Systems Biology: Models and Logics I

From Experiments to Models

Carla Piazza

Dipartimento di Matematica ed Informatica
Università di Udine

What is Systems Biology?

H. Kitano – Science 2002

System-level understanding, the approach advocated in systems biology, requires a shift in our notion of "what to look for" in biology. While an understanding of genes and proteins continues to be important, the focus is on understanding a system's structure and dynamics.

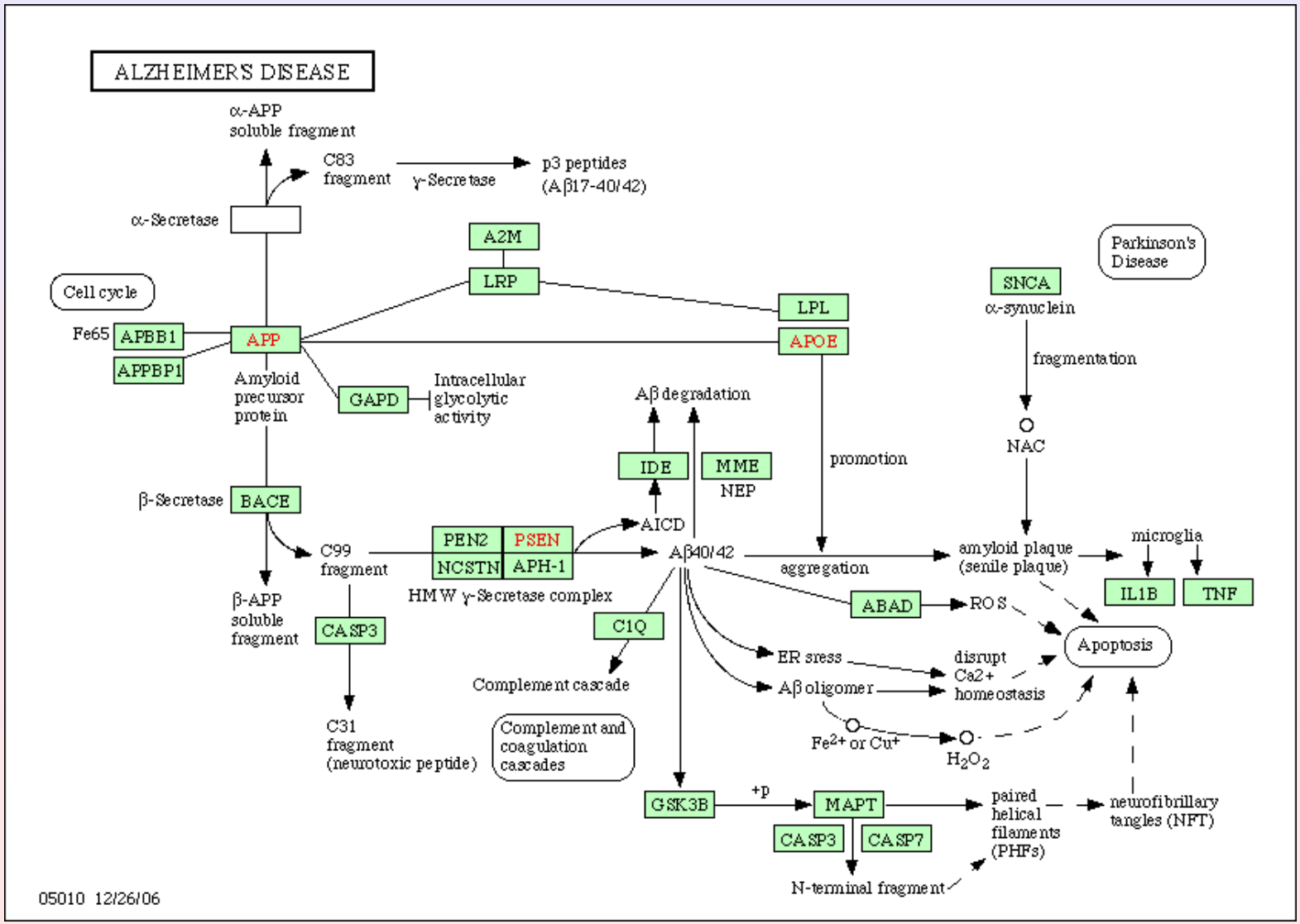
What can we do for Systems Biology?

P. Nurse – Nature 2003

An important part of the search for such explanations is the identification, characterization and classification of the logical and informational modules that operate in cells. For example, the types of modules that may be involved in the dynamics of intracellular communication include feedback loops, switches, timers, oscillators and amplifiers. Many of these could be similar in formal structure to those already studied in the development of machine theory, computing and electronic circuitry.

