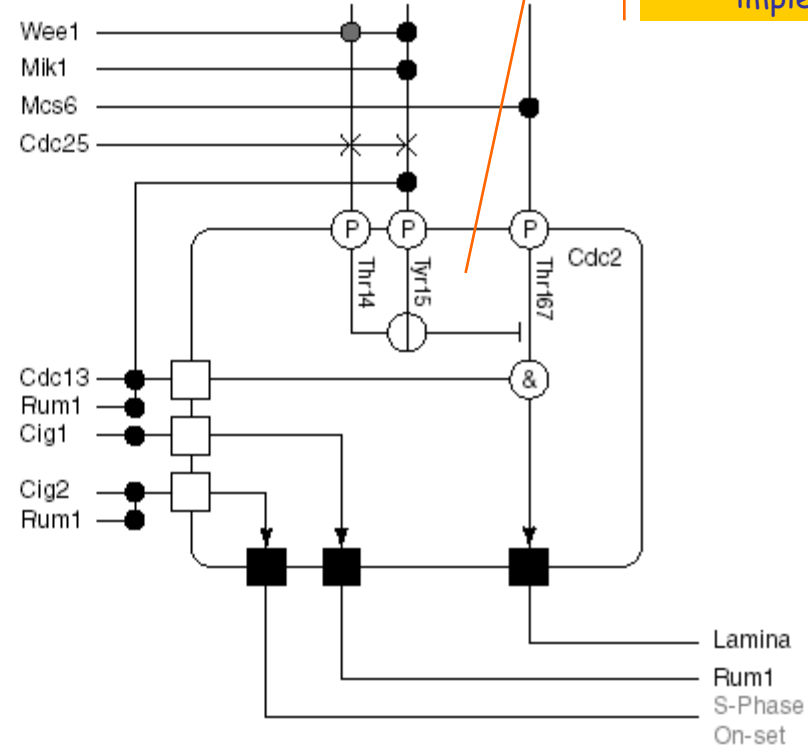
# Kitano Diagrams

From direct graphical representation of chemical reactions



(a) Graphical representation of fission yeast Cdc2 in Kohn diagram

To more abstract representation of the logic such reactions implement



(b) Proposed improvements of graphical representation of fission yeast Cdc2

BioSilico

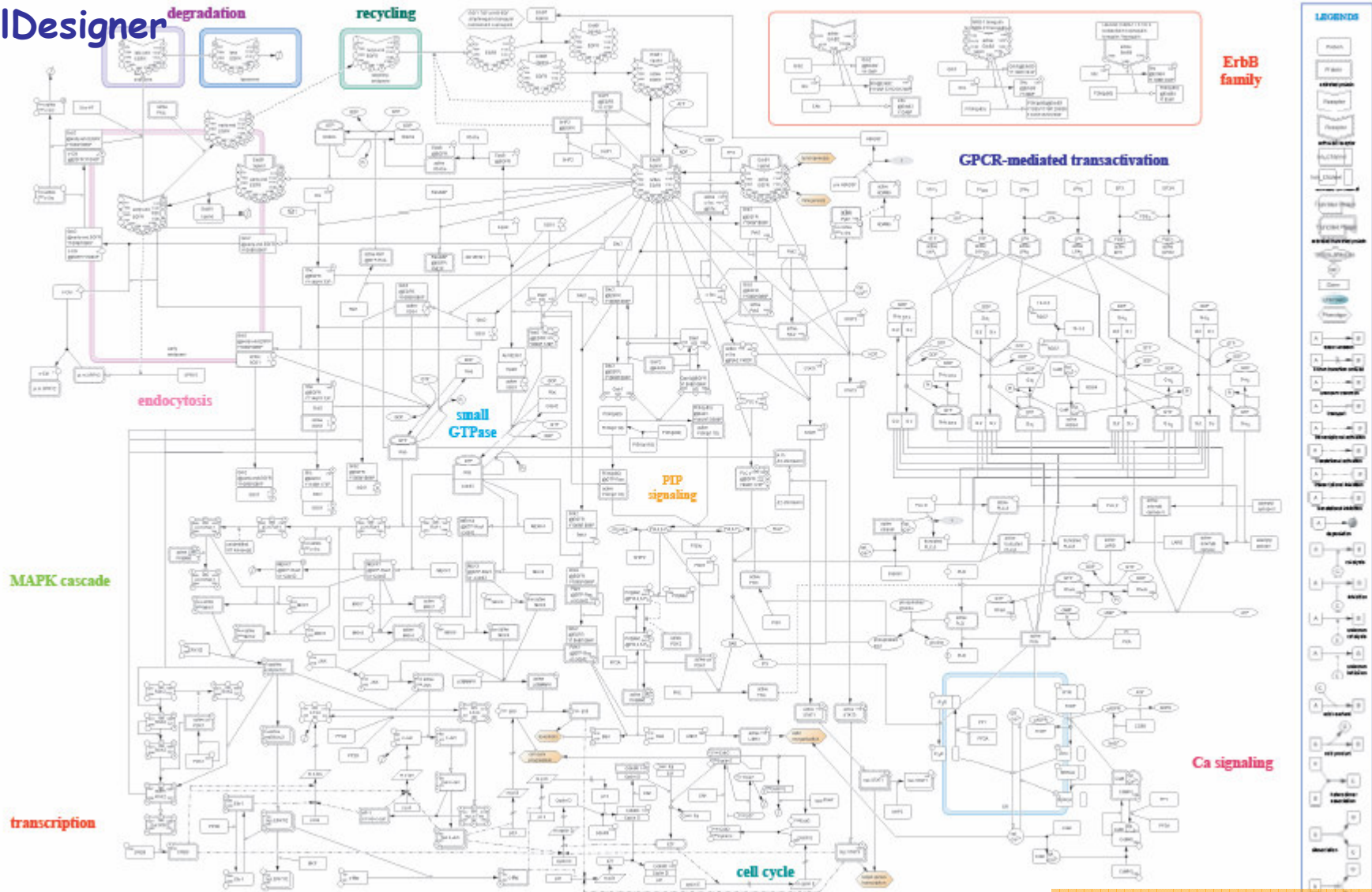
**Figure 1.** Representation of fission yeast Cdc2 protein in (a) the original MIM and (b) proposed improvements. Both diagrams represent interactions involving fission yeast Cdc2. Wee1 phosphorylates Thr14 and Tyr15, Mik1 phosphorylates Tyr15, Mcs6 phosphorylates Thr167, and Cdc25 dephosphorylates Thr14 and Tyr15. Cdc2 binds to either Cdc13, Cig1, or Cig2. When Cdc2 is forming a complex with Cdc13 and only Thr167 is phosphorylated, the complex interacts with Lamina. Phosphorylation of either Thr14 or Tyr15 inhibits activation of Cdc2 due to phosphorylation of Thr167. The complex auto-phosphorylates Tyr15 of its Cdc2. The complex of Cdc2 and Cig1 interacts with Rum1. Cdc2-Cdc13 complex and Cdc-Cig2 complex form heterotrimeric involving Rum1.

# Molecular Interaction Maps (Kitano)

Epidermal Growth Factor Receptor Pathway Map

Kaneko Oda H, Yuhito Matsuda R, Hiroaki Kitano H (2003)  
 The Systems Biology Institute, 252-5282, 1-1-1 Higashi, Suita City, Osaka, Japan  
 © 2003 SIBICI. All rights reserved. This work is licensed under a Creative Commons Attribution License.

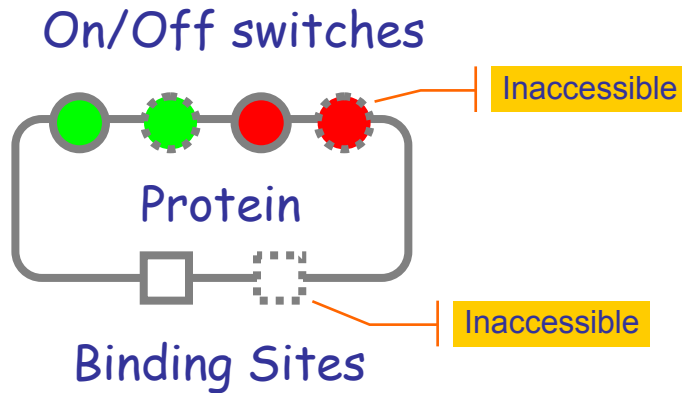
CellDesigner



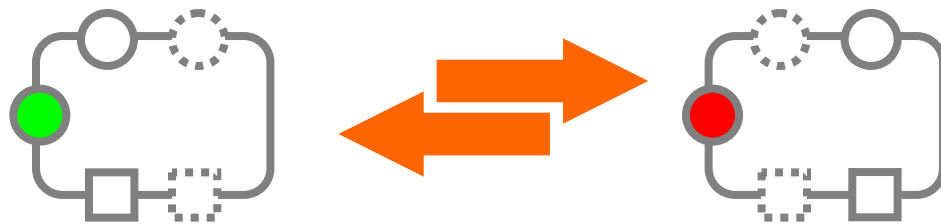
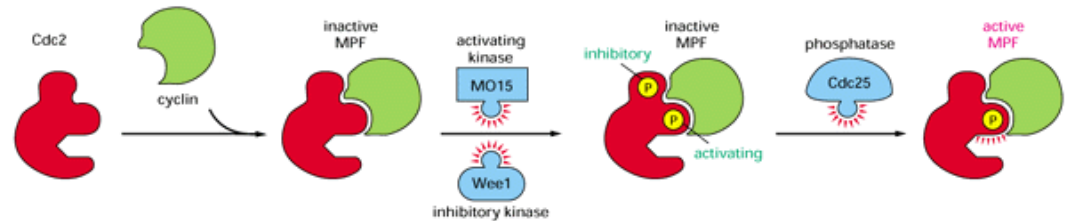


# The Protein Machine "Instruction Set"

cf. BioCalculus [Kitano&Nagasaki],  $\kappa$ -calculus [Danos&Laneve]

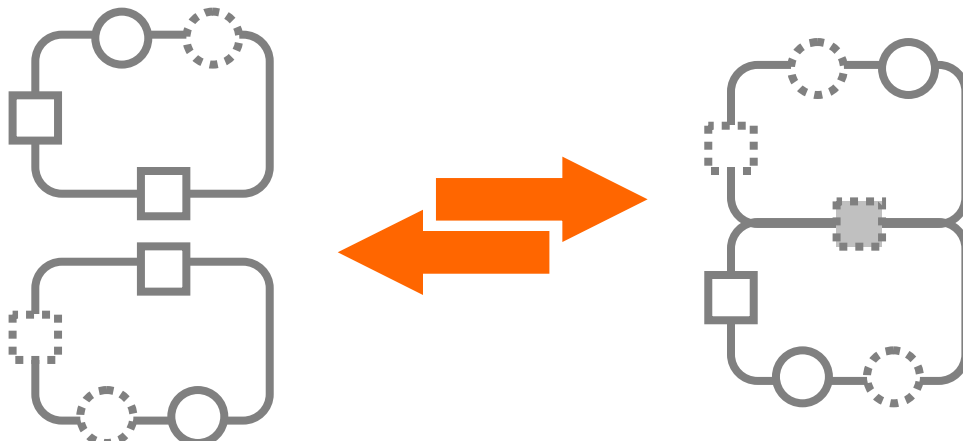


Each protein has a structure of binary switches and binding sites. But not all may be always *accessible*.



Switching of accessible switches.

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



Binding on accessible sites.

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

# Notations for the Protein Machine

- **Stochastic  $\pi$ -Calculus**
  - Priami (following Hillston's PEPA) formalizes a stochastic version of p-calculus where channels have communication *rates*.
- **BioSPi**
  - Regev-Shapiro-Silverman propose modeling chemical interactions (exchange of electrons and small molecules) as "communication".
  - Standard stochastic simulation algorithms (Gillespie) can be used to run in-silico experiments.
  - Complex formation is encoded via p-restriction.
- **PEPA**
  - Calder Gilmore and Hillston model the ERK pathway.
- **k-calculus**
  - Danos and Laneve (following Kitano's BioCalculus) define a calculus where complex formation is primitive.
- **(Stochastic) Petri Nets**
  - S.Reddy'94 modeling pathways.
  - Srivastava Perterson and Bentley analyze and simulate E.coli stress response circuit.
- **Bio State Charts**
  - Harel uses State Charts to model biological interactions via a semi-graphical FSM notation.
- **Pathway Logic**
  - Talcott-Eker-Knapp-Lincoln use term-rewriting.
- **BioCham**
  - ChabrierRivier-Fages-Soliman use term-rewriting and CLT modelchecking.
- **Kohn Diagrams, Kitano Diagrams**
- **SBML (Systems Biology Markup Language)**
  - XML dialect for MIM's:
    - Compartments (statically nested)
    - Reagents with concentrations
    - Reactions with various rate laws
  - Read and written by many tools via the Systems Biology Workbench protocol

