Languages & Notations for Systems Biology

Luca Cardelli

luca@microsoft.com
www.luca.demon.co.uk

Microsoft Research

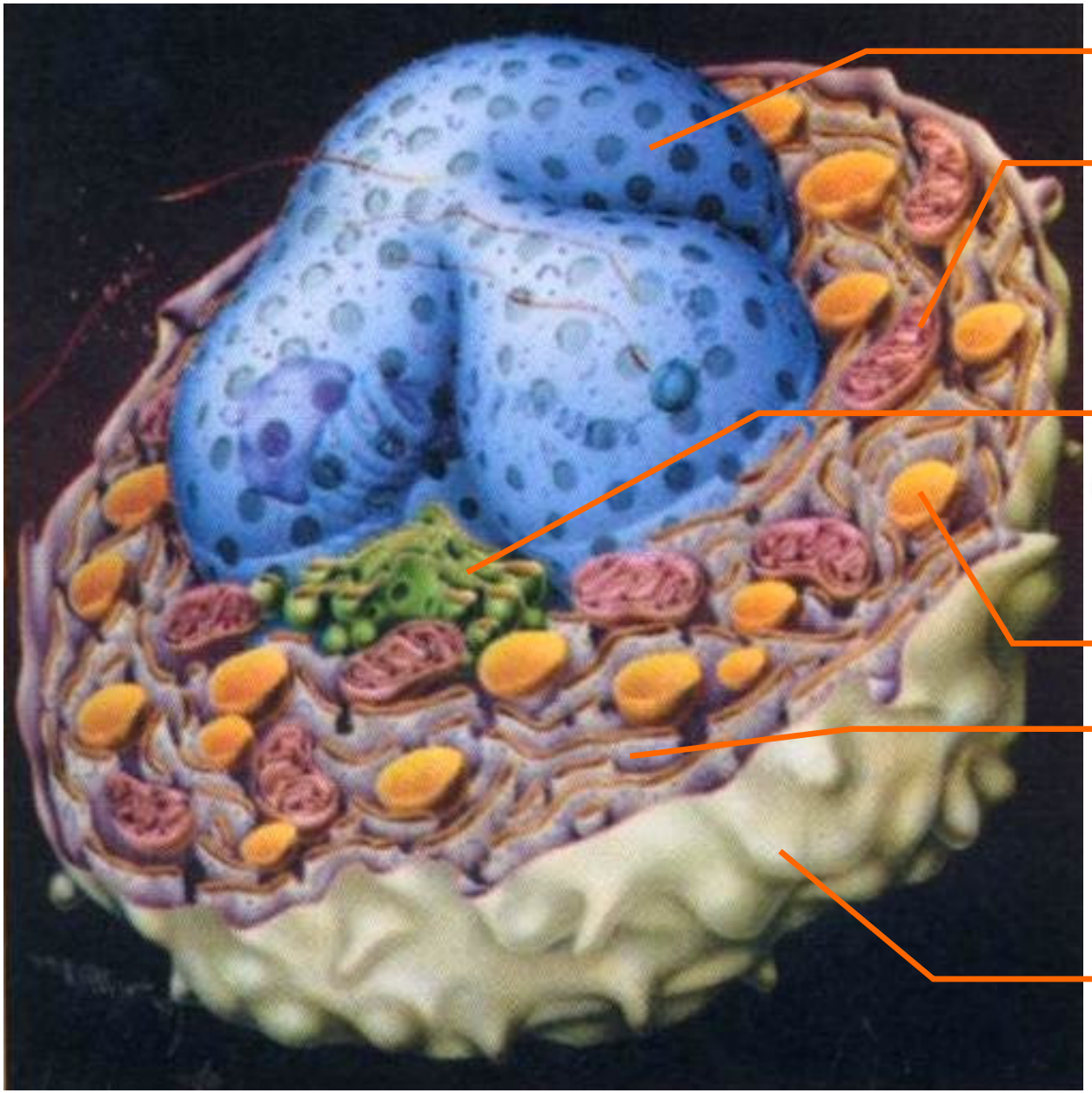
2004-05-12

Structural Architecture

Eukaryotic Cell

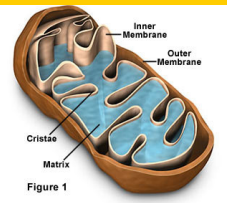
(10~100 trillion in human body)