Outline

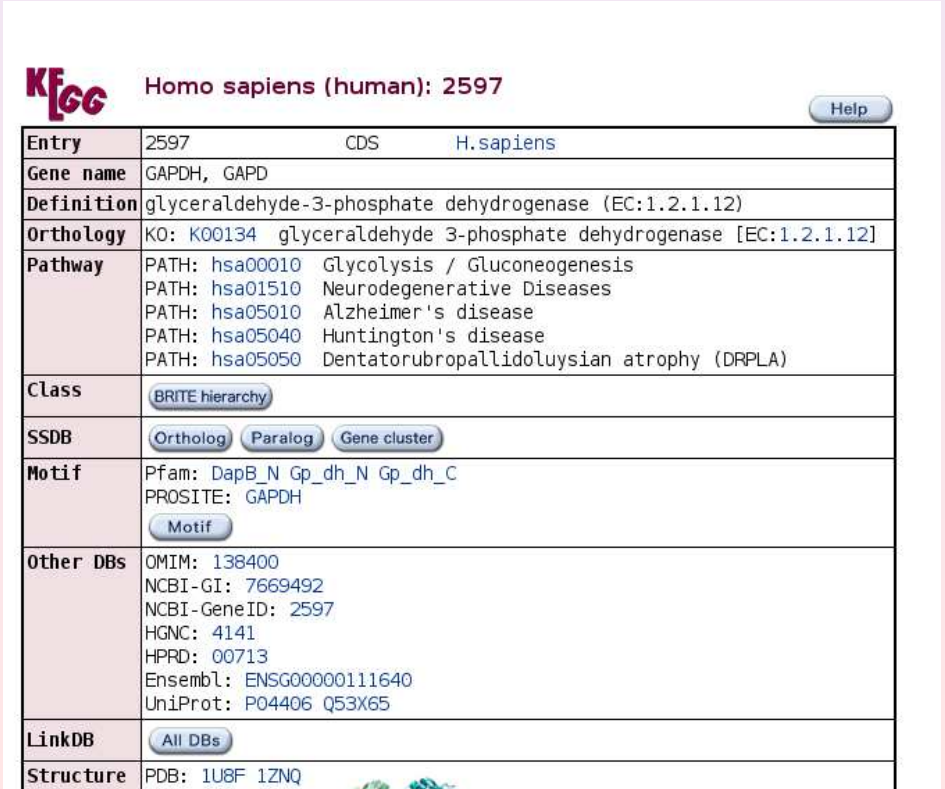
Input: Pathways



Pathways and DataBases

KEGG – Kyoto Encyclopedia of Genes and Genomes

KEGG PATHWAY is a collection of manually drawn pathway maps representing our knowledge on the molecular interaction and reaction networks.



KEGG Homo sapiens (human): 2597

Help

Entry	2597 CDS H.sapiens
Gene name	GAPDH, GAPD
Definition	glyceraldehyde-3-phosphate dehydrogenase (EC:1.2.1.12)
Orthology	KO: K00134 glyceraldehyde 3-phosphate dehydrogenase [EC:1.2.1.12]
Pathway	PATH: hsa00010 Glycolysis / Gluconeogenesis PATH: hsa01510 Neurodegenerative Diseases PATH: hsa05010 Alzheimer's disease PATH: hsa05040 Huntington's disease PATH: hsa05050 Dentatorubropallidoluysian atrophy (DRPLA)
Class	BRITE hierarchy
SSDB	Ortholog Paralog Gene cluster
Motif	Pfam: DapB_N Gp_dh_N Gp_dh_C PROSITE: GAPDH Motif
Other DBs	OMIM: 138400 NCBI-GI: 7669492 NCBI-GeneID: 2597 HGNC: 4141 HPRD: 00713 Ensembl: ENSG00000111640 UniProt: P04406 Q53X65
LinkDB	All DBs
Structure	PDB: 1U8F 1ZNQ

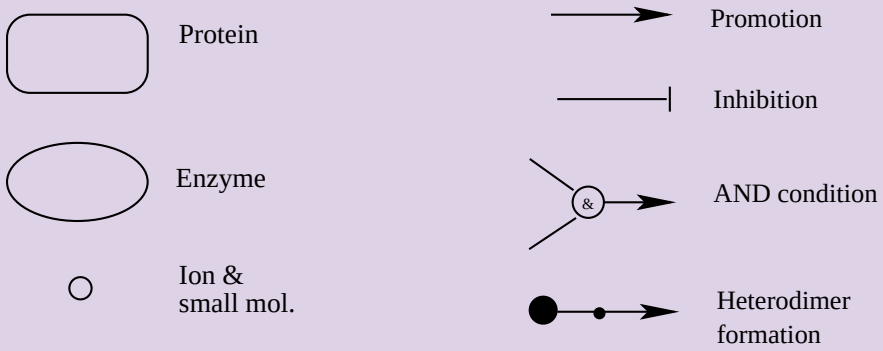
H. Kitano. *A graphical notation for biochemical networks.* BIOSILICO 2003