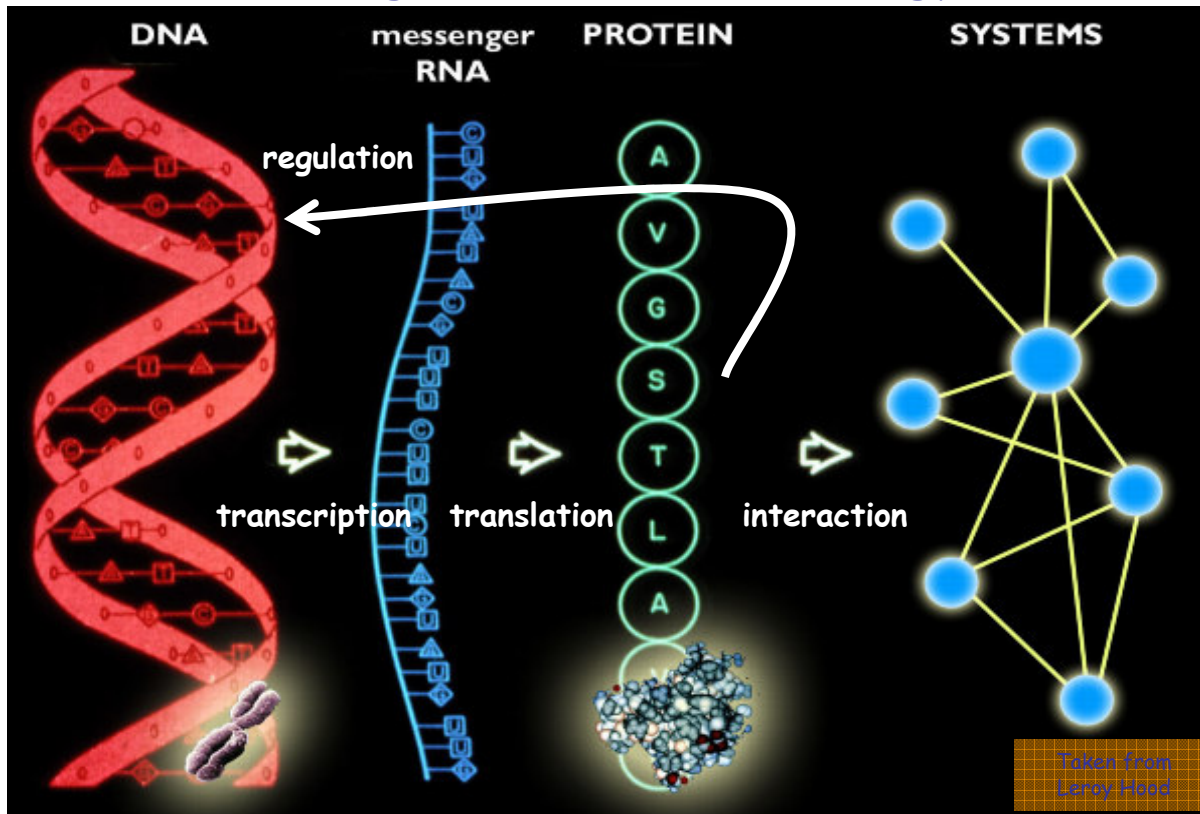
# The Gene Machine



# 2. The Gene Machine

*Pretty far from the atoms.*

The "Central Dogma" of Molecular Biology

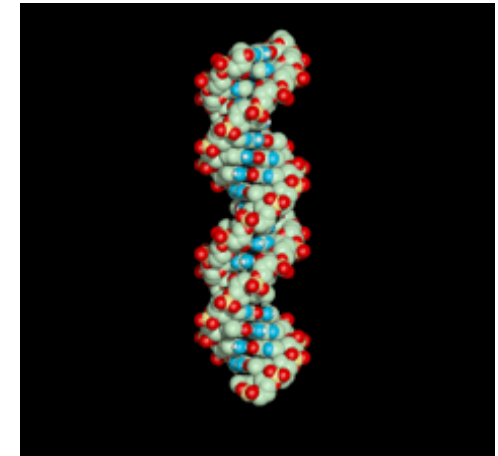


4-letter digital code

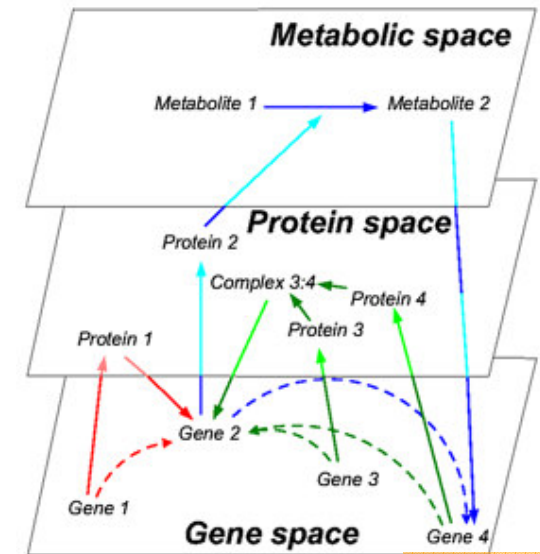
4-letter digital code

20-letter digital code

50.000(?) shapes

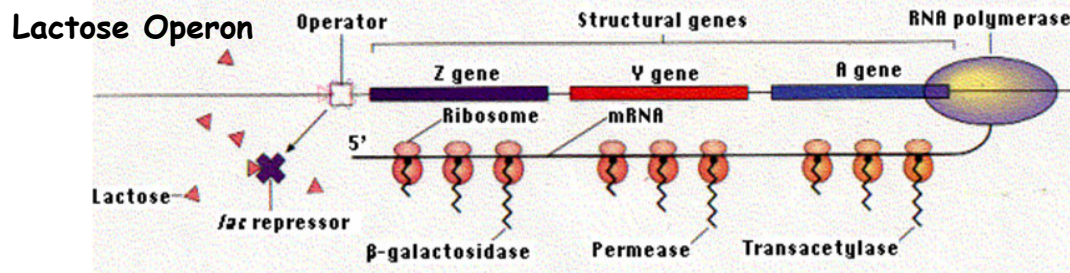


[DNA Tutorial](#)



Taken from Pedro Mendes

2006-05-26



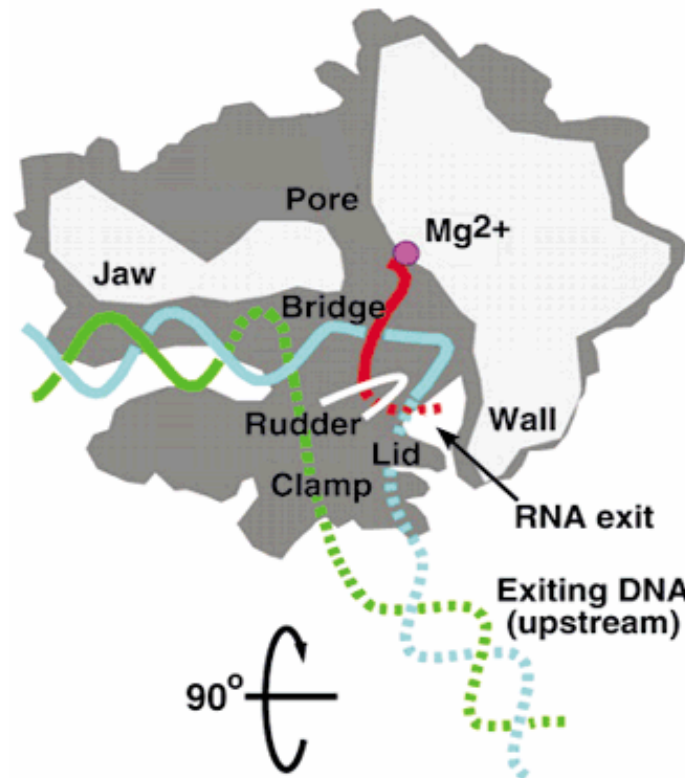
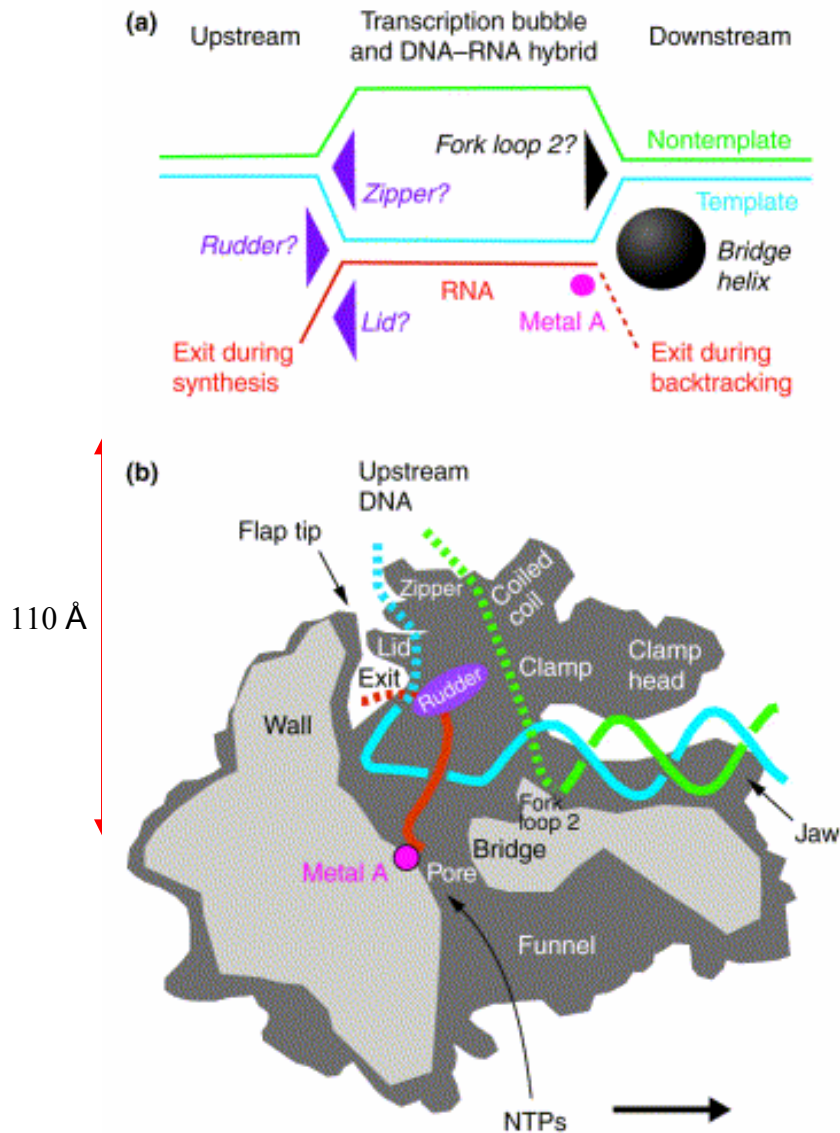
Luca Cardelli

# RNA Polymerase: DNA to RNA

## Multisubunit RNA polymerases.

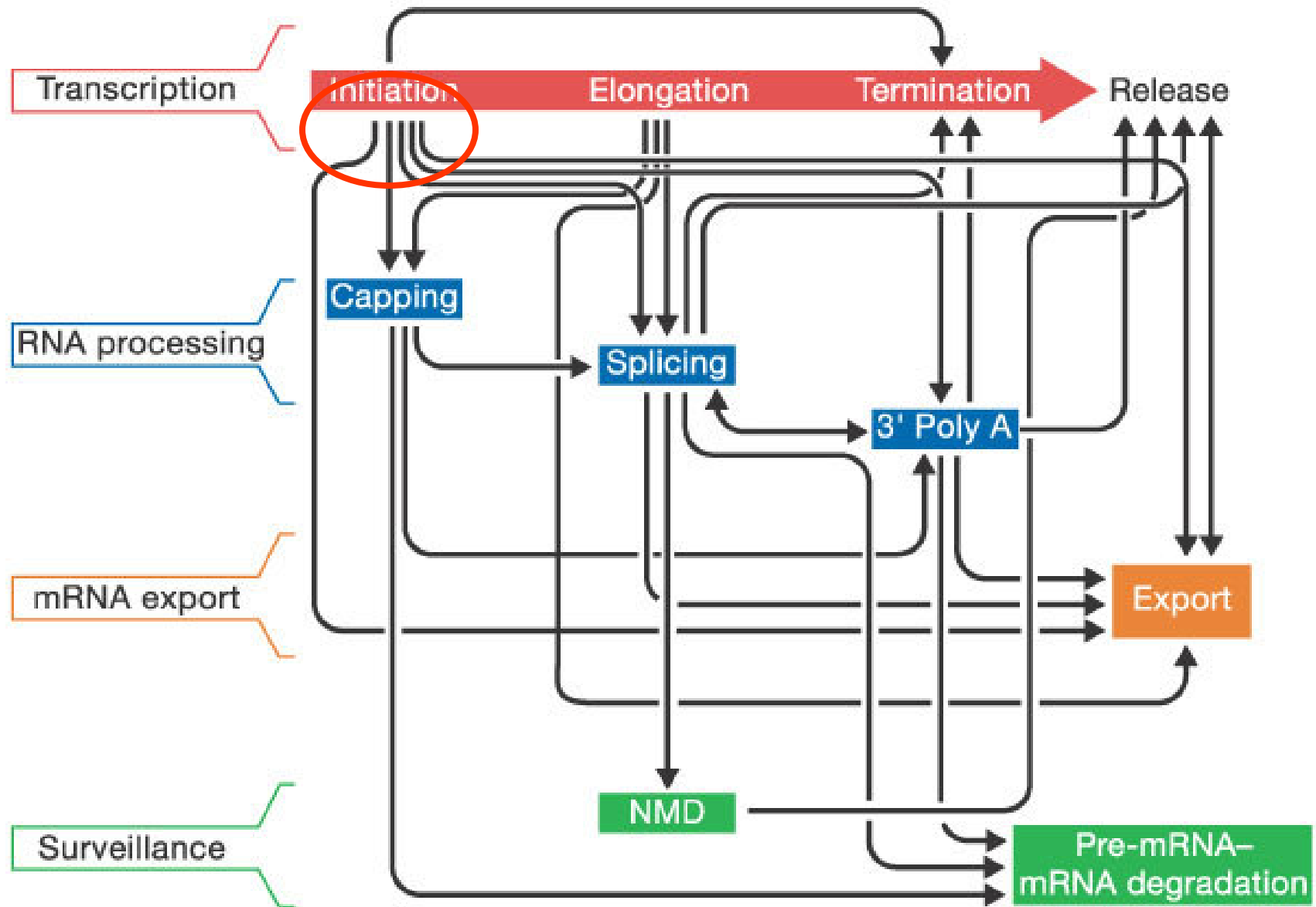
Cramer P (2002) *Curr Opin Struct Biol.* 12:89-97

Yeast RNA polymerase II is a 12 subunits complex (mwt. 400 kDa, size 140 Å x 140 Å x 110 Å). Its structure at 2.8 Å resolution reveals a division of the polymerase into four mobile modules, including a clamp, shown previously to swing over the active center. The clamp is in an open state, allowing entry of straight promoter DNA for the initiation of transcription



RNA polymerase II takes 30 seconds to one minute to transcribe a 1 kilobase long gene

# Transcription



# Transcription

	E. Coli	H. Sapiens
Number of genes	3.200	25.000
Number of different transcription factors	150	2.000

Gene length	$10^2$ to $10^3$ bp :	400 genes	
	$10^3$ to $10^4$ bp :	6000 genes,	1 to 10 minutes
	$10^4$ to $10^5$ bp :	12.200 genes	10 to 100 min.
H. Sapiens	$10^5$ to $10^6$ bp :	3200 genes	
	$10^6$ to $2.4 \cdot 10^6$ bp :	62 genes	16 to 24 h

Number of transcripts per gene : 0,1 to  $10^6$ .