Membranes everywhere

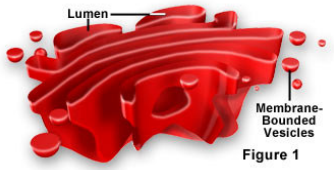


Nuclear membrane

Mitochondria

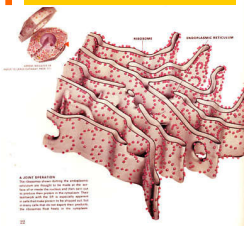


Golgi



Vesicles

E.R.



Plasma membrane (<10% of all membranes)

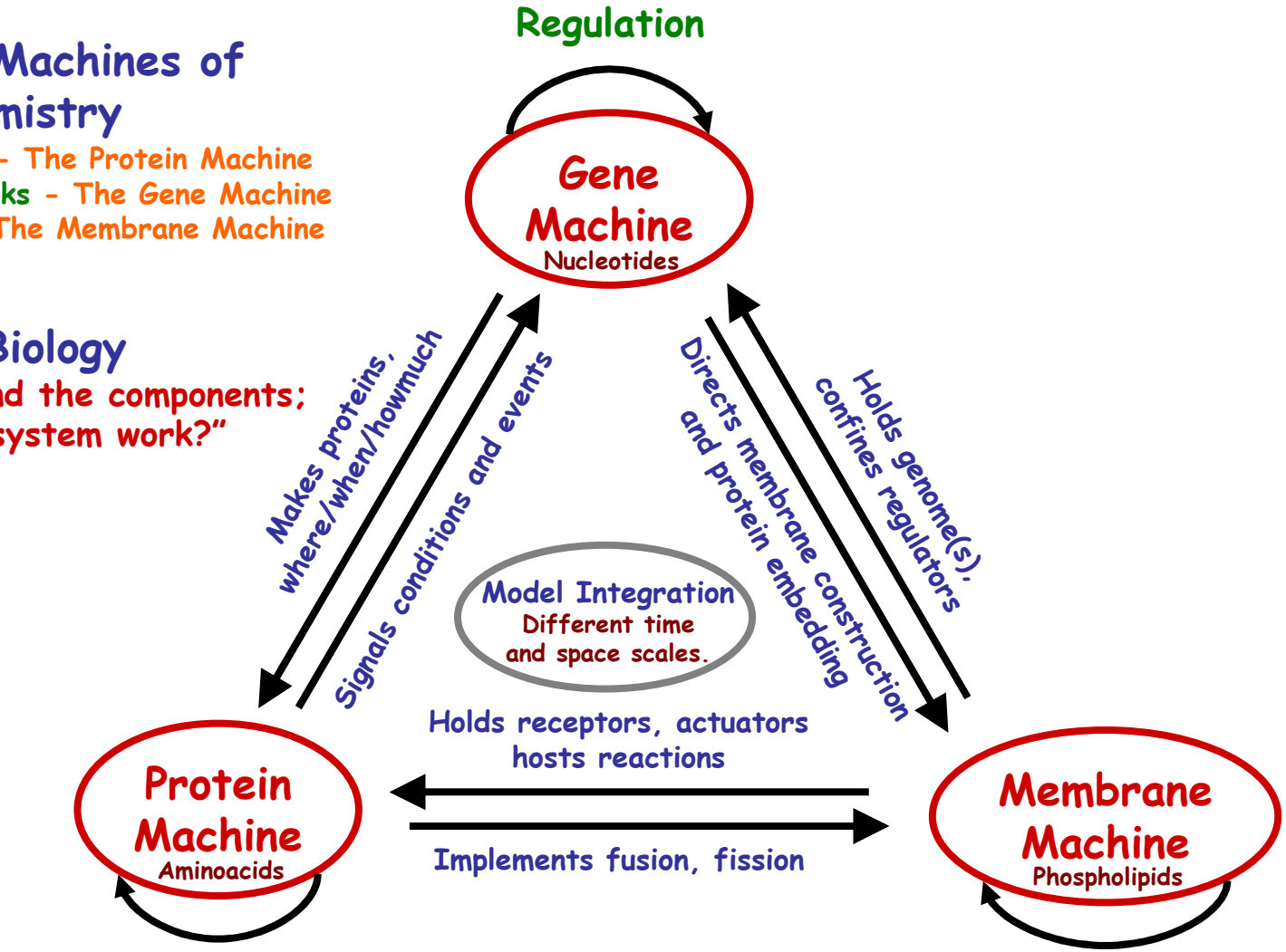
Functional Architecture

The 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?"