This is at the basis of standard graphical notation for Systems Biology Mark-up Language (SBML) Level-III:

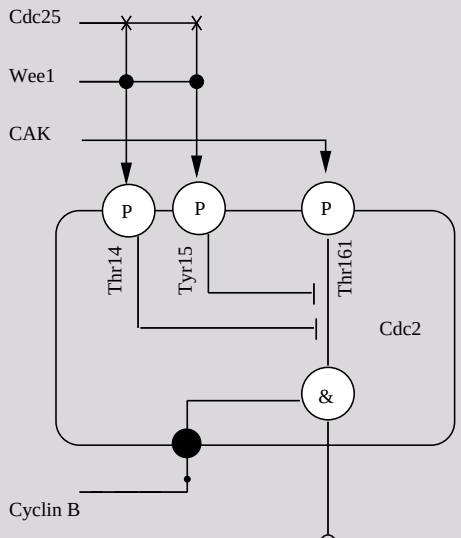
- State Transition Diagrams. Represent the evolution of the molecules during the reaction. They are graphs in which each node is a state of a component and each edge represents a transition between states.
- Block Diagrams. Represent the relationships among molecular species. Each molecule is represented only once. Each node is a molecule and each edge represents the interaction between two or more nodes.
- Flow Charts. Describe the evolution of the biological events. They abstract state-transition diagrams.