Up to  $10^6$  mRNA globin in erythroid cells

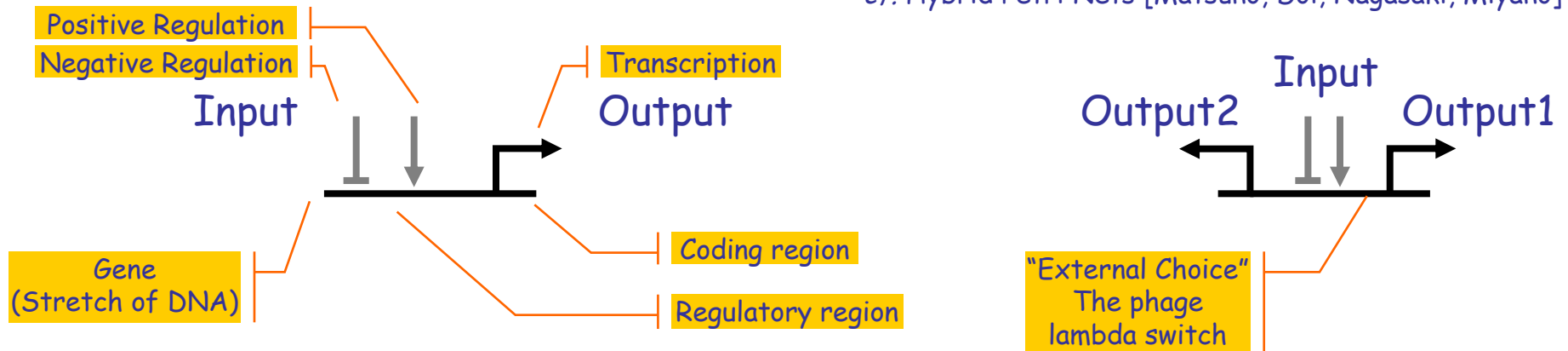
15.000 mRNA actin / nucleus in muscle fibers

10 to 100 copies for mRNAs encoding transcription factors

In an HeLa cell, there are approx. 100 mol. RNA pol / gene coding rRNA

# 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)

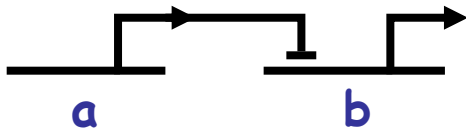
580,073bp 145KB (eBook)

E.Coli (bacteria): 4Mbp 1MB (floppy)

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

Wheat 17Gbp 4.25GB (DVD)

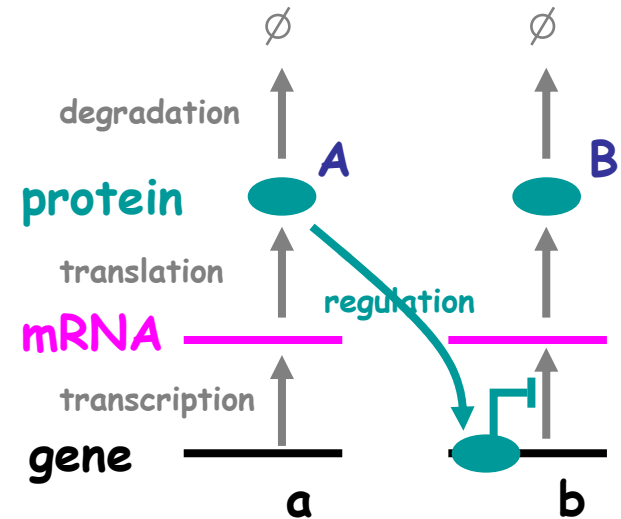
# Gene Composition



Is a shorthand for:

Under the assumptions [Kim & Tidor]

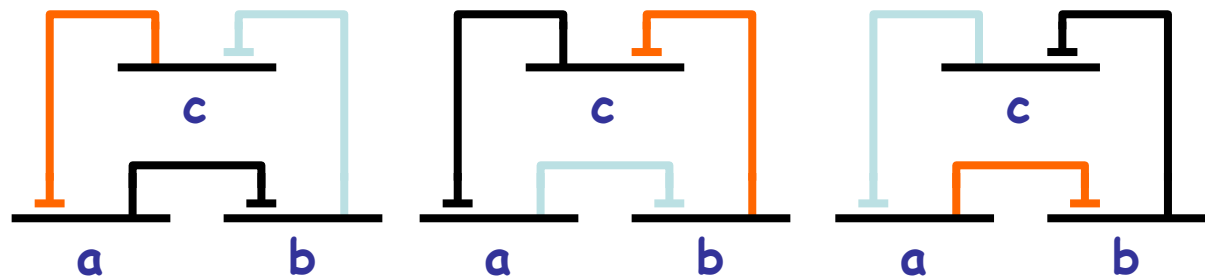
- 1) The solution is well-stirred  
(no spatial dependence on concentrations or rates).
- 2) There is no regulation cross-talk.
- 3) Control of expression is at transcription level only  
(no RNA-RNA or RNA-protein effects)
- 4) Transcriptions and translation rates monotonically affect mRNA and protein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



Ex: Bistable Switch



Ex: Oscillator

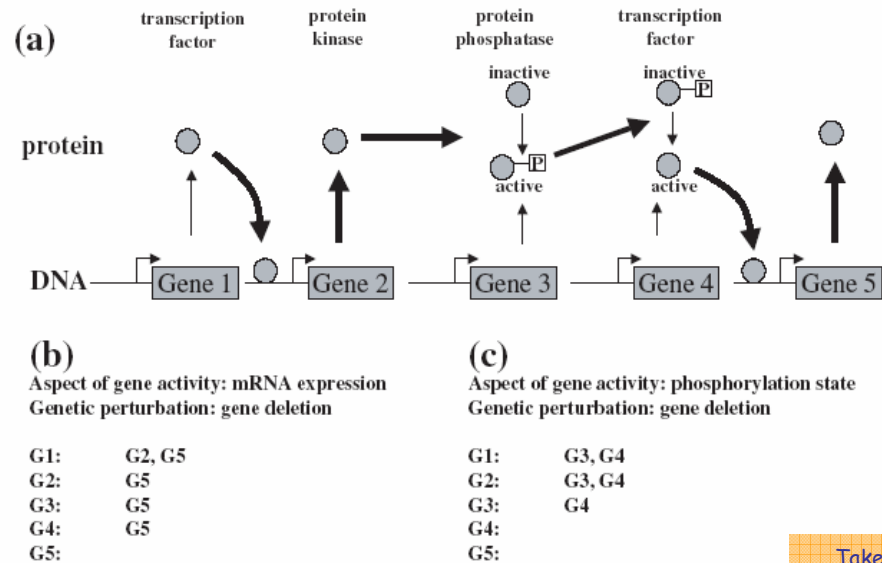


Expressed  
Repressed  
Expressing



# Indirect Gene Effects

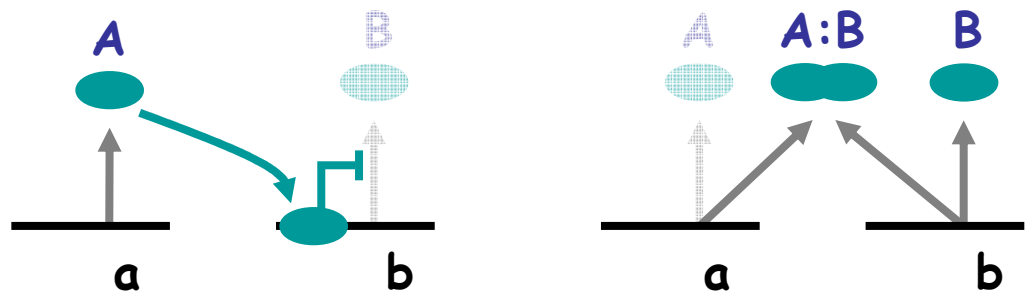
No combination of standard high-throughput experiments can reconstruct an a-priori known gene/protein network [Wagner].



Taken from Andreas Wagner

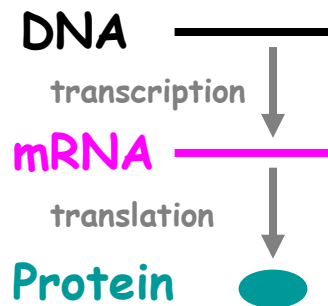
**Fig. 1.** The importance of specifying gene activity when reconstructing genetic networks. (a) A hypothetical biochemical pathway involving two transcription factors, a protein kinase, and a protein phosphatase, as well as the genes encoding them. See text for details. (b) Shown is a list of perturbation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



RNA is not just an intermediary; it can:

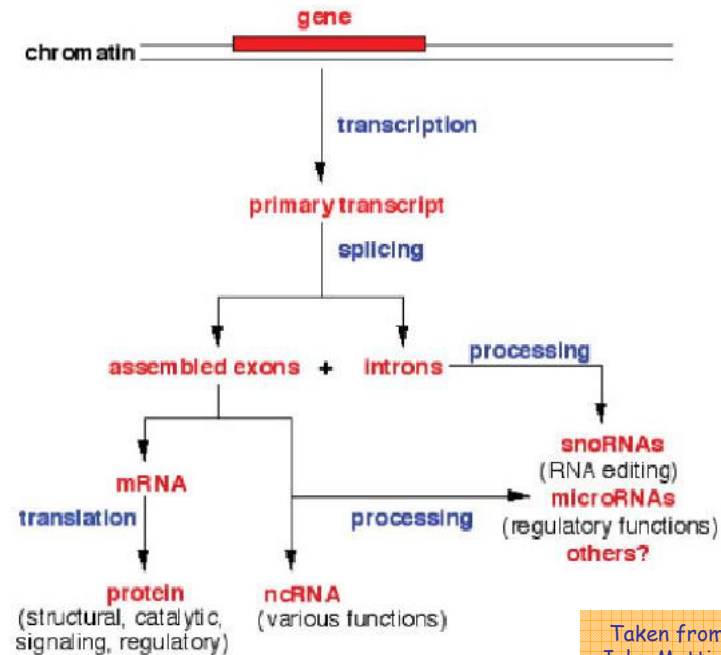
- Fold-up like a protein
- Act like an enzyme
- Regulate other transcribed RNA
- Direct protein editing
- ...

97-98% of the transcriptional output of the human genome is non-protein-coding RNA.

30-40,000 "protein genes" (1.5% of genome)

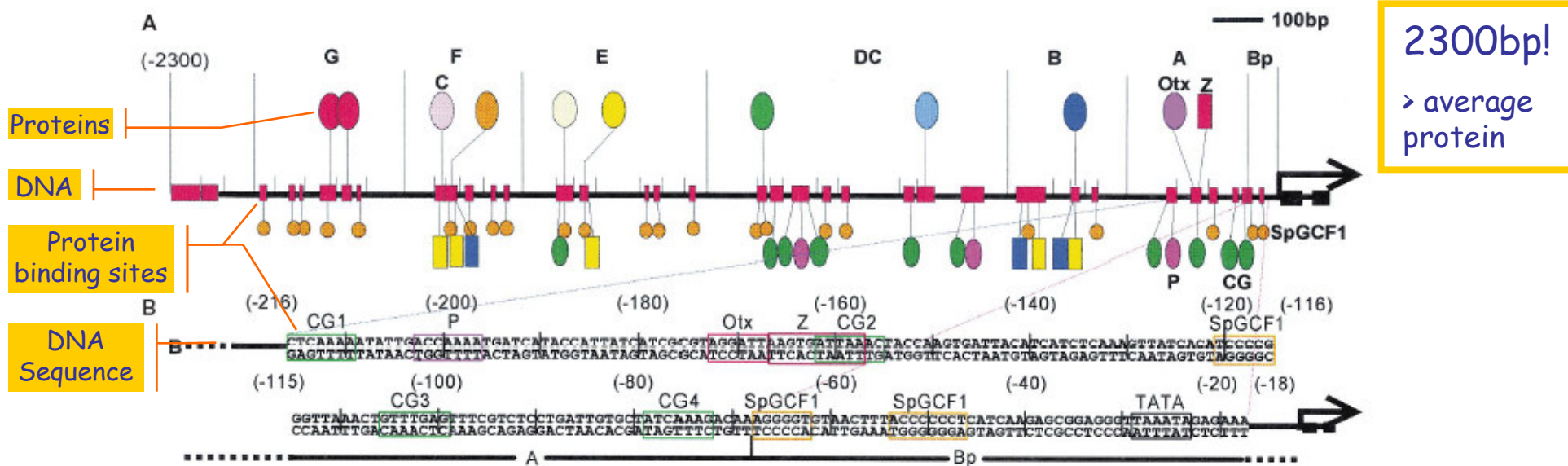
60-100,000 "transcription units" (>30% of genome is transcribed)

## Challenging the Dogma (in higher organisms)



**Figure 1.** A revised view of the flow of genetic information in the higher eukaryotes. Primary transcripts may be (alternatively) spliced and further processed to produce a range of protein isoforms and/or ncRNAs of various types, which are involved in complex networks of structural, functional and regulatory interactions.

# Structure of a Regulatory Region



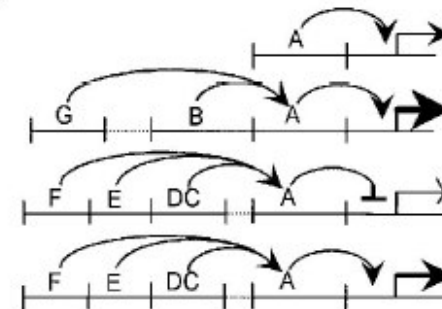
## C Module A functions:

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):

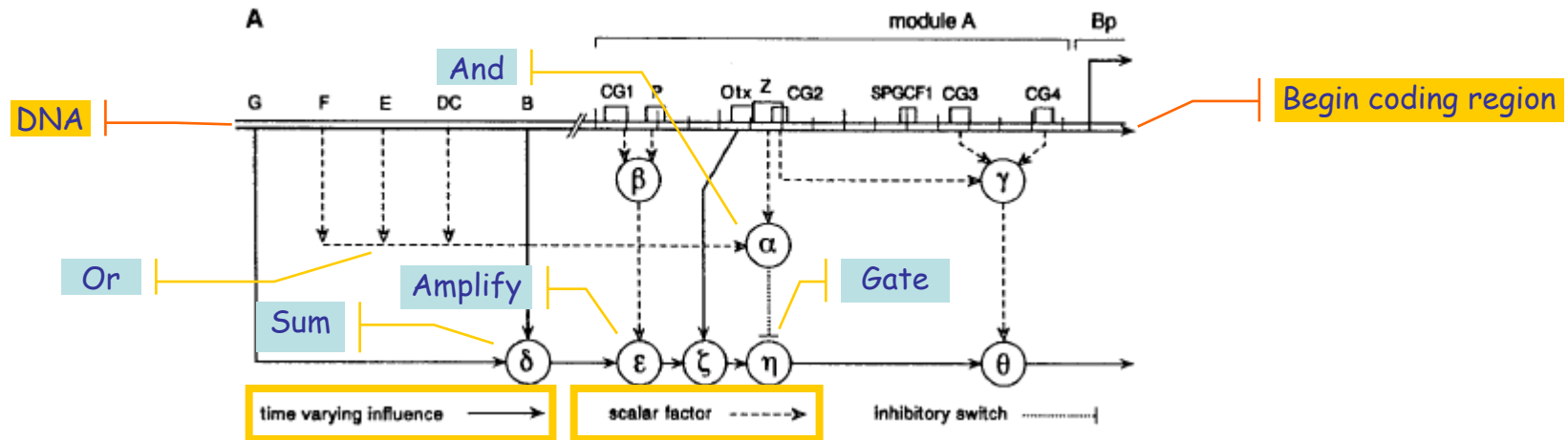
Modules E, F and DC with LiCl treatment:



**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<sub>3</sub> and CG<sub>4</sub> 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