EU Commission, Health Research Report on Computational Systems Biology

- General Modelling Requirements
  - Research projects should focus on integrated modelling of several cellular processes leading to as complete an understanding as possible of the dynamic behaviour of a cell. Several projects may be required to develop modules (metabolism, signalling, trafficking, organelles, cell cycle, gene expression, replication, cytoskeleton) in model organisms. This modelling should involve realistic analysis of experimental data, including a wide range of data for transcriptomics, proteomics and functional genomics, and interactions with cellular pathways including signal transduction, regulatory cascades, metabolic pathways etc. It should involve:
    - Coherent, high-quality, quantitative, heterogeneous and dynamic data sets as a basis for novel model constructions to advance from analytical to predictive modelling.
    - Experimental functional analysis tools (in-situ proteomics, protein-protein interactions, metabolic fluxes, etc)

Challenges for Formal Notations in Biology

- Describe biological systems precisely
  - For analysis (discovering principles of operation)
  - For simulation (drug development, etc.)
  - For engineering (optimizing output, etc.)
- *New* working hypothesis:
  - Describe these complex deeply-layered systems as if they were software systems.
    - They are based on digitally-coded information.
    - They are largely information processing systems.
    - Hence code them up in some analyzable language or notation.
  - Claim (to be validated):
    - modularity and compositionality advantages, just as in software, for scaling-up, w.r.t. traditional methods (chemical equations, differential equations).

Biochemical Process Notations

- Chemical reactions *is* a process calculus!
  - A long, long, flat list of thousands of reactions… highly concurrent and nondeterministic.
  - But there is also structure and modularity in biochemistry.
- Representing structure
  - Process calculi are the modular representation of discrete concurrent processes.
  - They can be seen as an input language for Petri Nets or for Continuous Time Markov Chains.
- Just like a sequence of assignments and goto's *is* a programming language.
  - There are better (yes?) programming languages.
  - But no ordinary programming language has that level of concurrency and nondeterminism.
- Let's take a look at the high-level process notations of biochemistry (mostly diagrams and pictures) …