Block Diagrams

Building Blocks

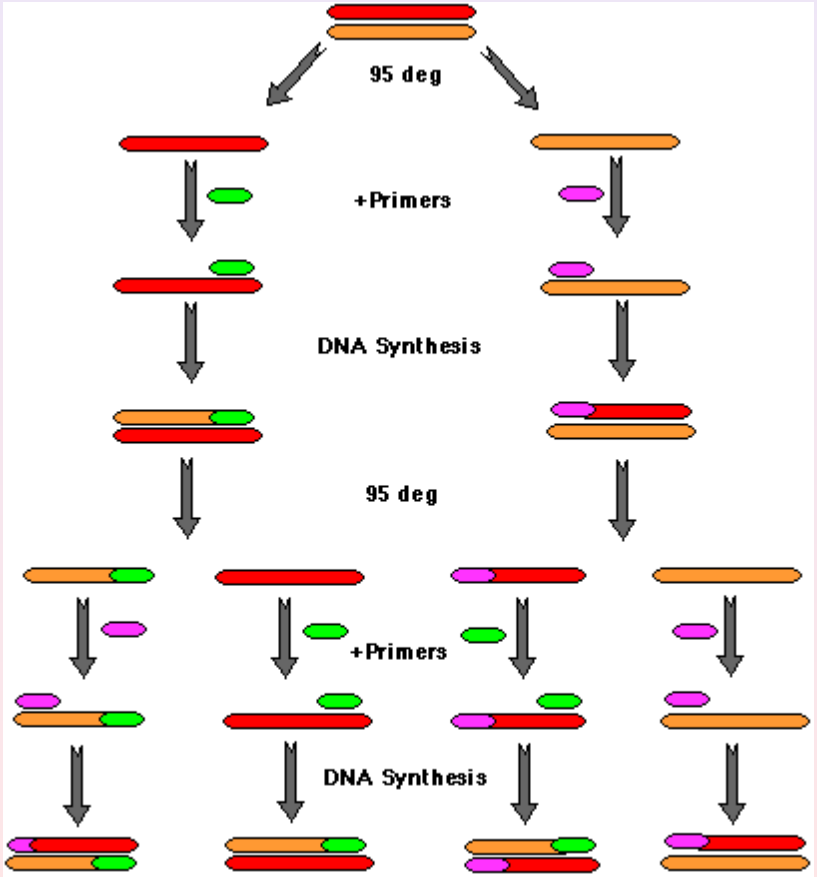


Example



Input: (1) Polymerase Chain Reaction

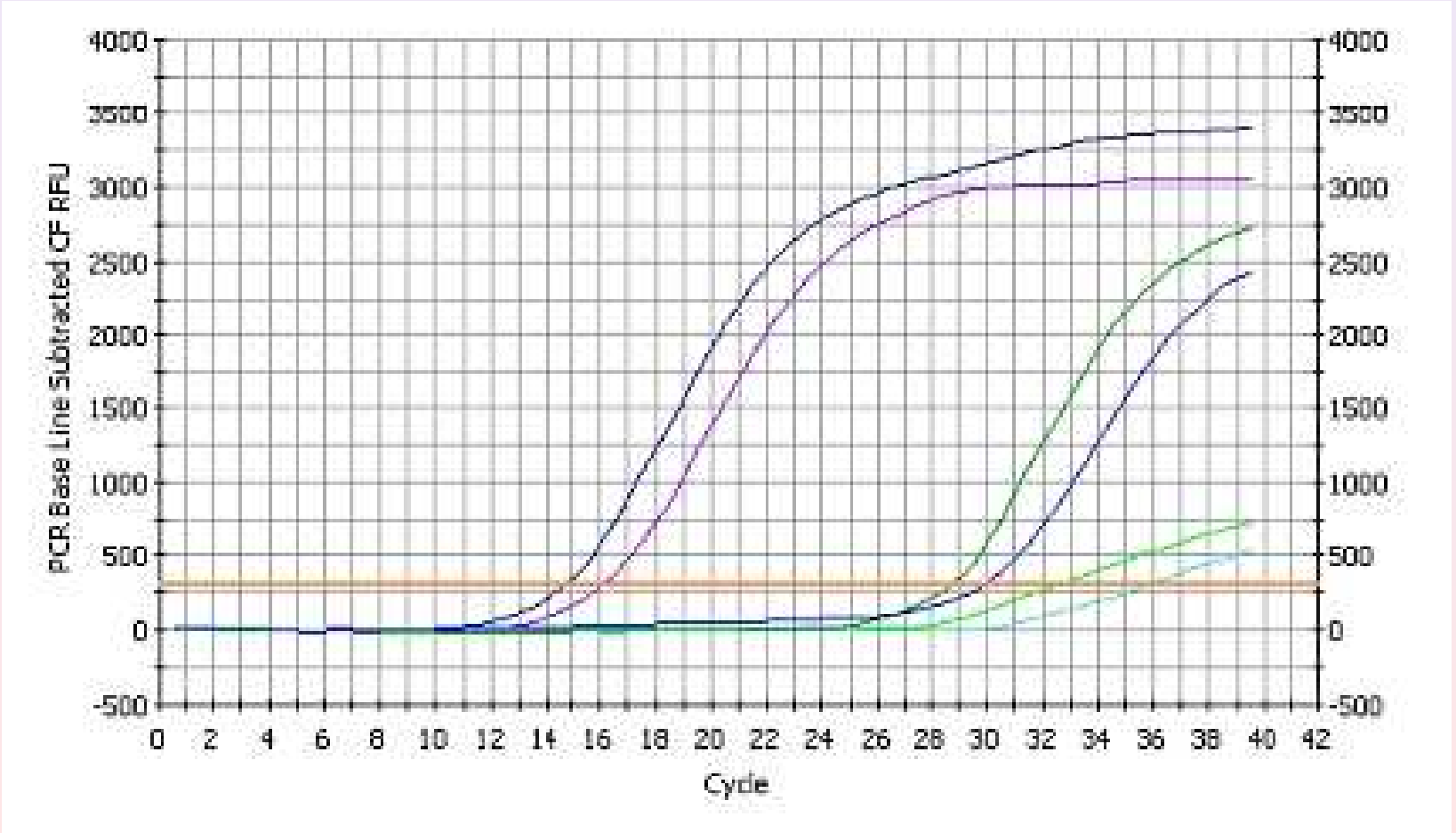
Polymerase chain reaction (PCR) is a technique for exponentially amplifying a fragment of DNA (RNA).



Multiplying a DNA template with PCR

Input: (1) Polymerase Chain Reaction

Polymerase chain reaction (PCR) is a technique for exponentially amplifying a fragment of DNA (RNA).