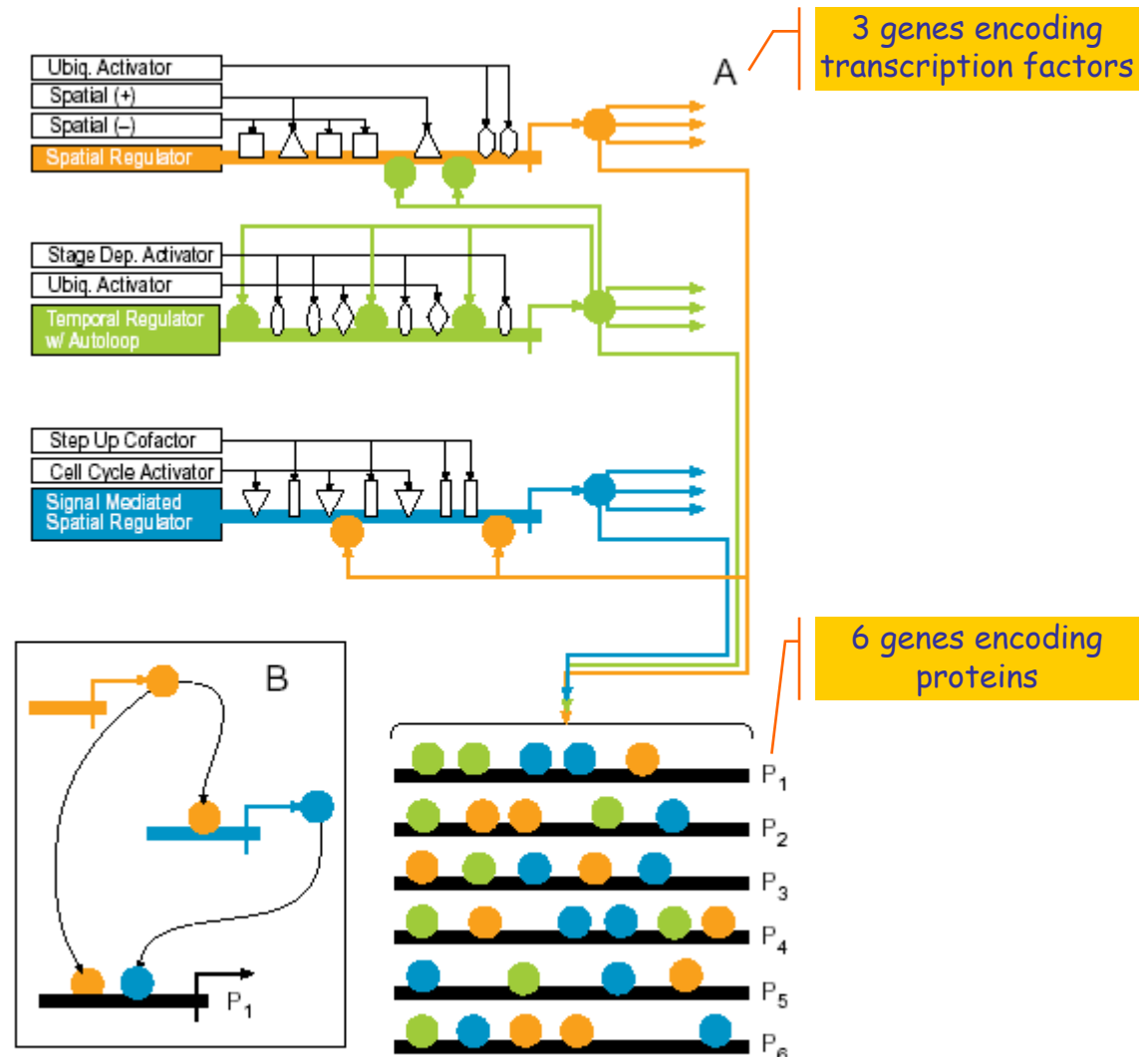
if (F = 1 or E = 1 or CD = 1) and (Z = 1)      Repression functions of modules F, E, and DC mediated by Z site  
 $\alpha = 1$   
 else  $\alpha = 0$   
 if (P = 1 and CG<sub>1</sub> = 1)      Both P and CG<sub>1</sub>, needed for synergistic link with module B  
 $\beta = 2$   
 else  $\beta = 0$   
 if (CG<sub>2</sub> = 1 and CG<sub>3</sub> = 1 and CG<sub>4</sub> = 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 CG<sub>1</sub>-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

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



# Regulation: Where/When/HowMuch

**Fig. 3.** A sector of an imaginary developmental gene regulatory network. (A) Network sector. Three genes encoding transcription factors are shown at the top. These are a spatial regulator (orange), a temporal regulator (green) and a signal-mediated regulator (blue). Genes encoding other transcription factors that originate off the diagram are indicated in black type on open backgrounds. A battery of six genes encoding some differentiation proteins (P1-P6) is shown below. Connections between the three genes encoding transcription factors and target sites in the P1-P6 genes are indicated by respectively colored bent arrows and the transcription factors as solid circles. The spatial regulatory gene is controlled by positive and negative interactions, which establish the limited spatial domain where it will be expressed, and it utilizes a ubiquitous ancillary activator to achieve an appropriate level of expression. This gene would be expressed only at certain stages due to requirement for the factor produced by the green temporal regulator, shown below the line binding to its target sites in the *cts*-regulatory DNA. The *cts*-regulatory system of the temporal regulator responds to its own transcription factor, and also depends on a factor appearing only after a certain stage of development, and on another ubiquitous ancillary activator. The signal-mediated regulator produces a factor that is activated by signals. For example, if this were a short-range signal produced by cells adjacent to the domain of expression of the spatial orange regulator, P1-P6 would be expressed only near the boundary. The *cts*-regulatory system controlling expression of the signal-mediated transcription factor includes target sites for the product of the orange spatial regulator, shown binding below the line representing the DNA, and also for two factors that work together to promote transcription during growth, one imagined as a regulator produced when cells are cycling, the other as a ubiquitous co-factor. The arrows at the right indicate that each of the three genes encoding transcription factors have many downstream targets besides the P1-P6 gene battery. Any resemblance between this network sector and a known regulatory element is purely coincidental. (B) A single relationship extracted from the network. A causal diagram is shown portraying the multilevel function of the orange spatial regulator, which controls both the gene encoding the blue signal-mediated regulator and the P1 gene; the latter, however, is also directly responsive to the spatial regulator.



3 genes encoding transcription factors

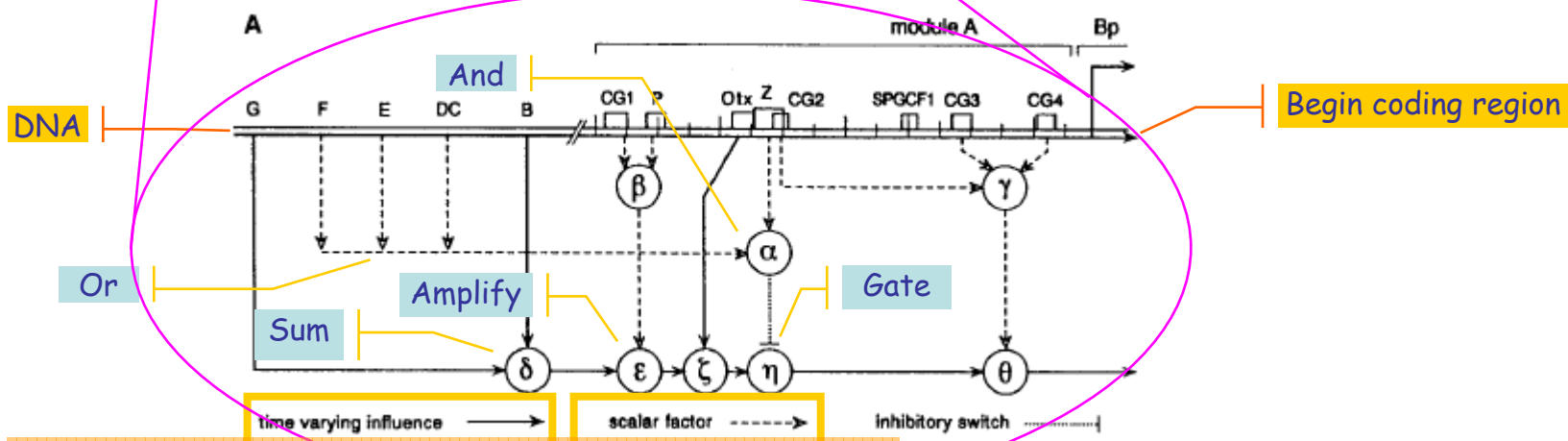
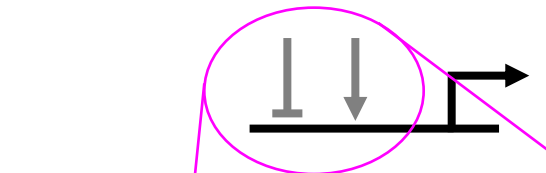
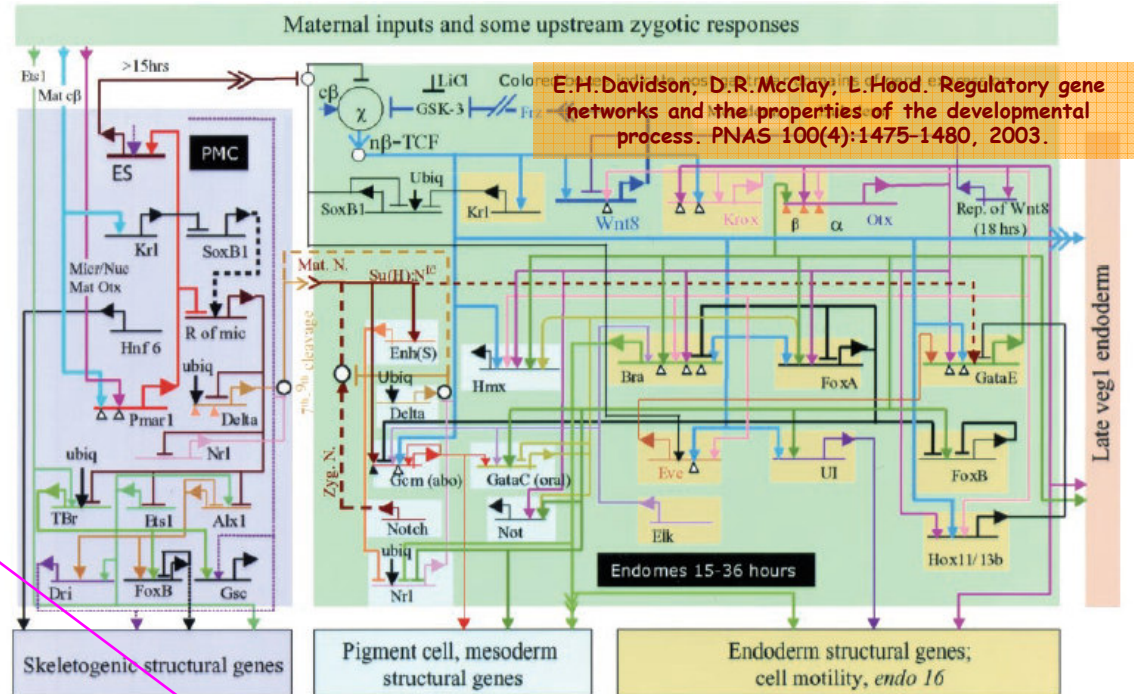
6 genes encoding proteins



# Gene Regulatory Networks

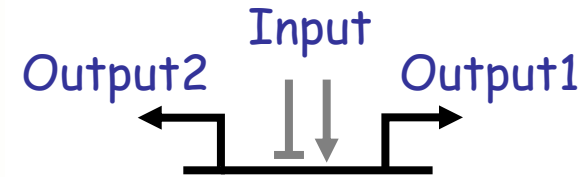
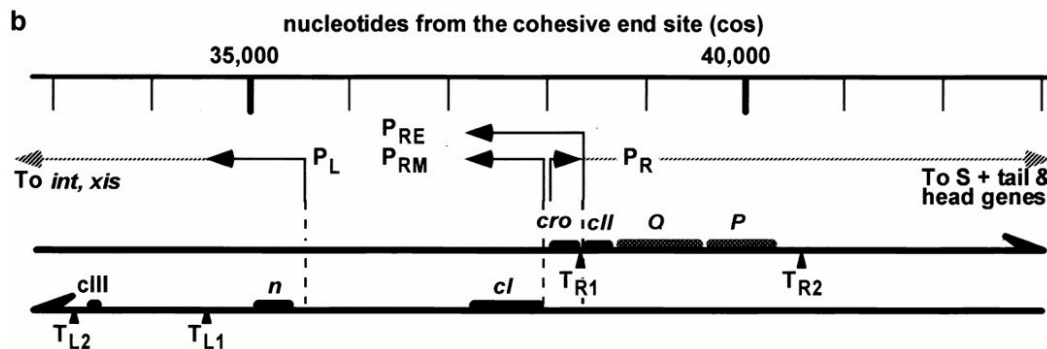
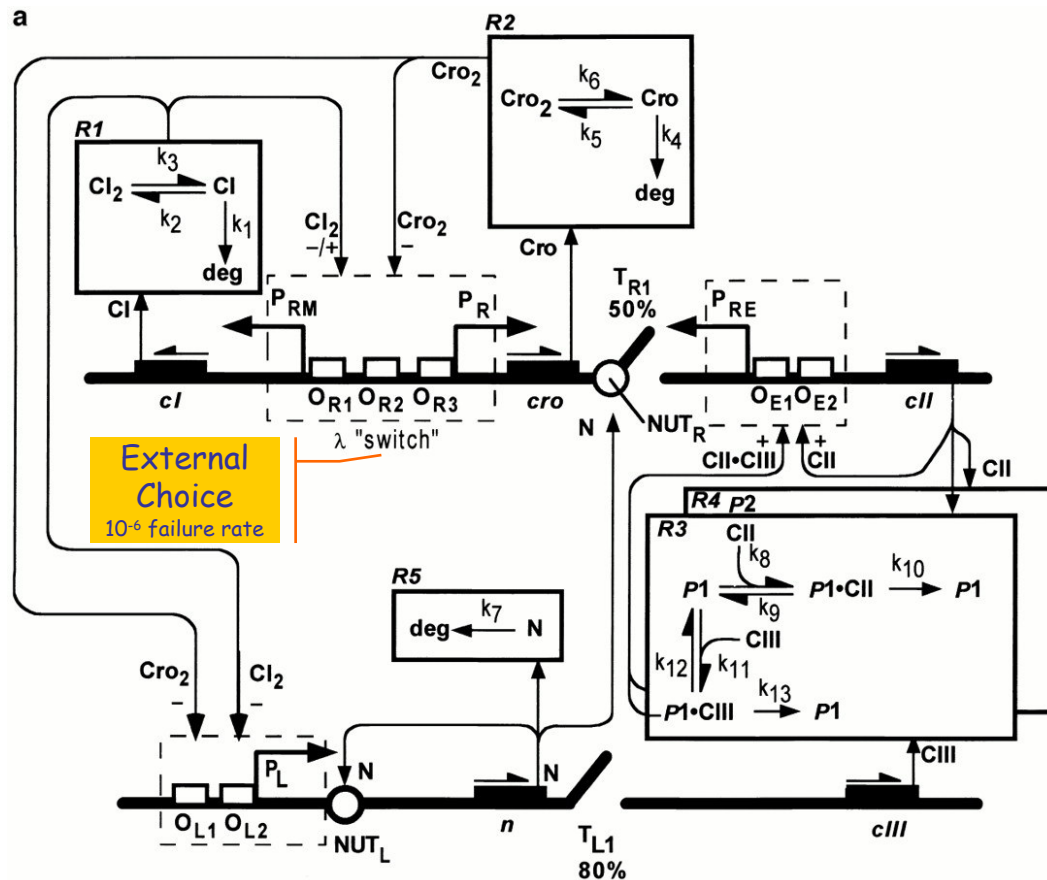
<http://strc.herts.ac.uk/bio/maria/NetBuilder/>

NetBuilder



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

# Phage Lambda Decision Circuit



**Figure 1.** The phage lysis-lysogeny decision circuit. (a) Bold horizontal lines indicate stretches of double-stranded DNA. Arrows over genes indicate direction of transcription. Dashed boxes enclose operator sites that comprise a promoter control complex. The three operator sites,  $O_{R1-3}$ , of the "lambda switch" implement concentration-dependent logic controlling promoters  $P_{RM}$  and  $P_R$ . Cro and CI dimers bind to the three sites with different affinities and in opposite order to control the activation level of the  $P_{RM}$  and  $P_R$  promoters (PTASHNE 1992 ; SHEA and ACKERS 1985 ). The five boxes R1–R5 contain nongenetic protein reaction subsystems. In R1, R2, and R5, "deg" indicates degradation. When protein N is available, transcribing RNAPs can be antiterminated at the NUTR and NUTL sites; termination sites  $T_{R1}$  and  $T_{L1}$  are inoperative for antiterminated RNAPs. The CI dimer acts as either a repressor or activator of promoter  $P_{RM}$ , depending on its concentration. See text for discussion of the proteases labeled as P1 and P2 in R3 and R4. (b) decision circuit DNA organization. Phage-encoded genetic elements of the decision circuit are located in a 5000 nucleotide region of the phage DNA. Genes are separated onto leftward and rightward transcribed strands as indicated by the arrows. Rightward extensions of the antiterminated  $P_R$  transcript transcribe the O and P genes essential for phage genome replication and the Q gene that controls transcription of later genes on the lytic pathway. Leftward extension of the antiterminated  $P_L$  transcript transcribes *xis* and *int* genes essential for phage chromosome integration and excision into and out of the host chromosome. Locations of four termination sites are indicated by  $T_{R1-2}$  and  $T_{L1-2}$ .