Protein Machine
Aminoacids

Molecular Interaction Maps

Pretty close to the atoms

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

JDesigner

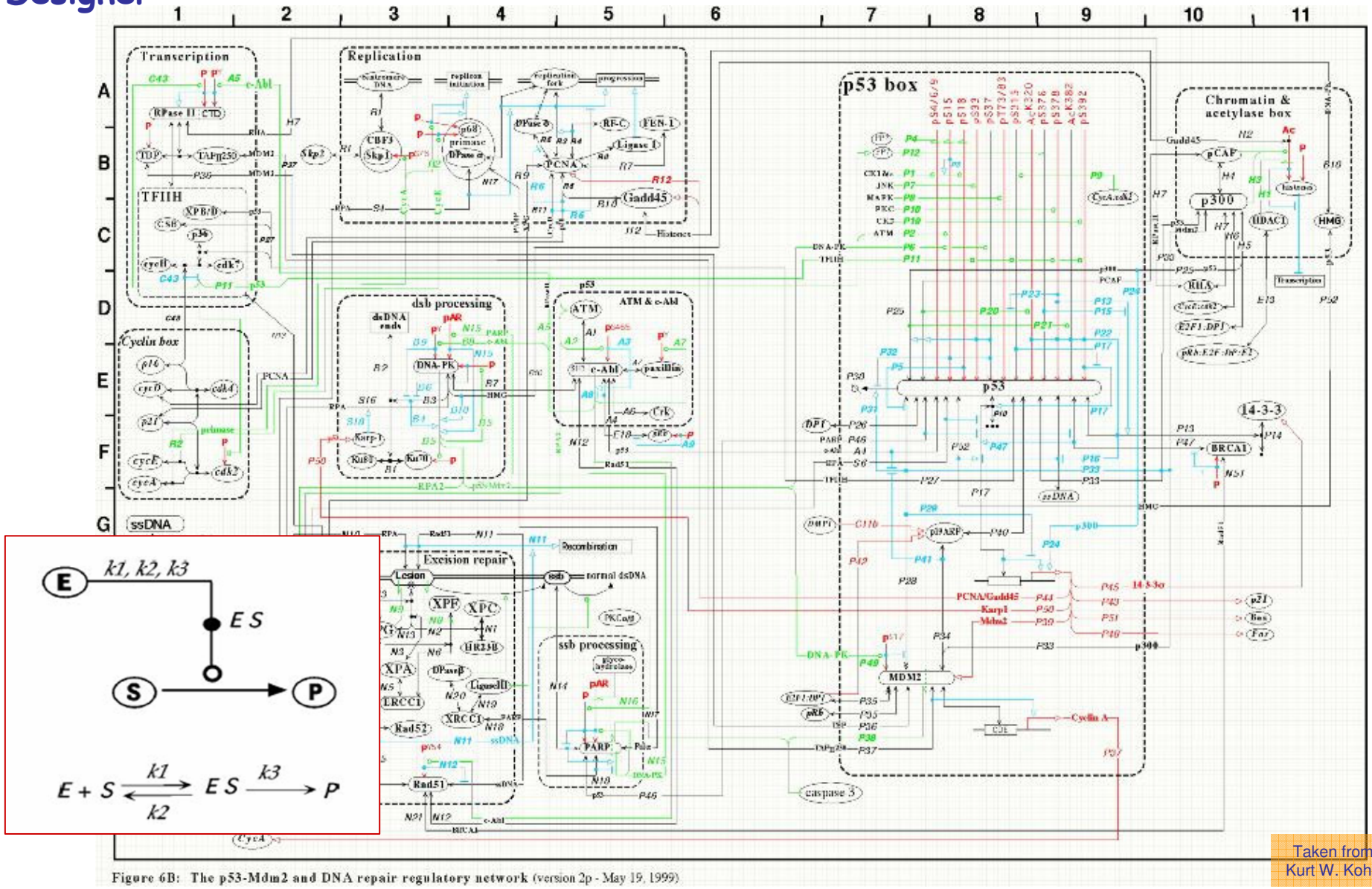


Figure 6B: The p53-Mdm2 and DNA repair regulatory network (version 2p - May 19, 1999)