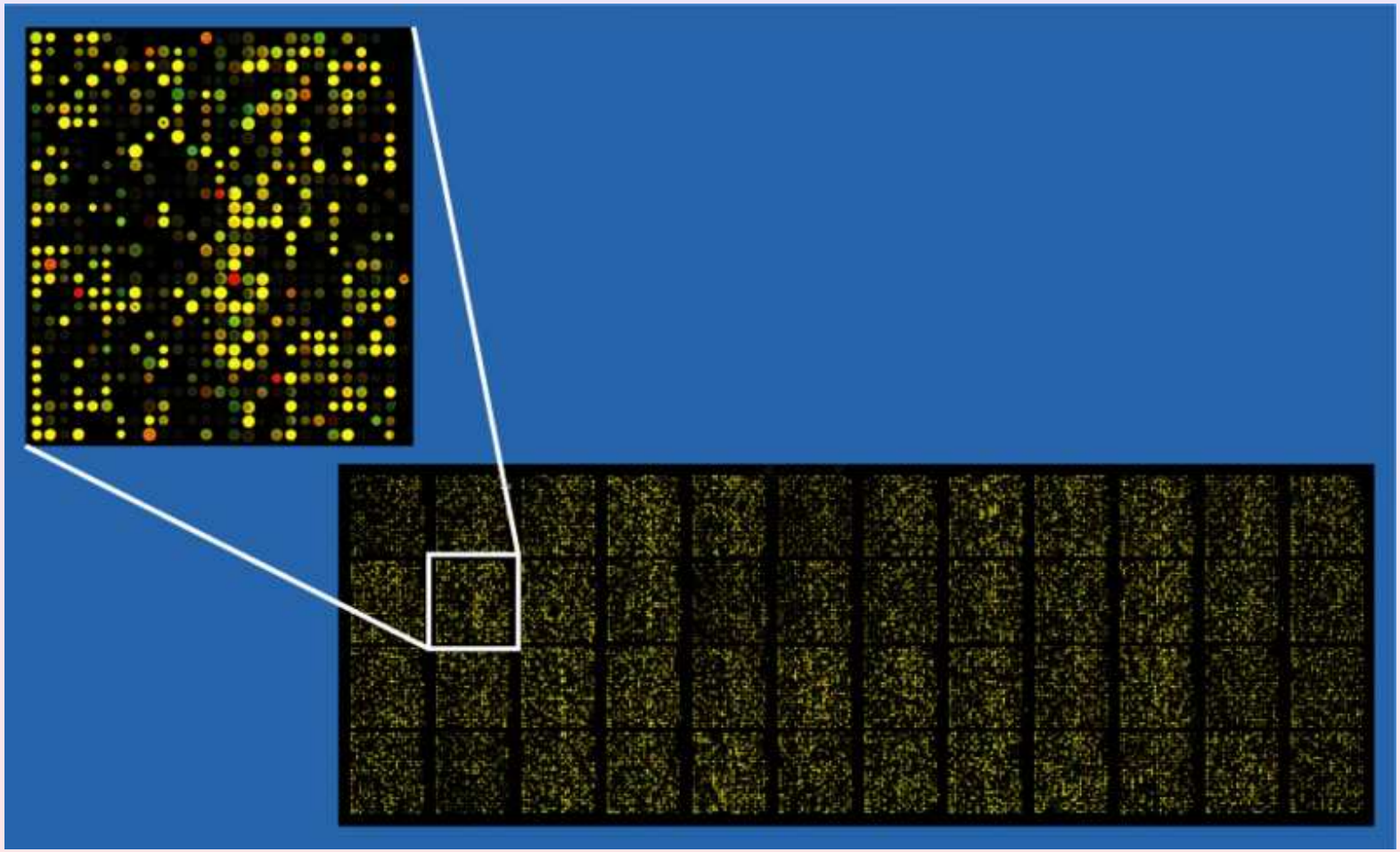


Input: (1) Polymerase Chain Reaction

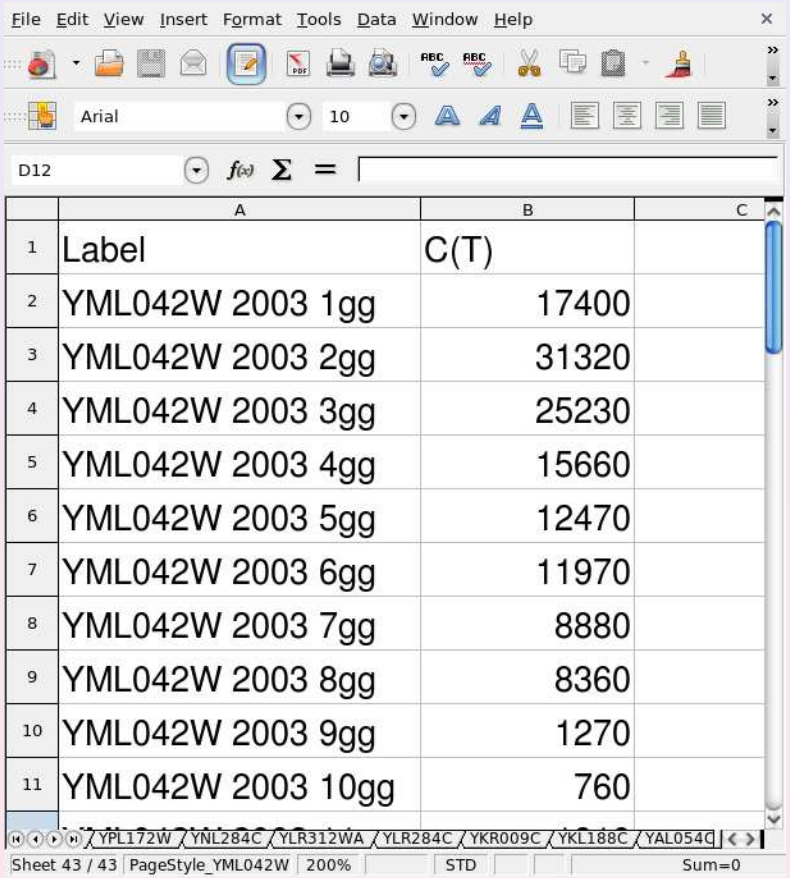
Polymerase chain reaction (PCR) is a technique for exponentially amplifying a fragment of DNA (RNA).