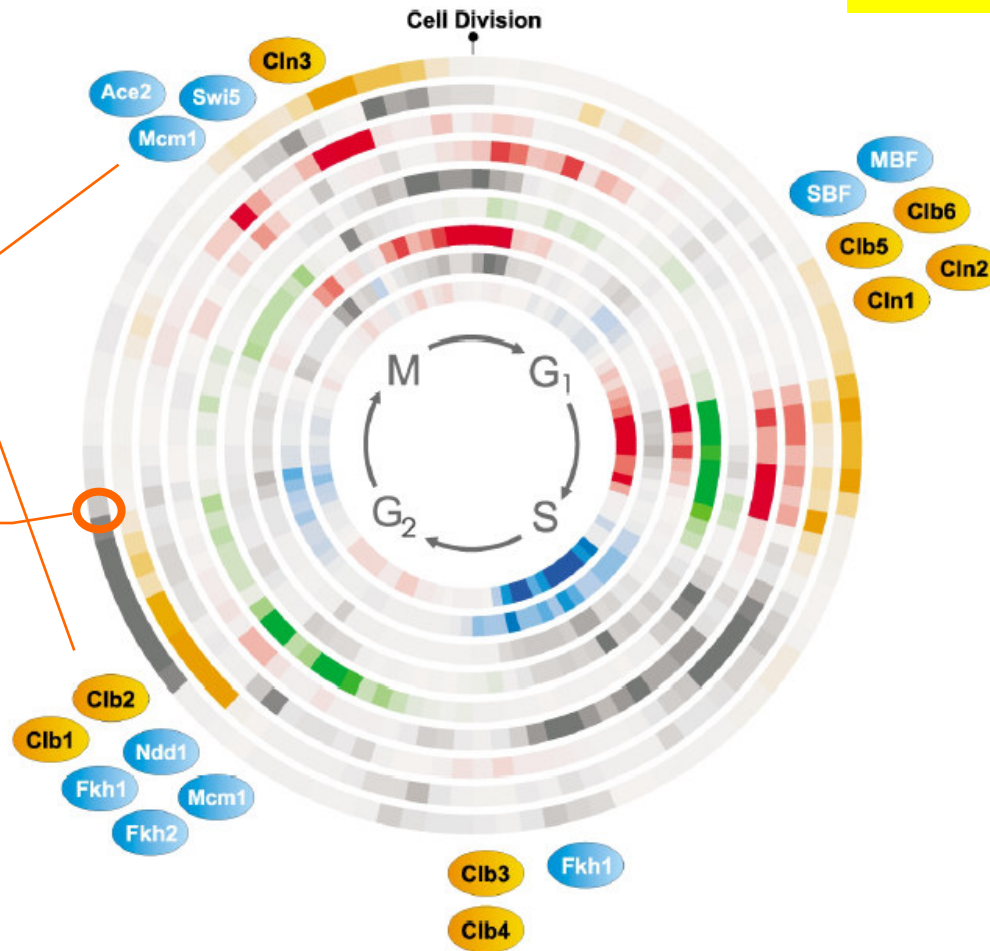
# Whole Genome Activity

## *S.cerevisiae* yeast cell cycle

Known cell cycle transcriptional activators

Cyclins

Average value of a given feature over all cell-cycle-proteins whose genes are maximally expressed at a given time in the cycle.



## Measured protein features

Y phosphorylation

S/T phosphorylation

O-glycosylation

N-glycosylation

Instability index

PEST regions

Extracellular

Nuclear

Isoelectric point

Figure 6. Feature variation during the cell cycle. The temporal variation in nine selected protein features during the cell cycle, with zero time (at the top of the plot) corresponding to the presumed time of cell division (M→G1 transition). The color scales correspond to  $\pm 2$  standard deviations from the cell cycle average. The concentric feature circles correspond to: isoelectric point, nuclear and extracellular localization predictions, PEST regions, instability index, N-linked glycosylation potential, O-GalNAc glycosylation potential, serine/threonine phosphorylation potential and tyrosine phosphorylation potential. The presumed positions of the four cell cycle phases G1, S, G2 and M are marked. Also depicted are known cell cycle transcriptional activators (marked in blue), positioned at the time where they are reported to function.

Protein Feature Based Identification of Cell Cycle Regulated Proteins in Yeast  
Ulrik de Lichtenberg, Thomas S. Jensen, Lars J. Jensen and Søren Brunak

# 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 automaton.
  - **Not Message-Passing between genes.** Messages themselves have behavior (e.g., they stochastically decay and combine), hence messages are processes as well.
  - **Not Data-Flow.** Any attempt to use data-flow-style modeling seems doomed because of widespread loops that lead to deadlocks or unbounded queues. Data-flow tokens do not "decay" like proteins.
- **How can it possibly work?**
  - **Stochastic broadcasting.** The apparently crude idea of broadcasting a whole bunch of asynchronous decaying messages to activate a future gate, means there are never any "pipeline full" deadlocks, even in presence of abundant feedback loops.
  - **Stochastic degradation.** Degradation is fundamental for system stability, and at the same time can lead to sudden instability and detection of concentration levels.

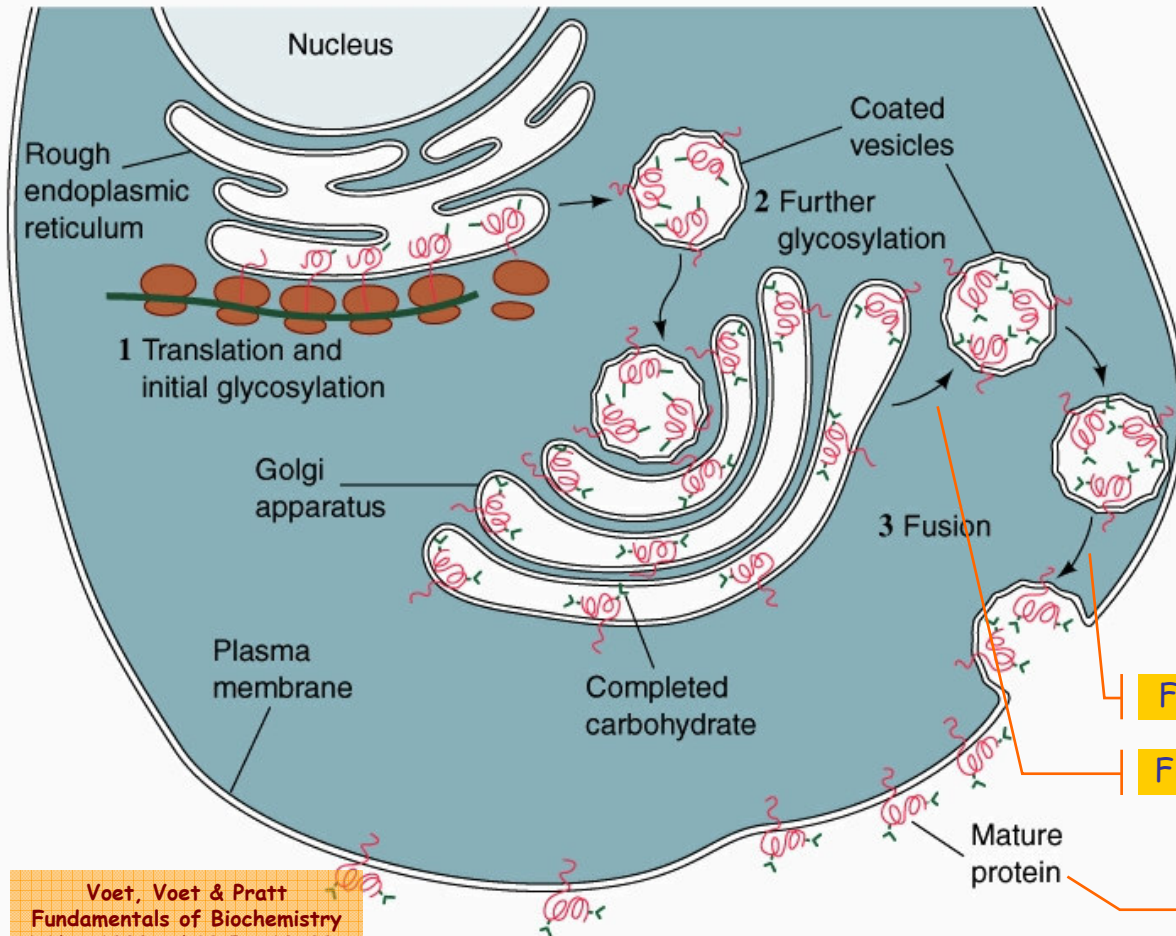
# Notations for the Gene Machine

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

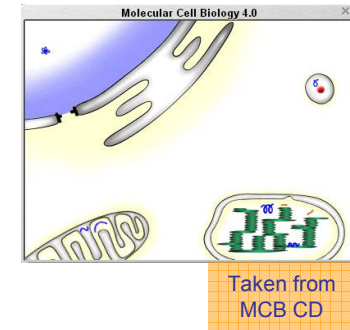


# **The Membrane Machine**

# 3. The Membrane Machine *Very far from the atoms.*



Molecular transport and transformation through dynamic compartment **fusion and fission**.



Fusion

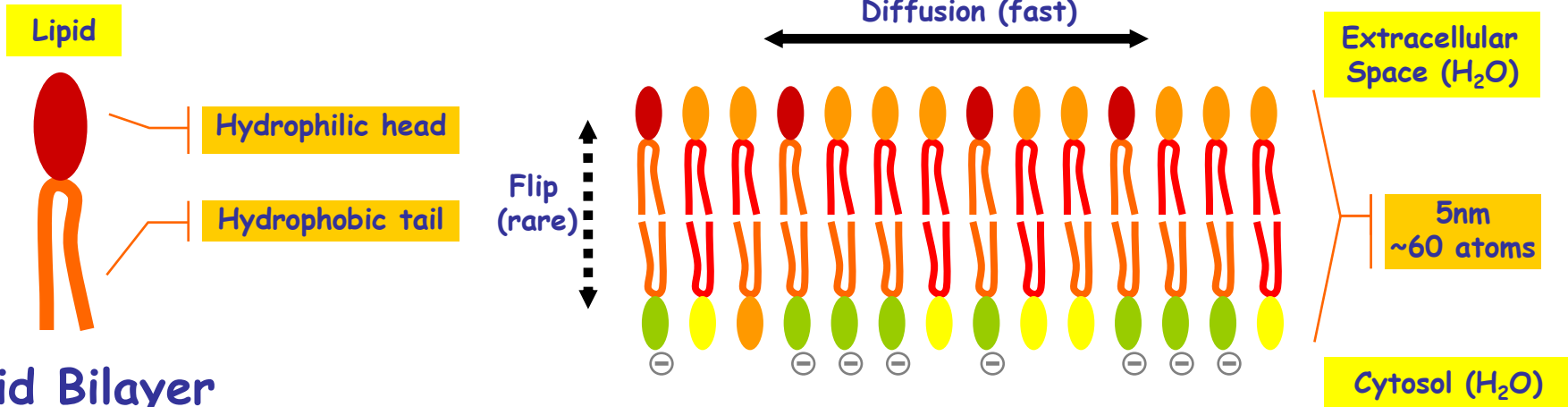
Fission

} The Instruction Set

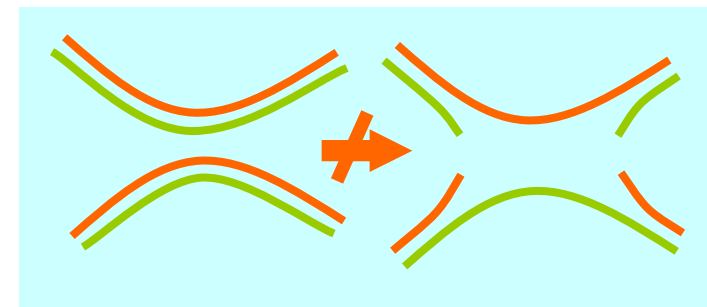
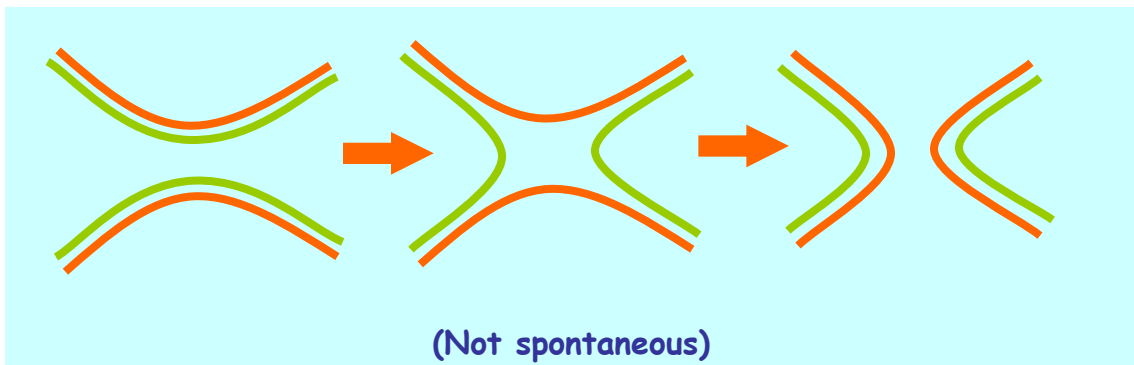
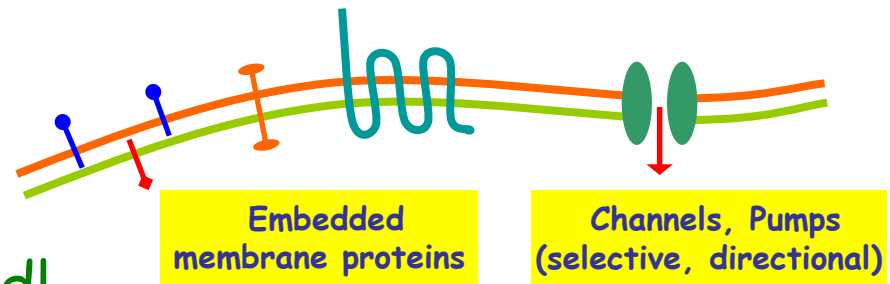
Voet, Voet & Pratt  
Fundamentals of Biochemistry  
Wiley 1999. Ch10 Fig 10-22.  
Copyright 1999, John Wiley and Sons, Inc. All rights reserved.