Taken from Kurt W. Kohn

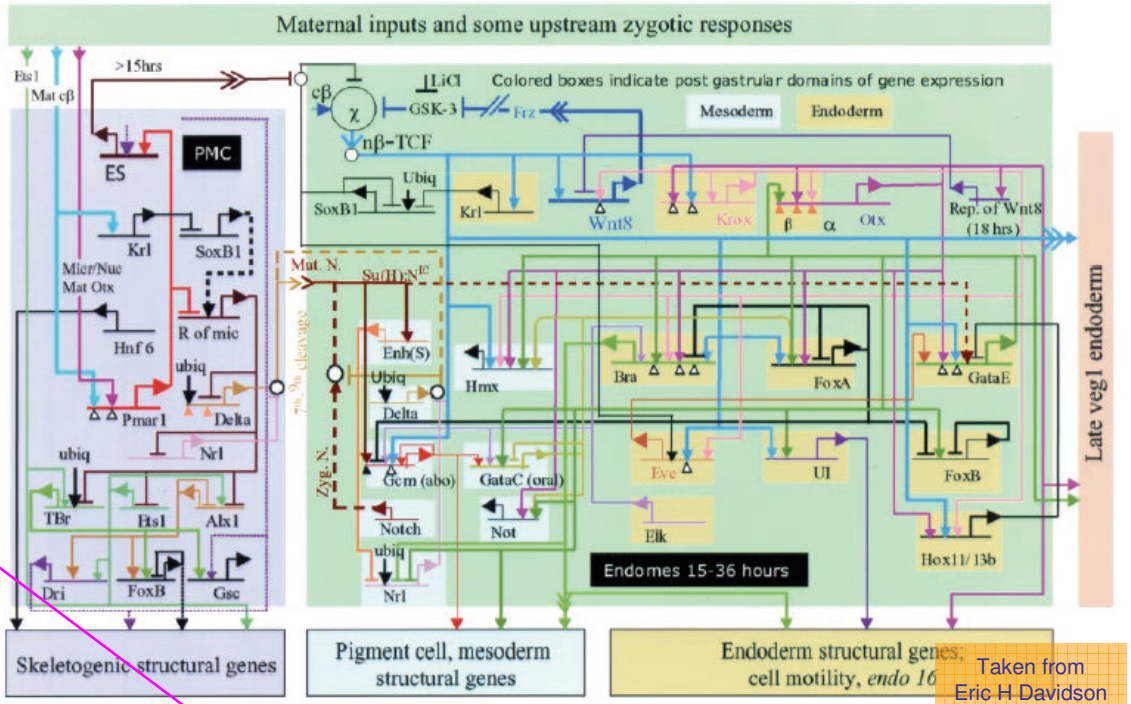
Gene Machine
Nucleotides

Gene Regulatory Networks

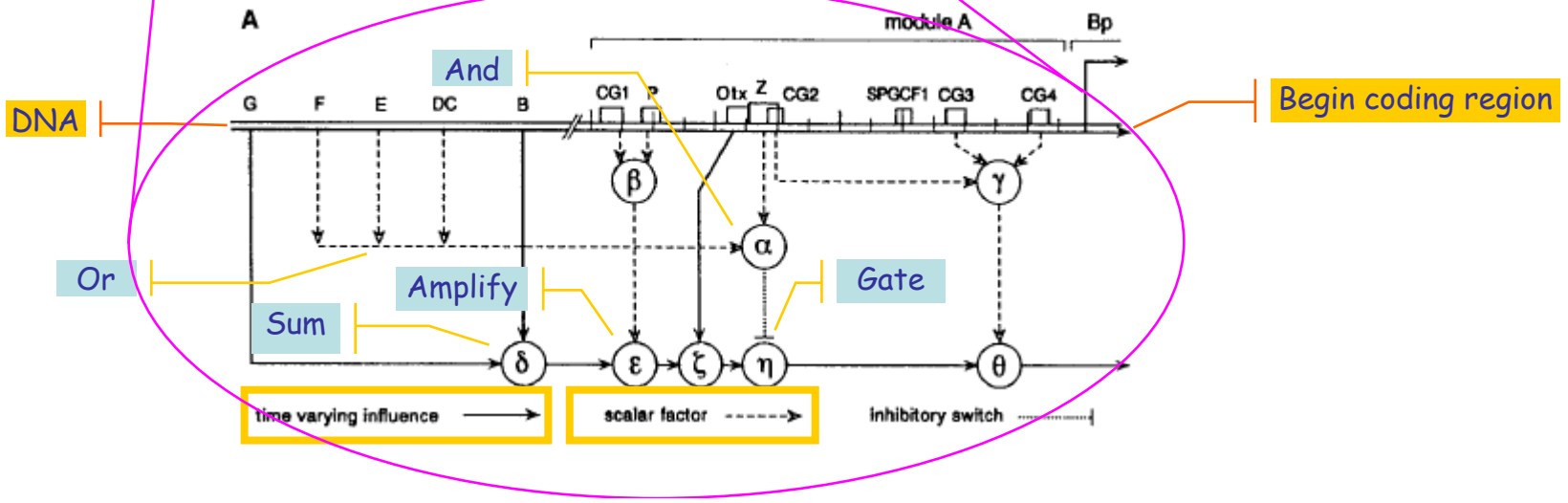
Pretty far from the atoms

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

NetBuilder



Taken from Eric H Davidson

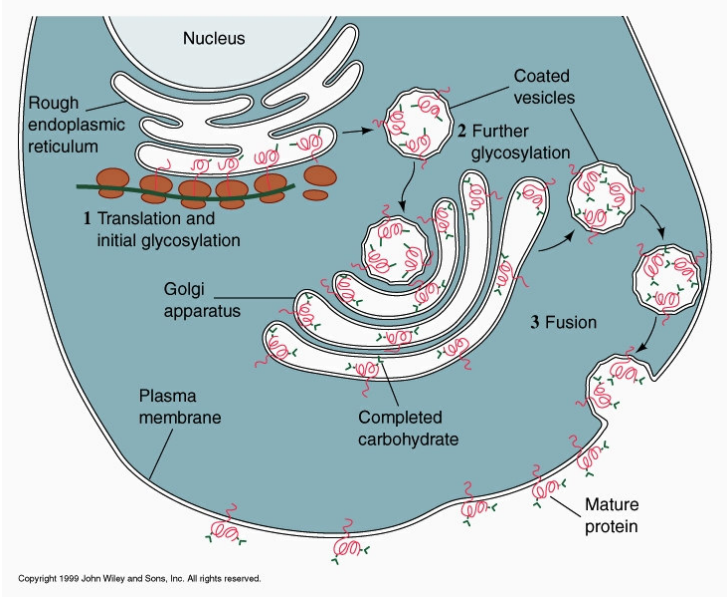


Membrane Machine
Phospholipids

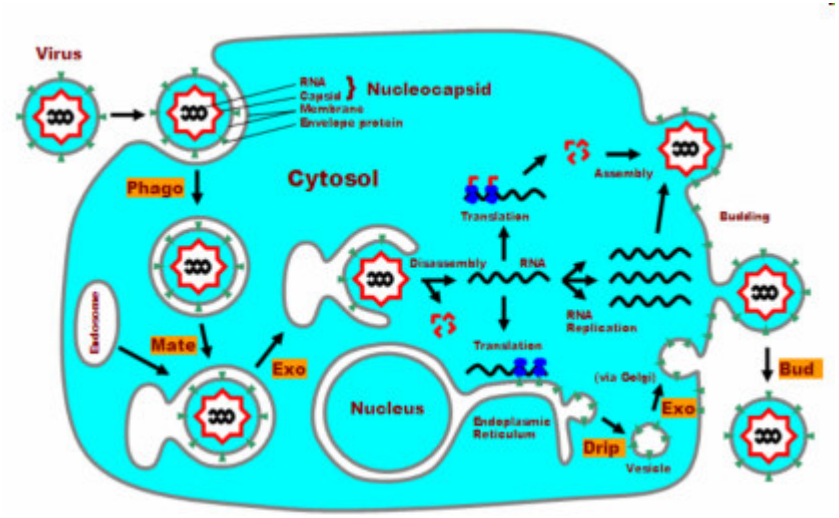