It can be used for Gene Profiling.

Input: (2) Microarrays



PCR and Microarray Data

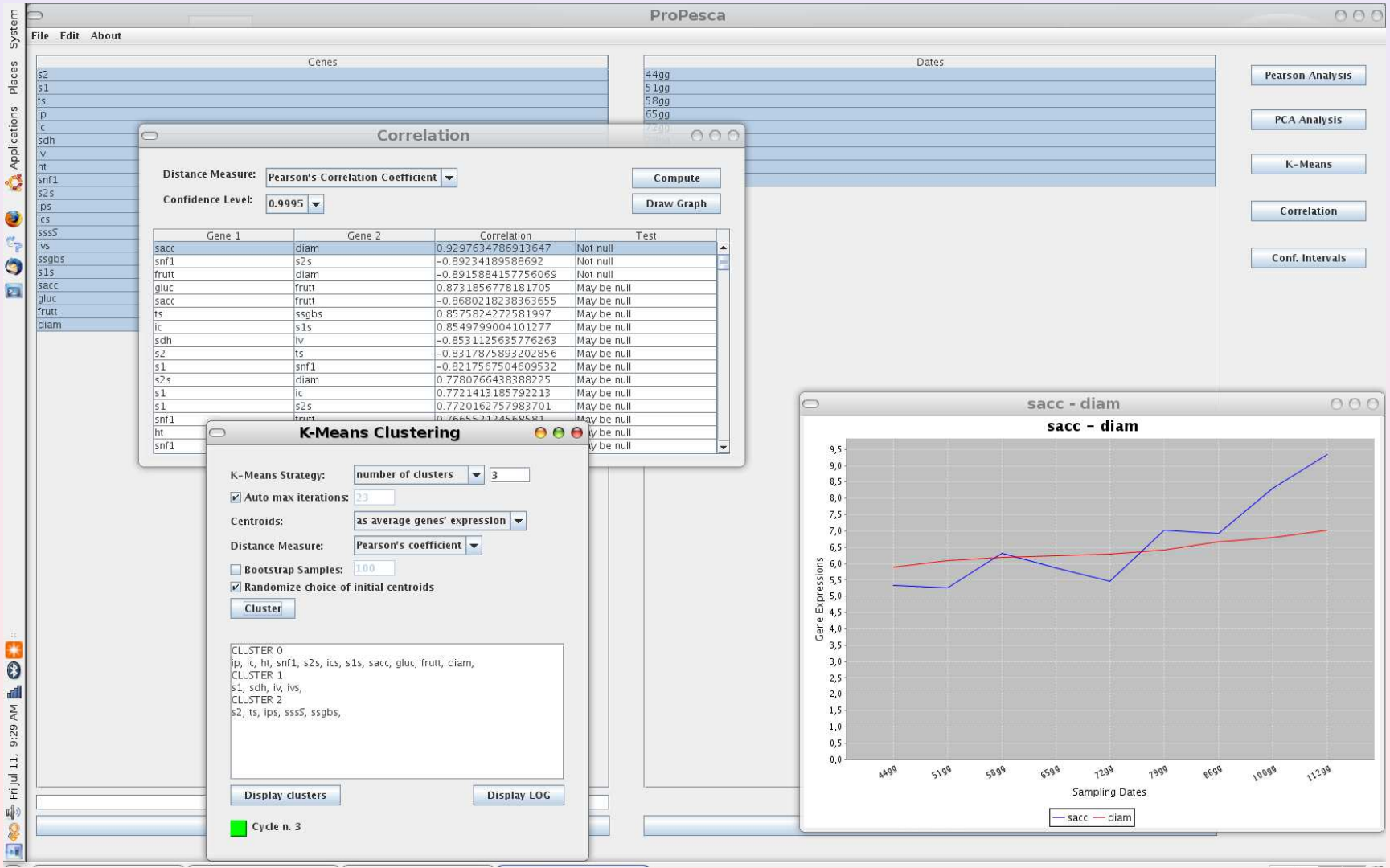


	A	B	C
1	Label	C(T)	
2	YML042W 2003 1gg	17400	
3	YML042W 2003 2gg	31320	
4	YML042W 2003 3gg	25230	
5	YML042W 2003 4gg	15660	
6	YML042W 2003 5gg	12470	
7	YML042W 2003 6gg	11970	
8	YML042W 2003 7gg	8880	
9	YML042W 2003 8gg	8360	
10	YML042W 2003 9gg	1270	
11	YML042W 2003 10gg	760	

YPL172W YNL284C YLR312WA YLR284C YKR009C YKL188C YAL054C

We can treat them as random variables and apply Statistical Methods. . . or even Information Theory.

ProPesca


ProPesca
File Edit About
Genes
s2
s1
ts
ip
ic
sdh
iv
ht
snf1
s2s
ips
ics
sssS
ivs
ssgbs
s1s
sacc
gluc
frutt
diam
Dates
44gg
51gg
58gg
65gg
Pearson Analysis
PCA Analysis
K-Means
Correlation
Conf. Intervals
Correlation
Distance Measure:
Pearson's Correlation Coefficient
Confidence Level:
0.9995
Compute
Draw Graph
Gene 1 | Gene 2 | Correlation | Test
sacc | diam | 0.9297634786913647 | Not null
snf1 | s2s | -0.89234189588692 | Not null
frutt | diam | -0.8915884157756069 | Not null
gluc | frutt | 0.8731856778181705 | May be null
sacc | frutt | -0.8680218238363655 | May be null
ts | ssgbs | 0.8575824272581997 | May be null
ic | s1s | 0.8549799004101277 | May be null
sdh | iv | -0.8531125635776263 | May be null
s2 | ts | -0.8317875893202856 | May be null
s1 | snf1 | -0.8217567504609532 | May be null
s2s | diam | 0.7780766438388225 | May be null
s1 | ic | 0.7721413185792213 | May be null
s1 | s2s | 0.7720162757983701 | May be null
K-Means Clustering
K-Means Strategy:
number of clusters
3
Auto max iterations:
23
Centroids:
as average genes' expression
Distance Measure:
Pearson's coefficient
Bootstrap Samples:
100
Randomize choice of initial centroids
Cluster
CLUSTER 0
ip, ic, ht, snf1, s2s, ics, s1s, sacc, gluc, frutt, diam,
CLUSTER 1
s1, sdh, iv, ivs,
CLUSTER 2
s2, ts, ips, sssS, ssgbs,
Display clusters
Display LOG
Cycle n. 3
sacc - diam
sacc - diam
Gene Expressions
Sampling Dates
44gg
51gg
58gg
65gg
72gg
79gg
86gg
100gg
112gg
sacc
diam

¿Output?

Biologists questions:

- Is this pathway complete? Are there missing nodes or edges?
- Which are the admissible equilibria? Is there a periodic behavior?

- If these are the PCR data, which is the hidden pathway?
- Which genes control this phenomenon?
- Which are the main phases in this behavior?

¿Output?

Biologists questions:

- Is this pathway complete? Are there missing nodes or edges?
- Which are the admissible equilibria? Is there a periodic behavior?

- If these are the PCR data, which is the hidden pathway?
- Which genes control this phenomenon?
- Which are the main phases in this behavior?

Our answers  : I cannot tell. I need more information. I need consistent data.

Robustness Property

H. Kitano – Nature 2004

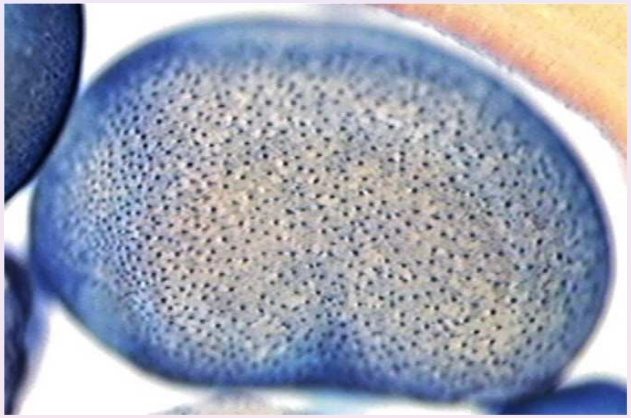
Robustness is a ubiquitously observed property of biological systems. It is considered to be a fundamental feature of complex evolvable systems. It is attained by several underlying principles that are universal to both biological organisms and sophisticated engineering systems. Robustness facilitates evolvability and robust traits are often selected by evolution.