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

# Membranes are Oriented 2D Surfaces

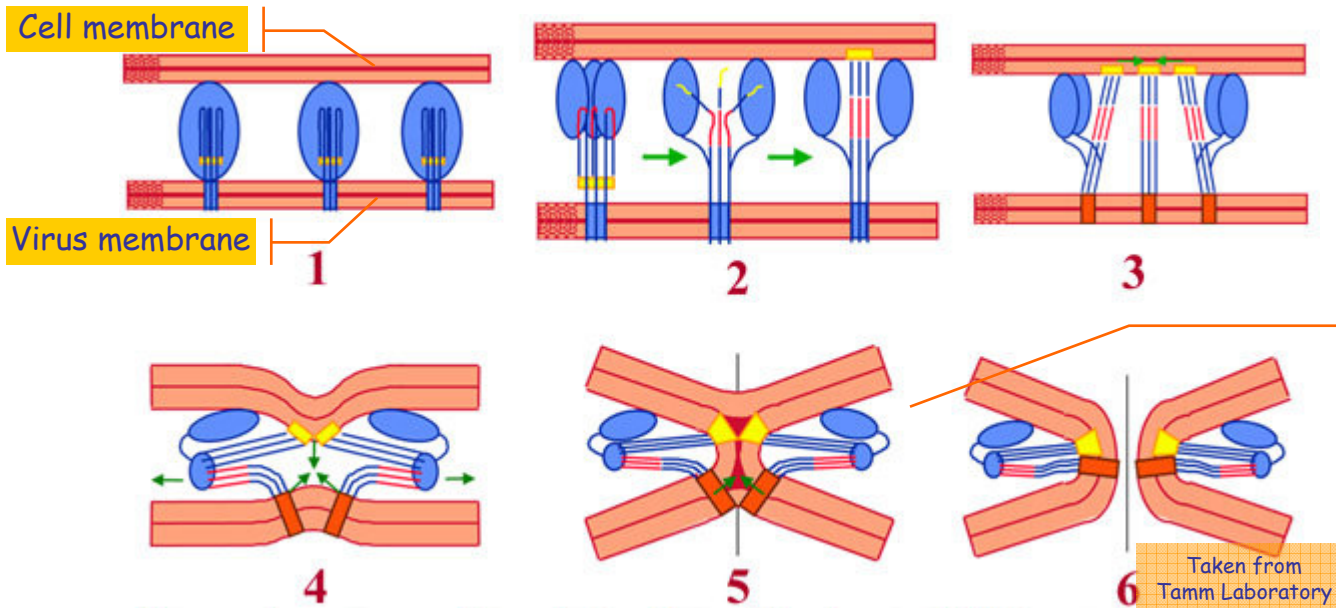


**Lipid Bilayer**  
 Self-assembling  
 Largely impermeable  
 Asymmetrical (in real cells)  
 With embedded proteins  
**A 2D fluid inside a 3D fluid!**



# Membrane Fusion

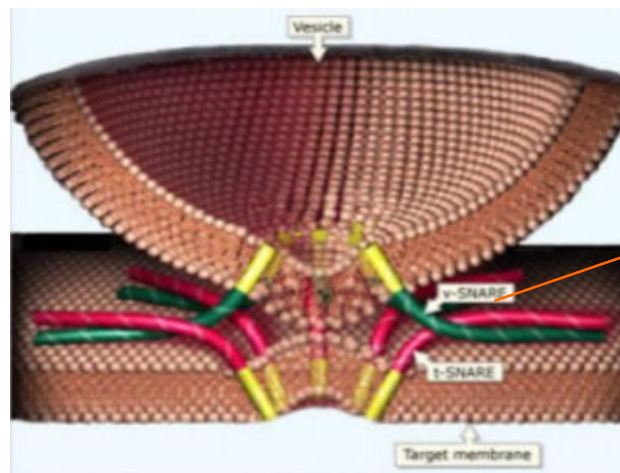
Positive curvature to  
Negative curvature  
transition in 3D



**Aggressive fusion  
(virus)**

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

*Proposed sequence of events in pH sensitive hemagglutinin membrane fusion*



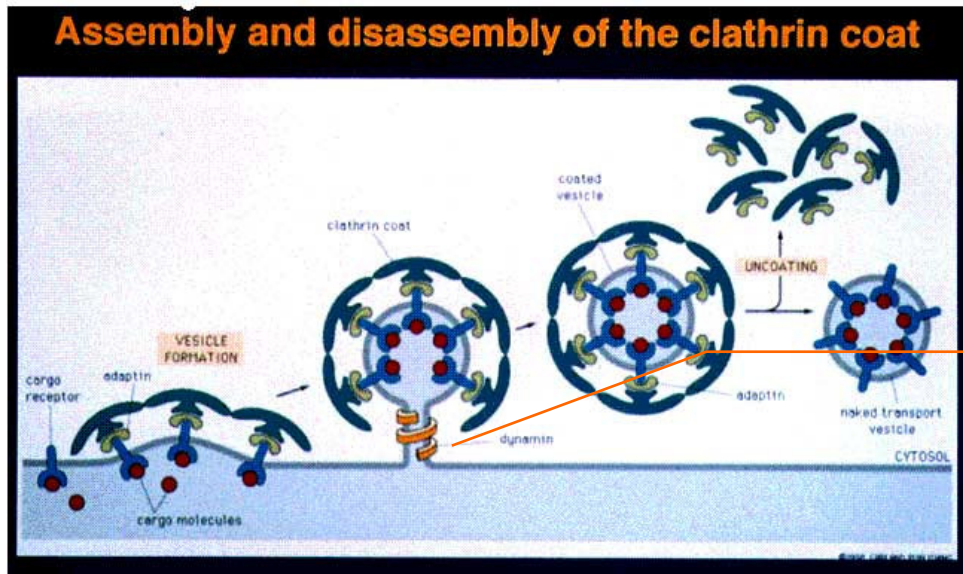
**Cooperative fusion  
(vesicle)**

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



# Membrane Fission

Negative curvature to Positive curvature transition in 3D

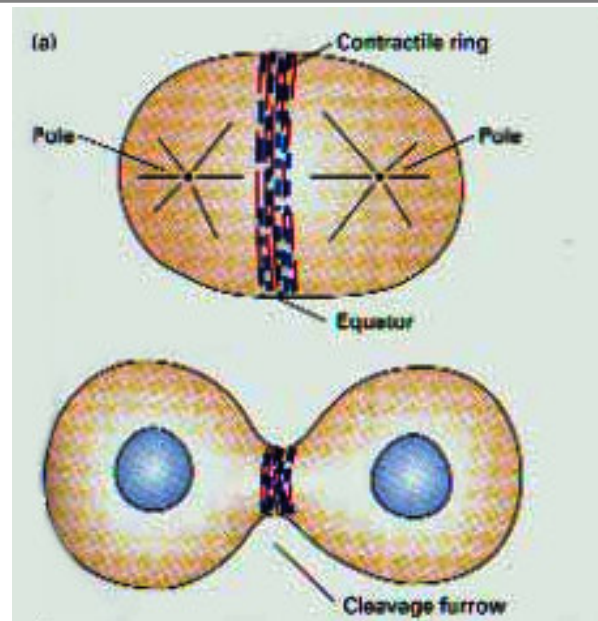


## Vesicle Formation



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]

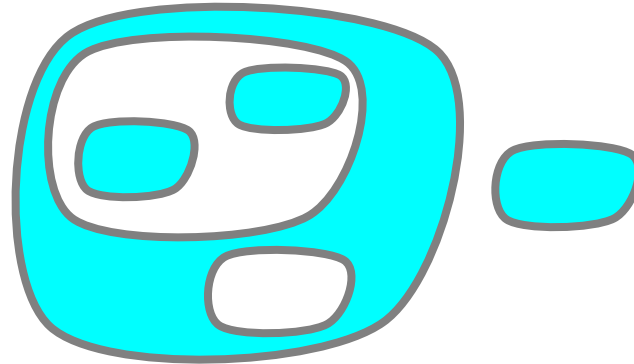


## Cytokinesis (Mitosis)



# Local Membrane Reactions

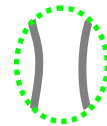
Membrane System



Local View



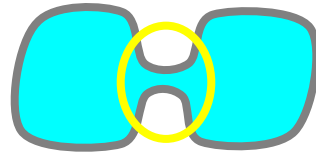
Switch



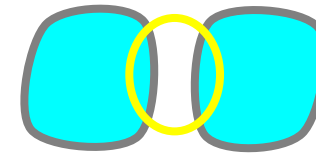
(Symmetric by 90° rotation.)

# Global Membrane Reactions

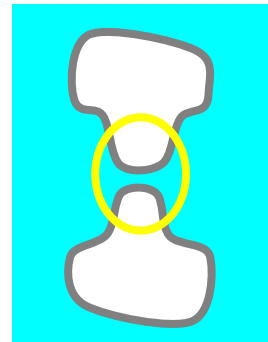
Global Views



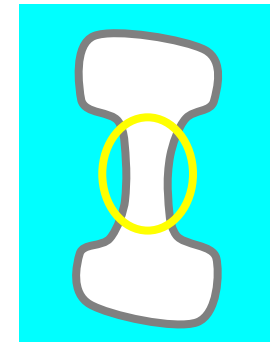
Mito →



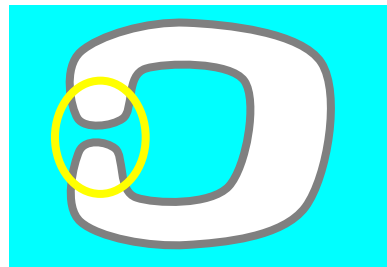
(Fission)



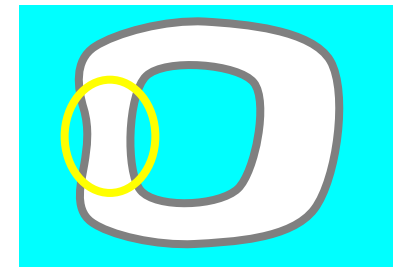
Mate →



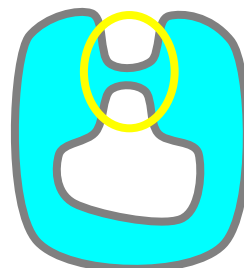
(Fusion)



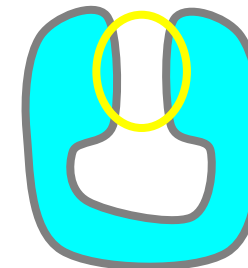
Endo →



(Fission)



Exo →

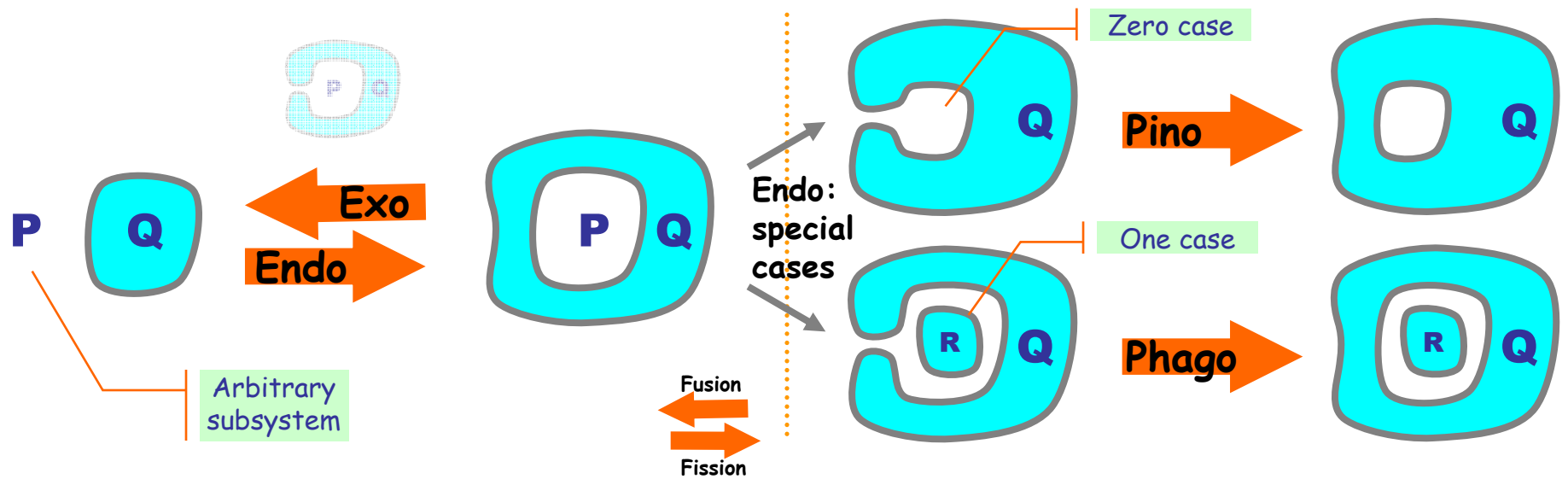
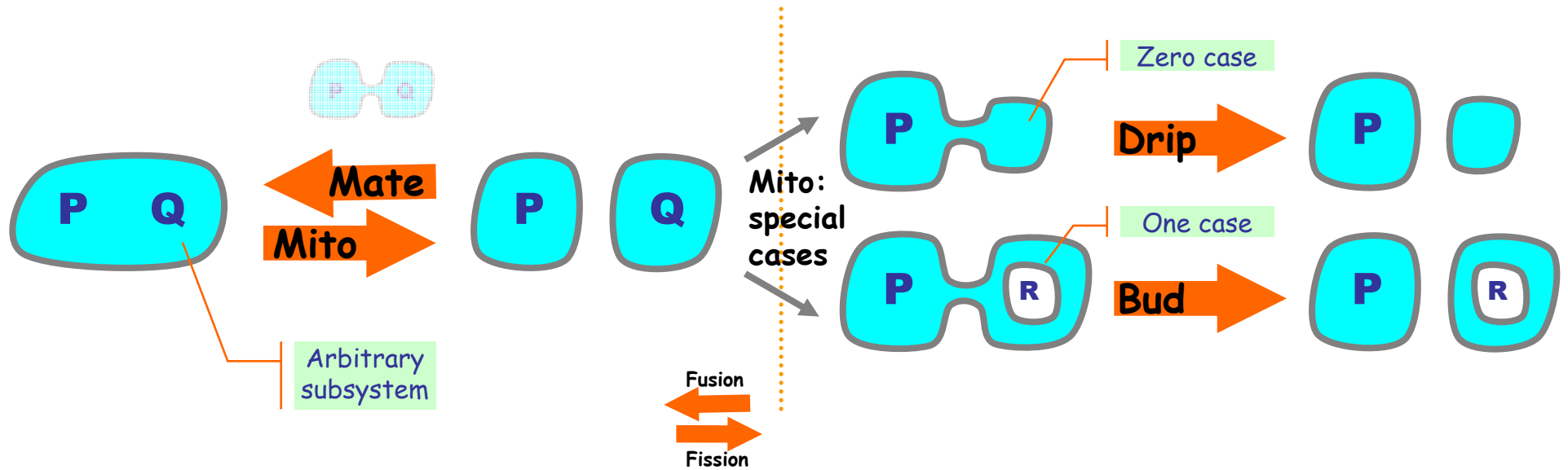