Transport Diagrams

Very far from the atoms.

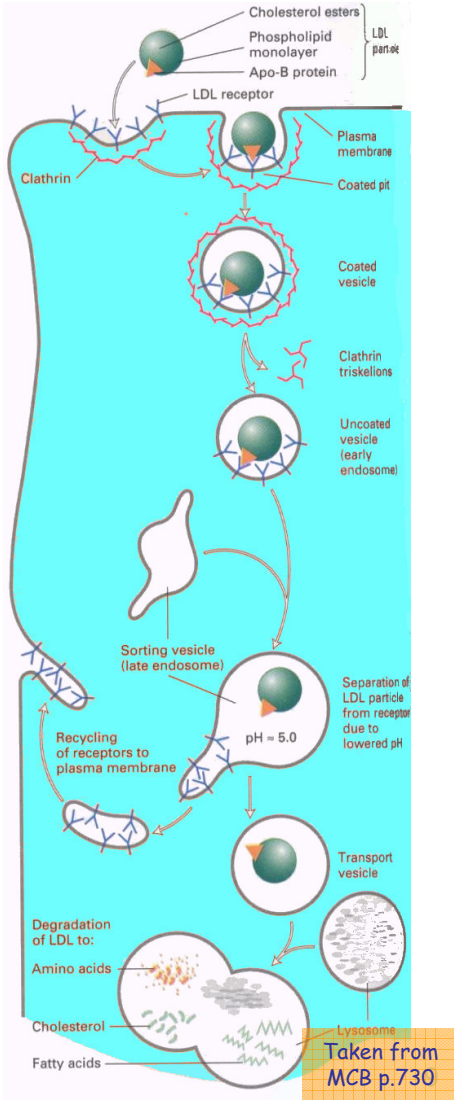
Protein Production and Secretion



Viral Replication



LDL-Cholesterol Degradation



Taken from MCB p.730

Promising Techniques and Technologies

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.
  - Stochastic Highwire? [Merdith]
  - 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.

"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.

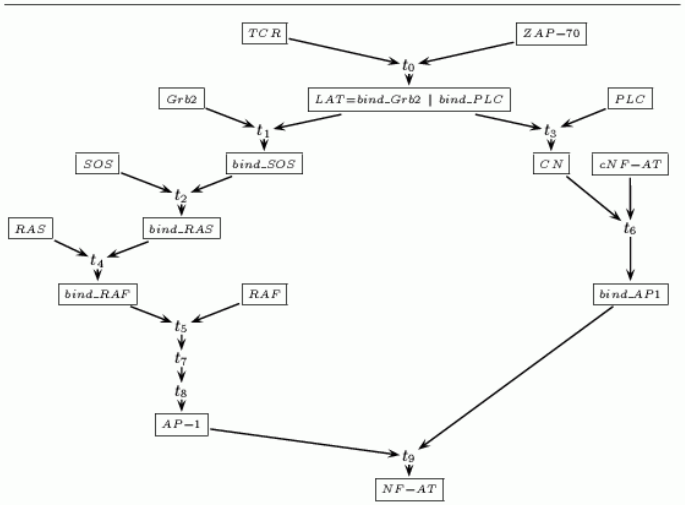


Fig. 2. A computation of *Sys*. For readability, the processes, enclosed in boxes, have no address. Causality (both on transitions and processes) is represented by the (Hasse diagram resulting from the) arrows; their absence makes it explicit concurrent activities.

- 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].

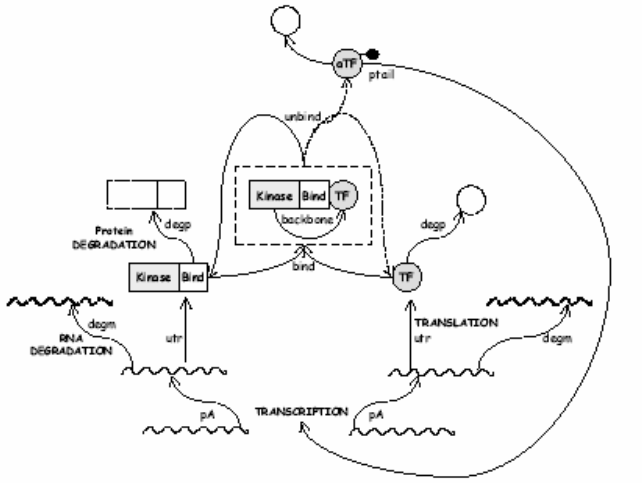


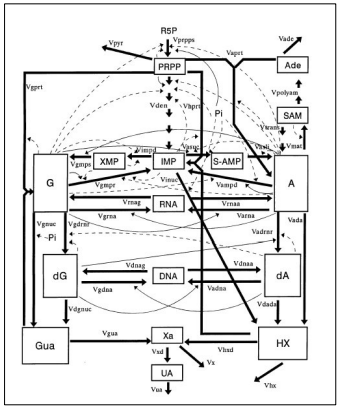
Fig. 1. Graphical presentation of Transcriptional Regulation by Positive Feedback [25].