(Fusion)

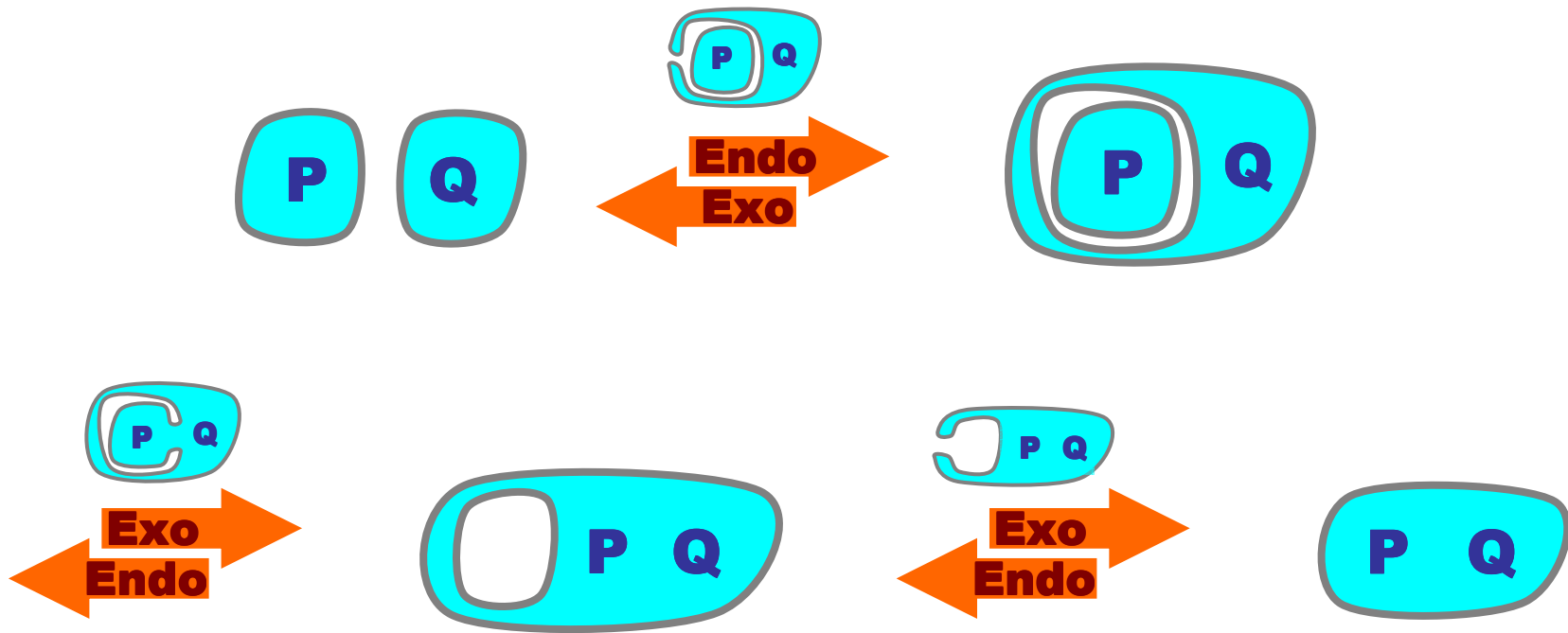


Same Local View!

# The Membrane Machine "Instruction Set"

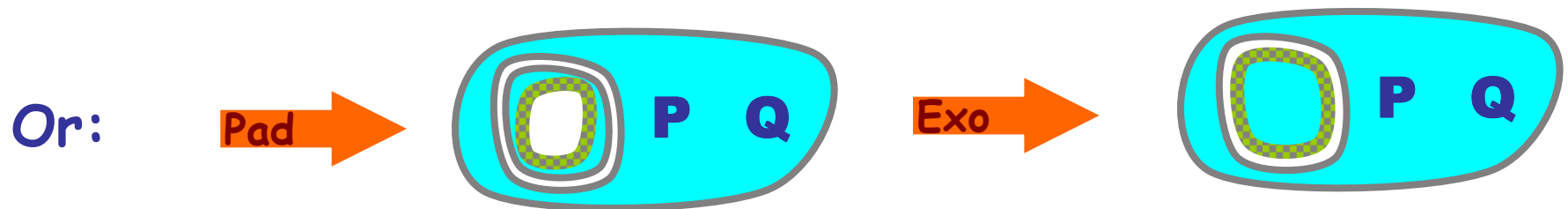
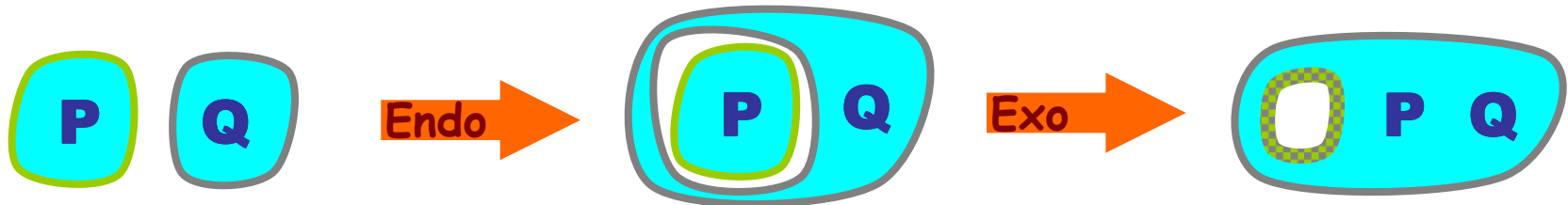
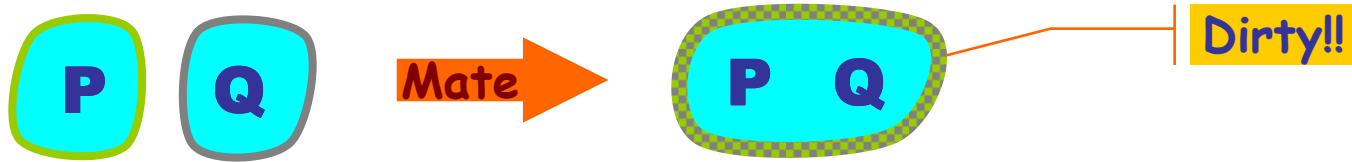


# Mito/Mate by 3 Endo/Exo



# (fake) Example: Clean Eating

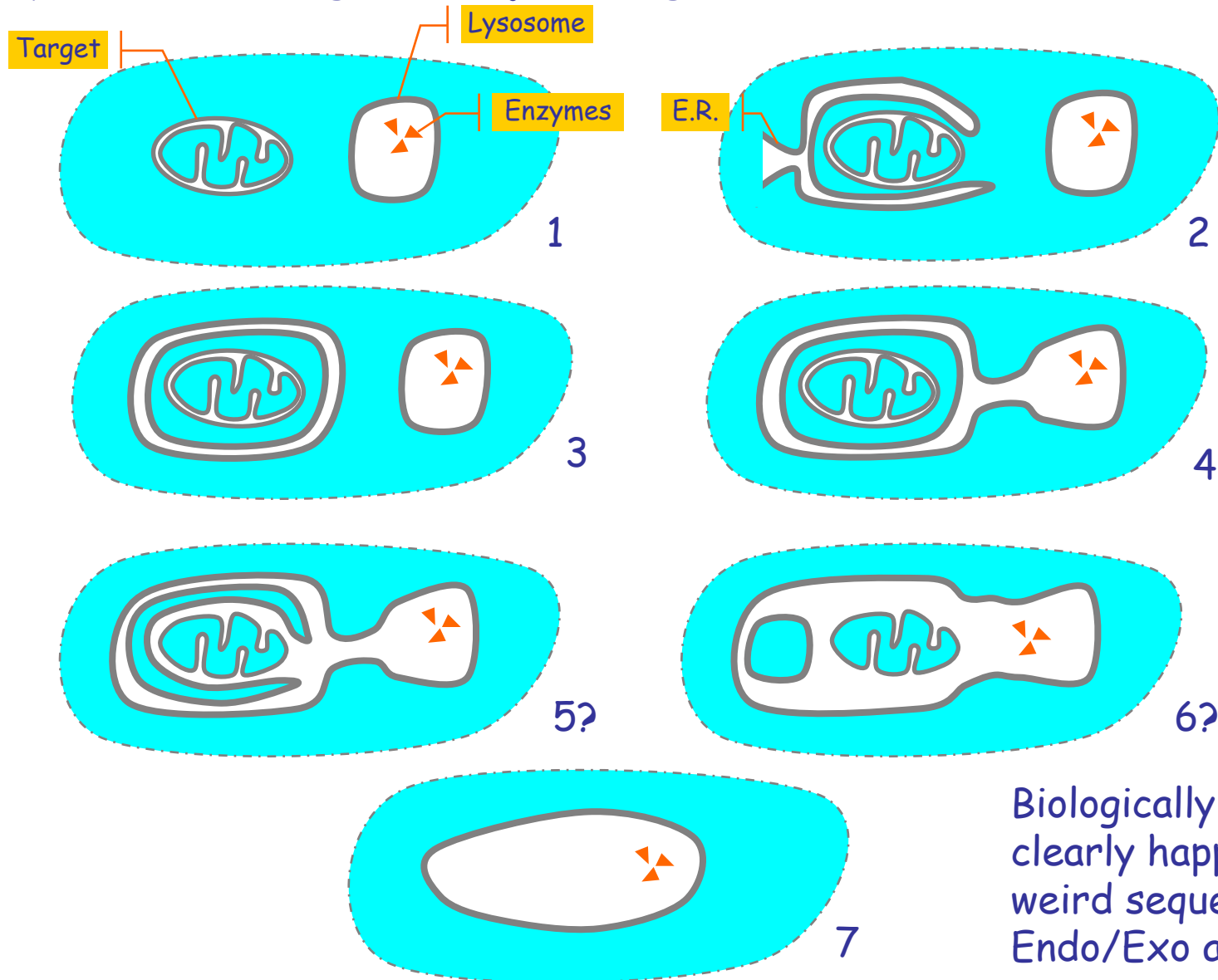
(why Endo/Exo is "healthier" than Mito/Mate)





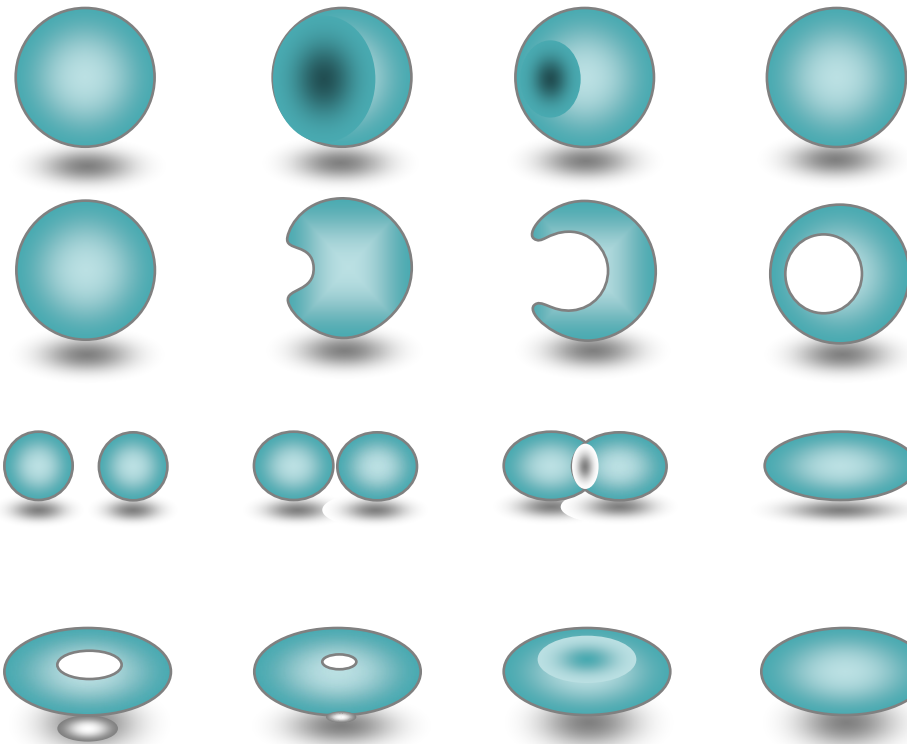
# Example: Autophagic Process

Lysosome and target don't just merge.



Biologically, Mito/Mate clearly happens. However, weird sequences of Endo/Exo are also common.