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)))
```

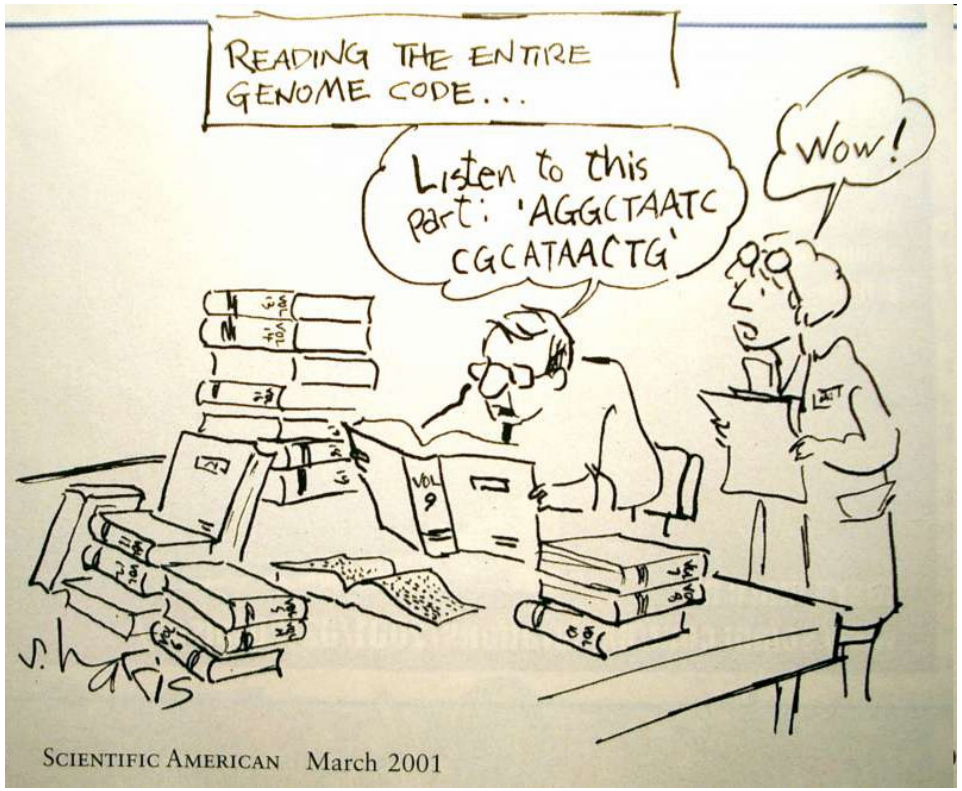


- Stochastic: Spring [Parker Normal Kwiatkowska]
  - Designed for stochastic (computer) network analysis
    - Discrete and Continuous Markov Processes.
    - Process input language.
    - Modelchecking of probabilistic queries.

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 abstraction levels and behaviors.
  - We can use equivalences for state minimization (symmetries).
- Disclaimers
  - Some of these technologies are basically ready (small-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.

END



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

"Although the road ahead is long and winding, it leads to a future where biology and medicine are transformed into precision engineering." - Hiroaki Kitano.

Process Calculi

- Unfortunately, there are *many* process calculi.
  - There are suitable general theories (Milner's BiGraphs) where we can hope to achieve at least partial *model integration*.
  - π-calculus is "canonical": has most interesting things in it (composition, interaction, and hiding), but not all.
  - There is a set of standard techniques (transition systems, equivalences, etc.) to build and study new calculi "on demand".
- Fortunately, there are *many* process calculi.
  - Some are better for modeling software or hardware.
  - We can look for the ones that best model biological processes.
  - Many kinds of processes = many kinds of calculi.

Composition(ality)	P | Q
Complexation (new!)	P:Q
Localization	[P]
Interaction	a.P
Hiding	(νn)P

In Their Own Words…

- On the nature of modeling
  - Sydney Brenner: *"When you want to have a predictive science, you have to be able to calculate."*
  - Hamid Bolouri & Eric H. Davidson: *"Abstract models have relatively few parameters and so … it is easier to explore their behavior and build models with them. … In contrast, more detailed models suffer from an explosion in the number of their parameters."*
  - Denis Noble: *"There will probably therefore be no unique model that does everything at all levels. … One of the first questions to ask of a model therefore is what questions does it answer best."*
  - Hiroaki Kitano: *"Molecular biology has uncovered a multitude of biological facts … but this alone is not sufficient for interpreting biological systems. … A system-level understanding should be the prime goal of biology."*
  - Al Hershey: *"Influential ideas are always simple. Since natural phenomena need not be simple, we master them, if at all, by formulating simple ideas and exploring their limitations."*