# ... in 3D



# A Membrane Algorithm

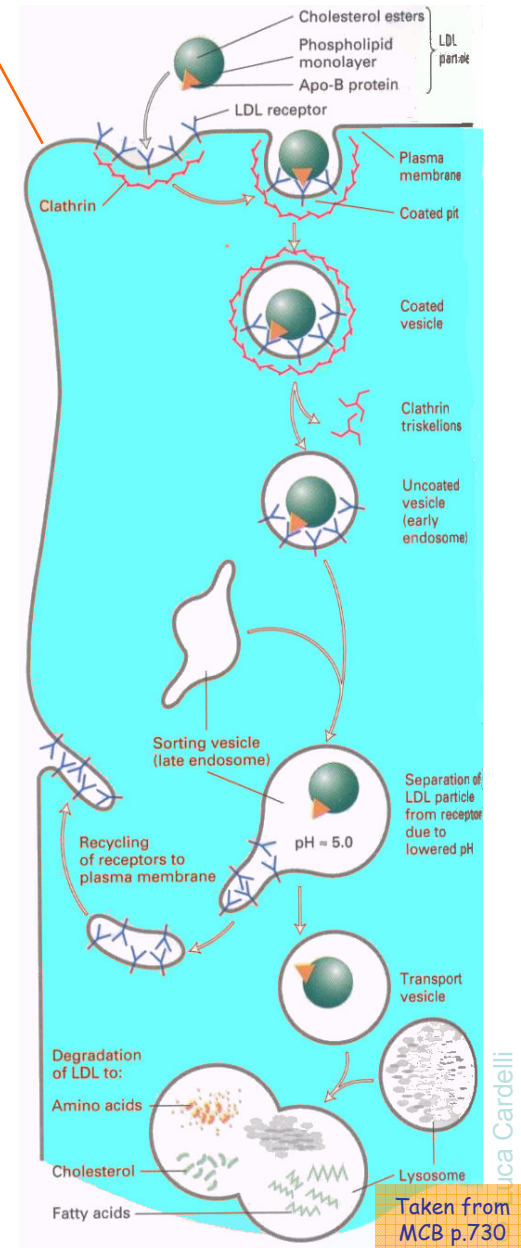
- **LDL-Cholesterol Degradation**

- 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**

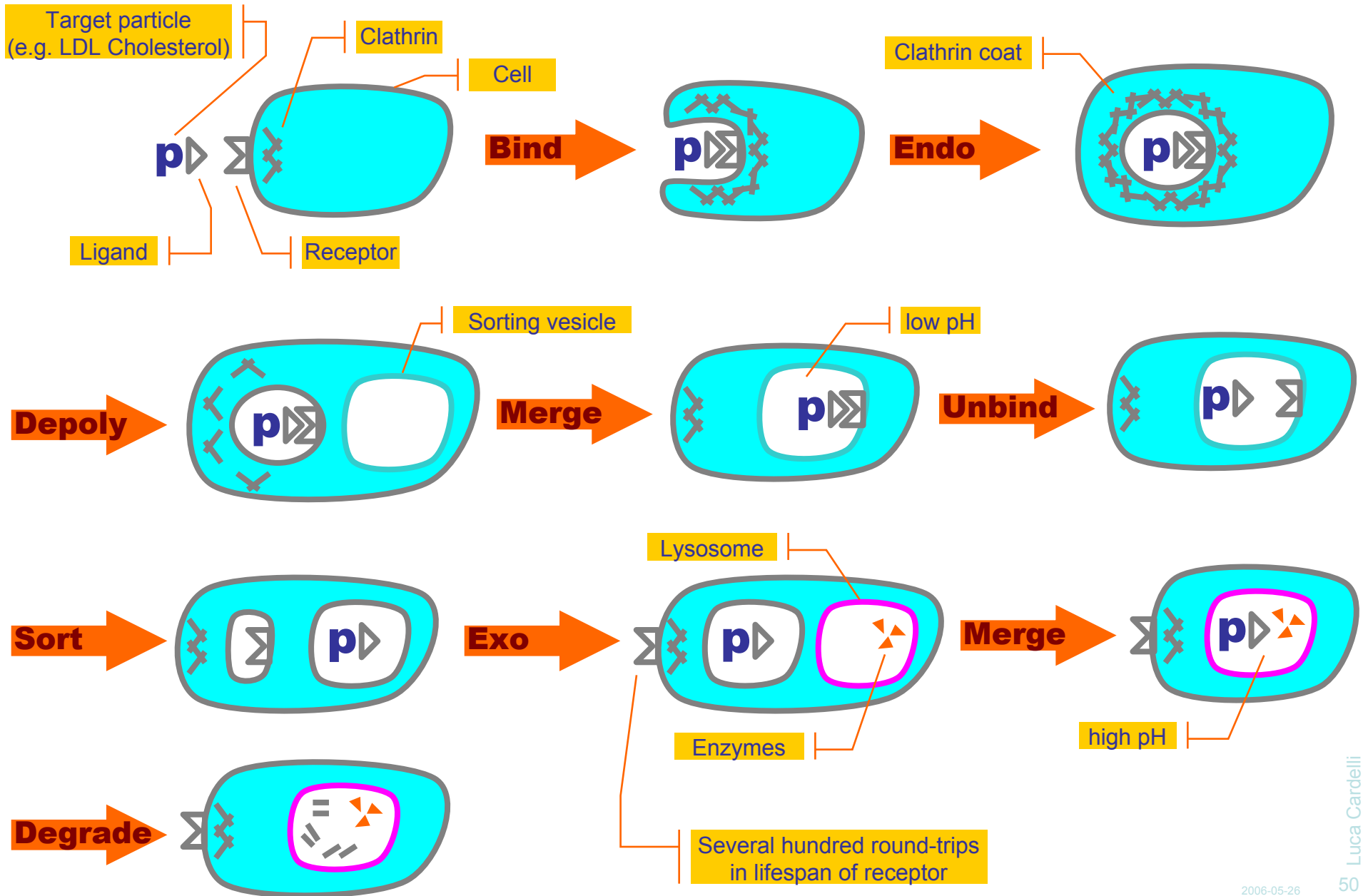
- But its "purpose" is to reliably implement a specific sequence of discrete steps.

Lipid bilayer



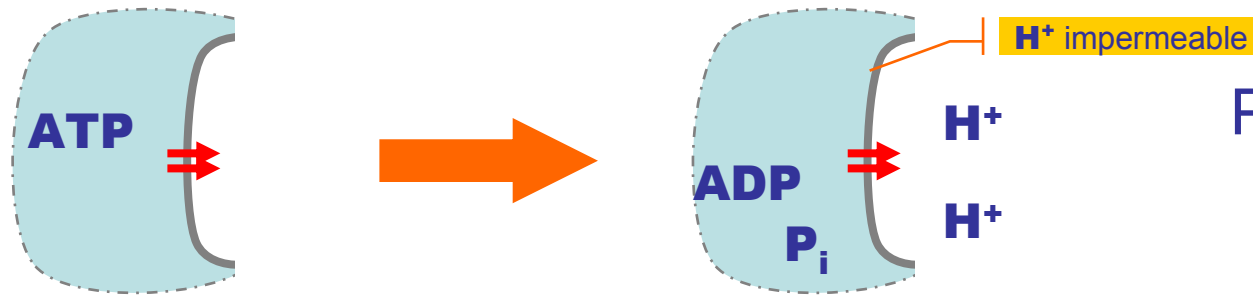
Taken from MCB p.730

# A more C.S.-style state transition diagram (Receptor-Mediate Degradation Pathway)



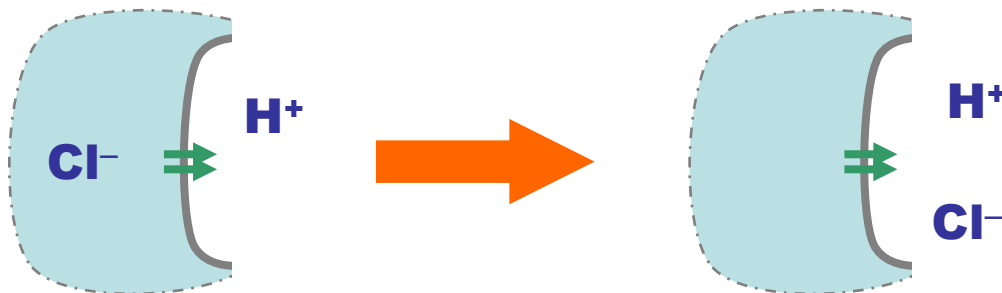
# Example: A Specialized Membrane

E.g. plant vacuole (white).

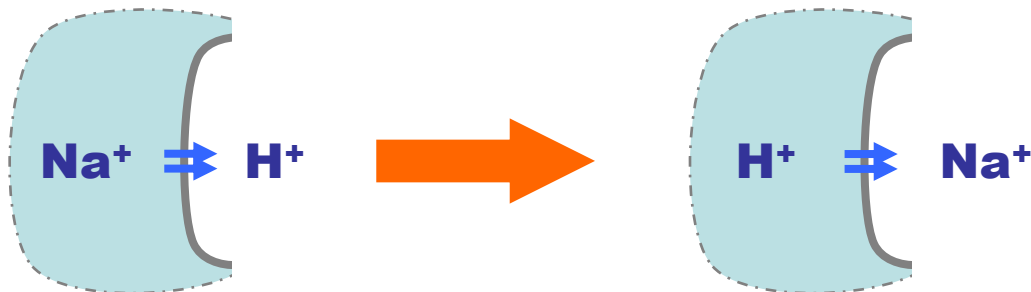


Proton Pump

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



Ion Channel



Proton Antiporter

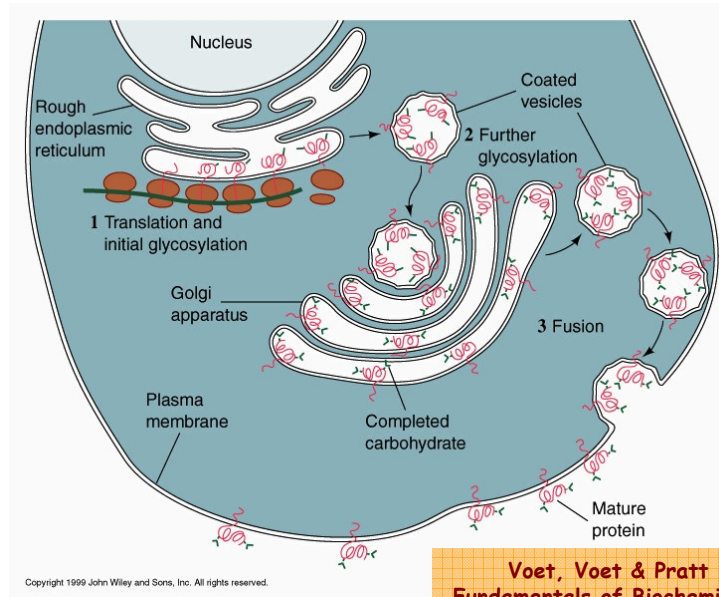
A plant vacuole has all those things on it, to accumulate NaCl.

2006-05-26



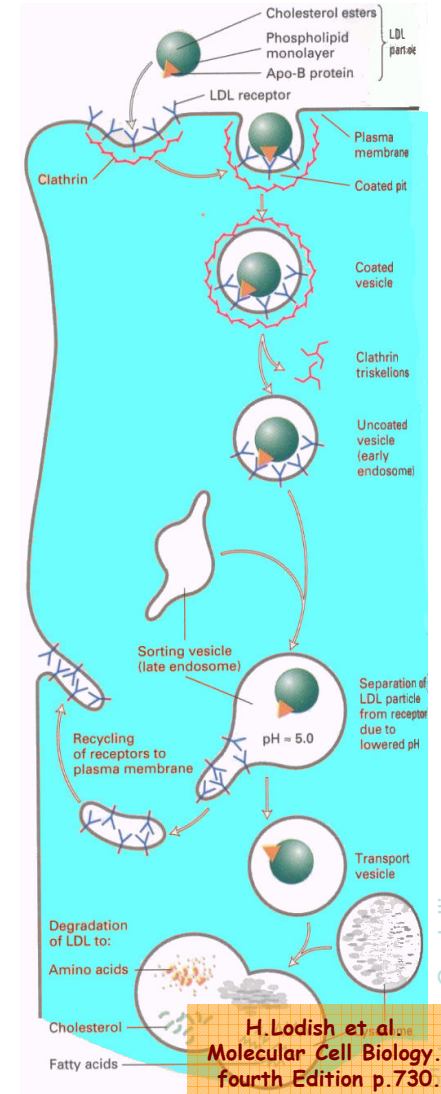
# Membrane Algorithms

## Protein Production and Secretion



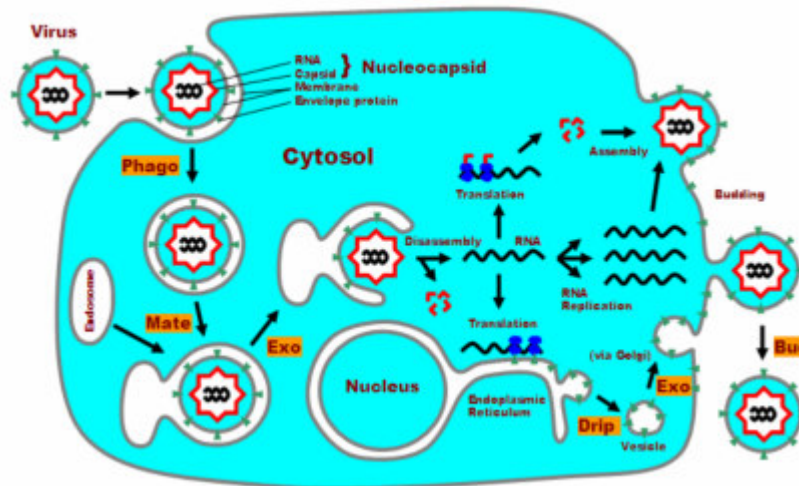
Voet, Voet & Pratt  
Fundamentals of Biochemistry  
Wiley 1999. Ch10 Fig 10-22.

## LDL-Cholesterol Degradation



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

## Viral Replication



Adapted from: B. Alberts et al.  
Molecular Biology of the Cell  
third edition p.279.

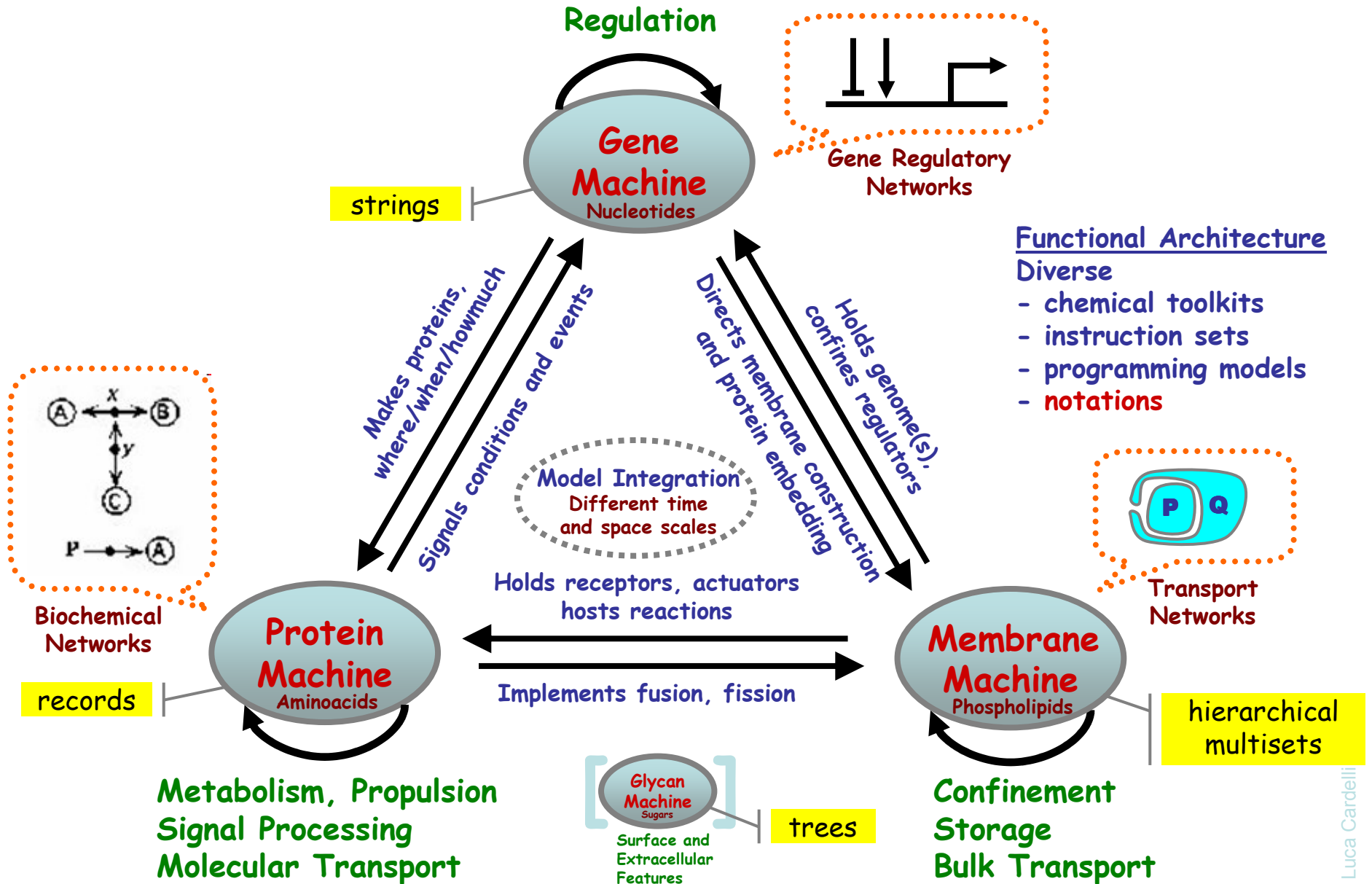
# What makes Endo happen?

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

# 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...

# Summary



Q?