. . . and Others Properties

- Spatial Information. Interactions occur only when the reactants are “close”.
- Fast/Slow Reactions. When different phenomena involve different time scales it seems necessary to study them separately, but sometime they are mutually dependent.
- Scalability. We are modeling cells. The challenge is to model tissues, organs, systems.
- . . . see, e.g., B. Mishra et al. *A Sense of Life*. OMICS 2003.

Delta Notch Example

Delta and Notch are proteins involved in cell differentiation (see, e.g., Collier et al., Ghosh et al.).



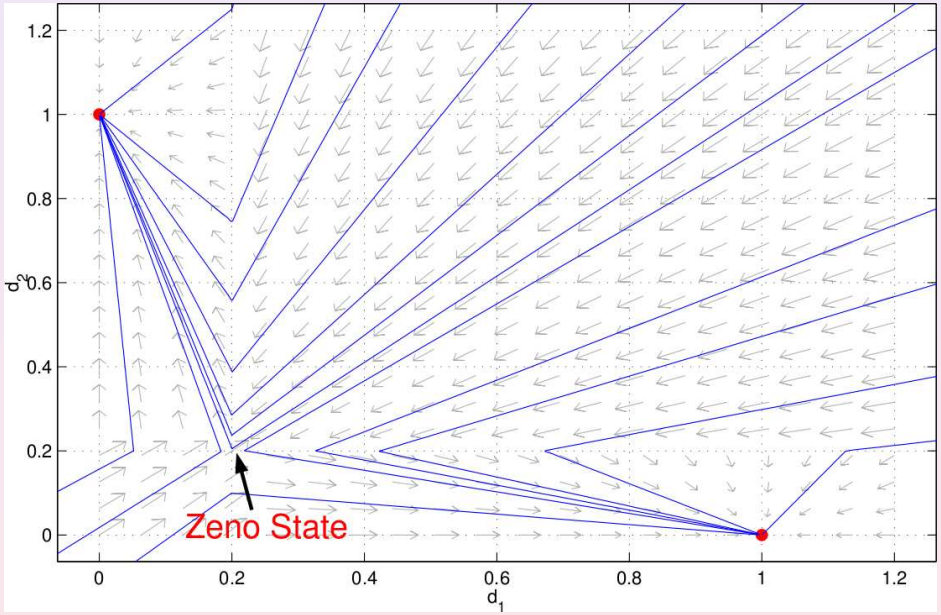
Notch production is triggered by high Delta levels in neighboring cells.

Delta production is triggered by low Notch concentrations in the same cell.

High Delta levels lead to differentiation.

Delta-Notch Example – Two Cells

Involves spatial information, scalability, . . .



. . . and robustness. A Zeno state occur if the cells have identical initial concentrations.

Which kind of Model/Logic?

- Quantitative vs Qualitative.
  Do we have enough data?
- Dense vs Discrete.
  Is nature discrete or dense?
- Stochastic vs (Non) Deterministic.
  Does (non) determinism exist in nature?

Which kind of Model/Logic?

- Quantitative vs Qualitative.
  Do we have enough data?
- Dense vs Discrete.
  Is nature discrete or dense?
- Stochastic vs (Non) Deterministic.
  Does (non) determinism exist in nature?

- Mathematical vs Computational.

J. Fisher and T. A. Henzinger. *Executable cell biology*. Nat. Biotech. 2007

Which kind of Model/Logic?

- Quantitative vs Qualitative.
  Do we have enough data?
- Dense vs Discrete.
  Is nature discrete or dense?
- Stochastic vs (Non) Deterministic.
  Does (non) determinism exist in nature?

- Mathematical vs Computational.
  J. Fisher and T. A. Henzinger. *Executable cell biology*. Nat. Biotech. 2007

- We should mix things up: Hybrid Models

Mathematical and Computational Models

J. Fisher and T. A. Henzinger 2007.

Computational Models

A computational model is a formal model whose primary semantics is operational; that is, the model prescribes a sequence of steps or instructions that can be executed by an abstract machine, which can be implemented on a real computer.

Mathematical Models

A mathematical model is a formal model whose primary semantics is denotational; that is, the model describes by equations a relationship between quantities and how they change over time.

Quantitative and Qualitative Models

J. Fisher and T. A. Henzinger 2007.

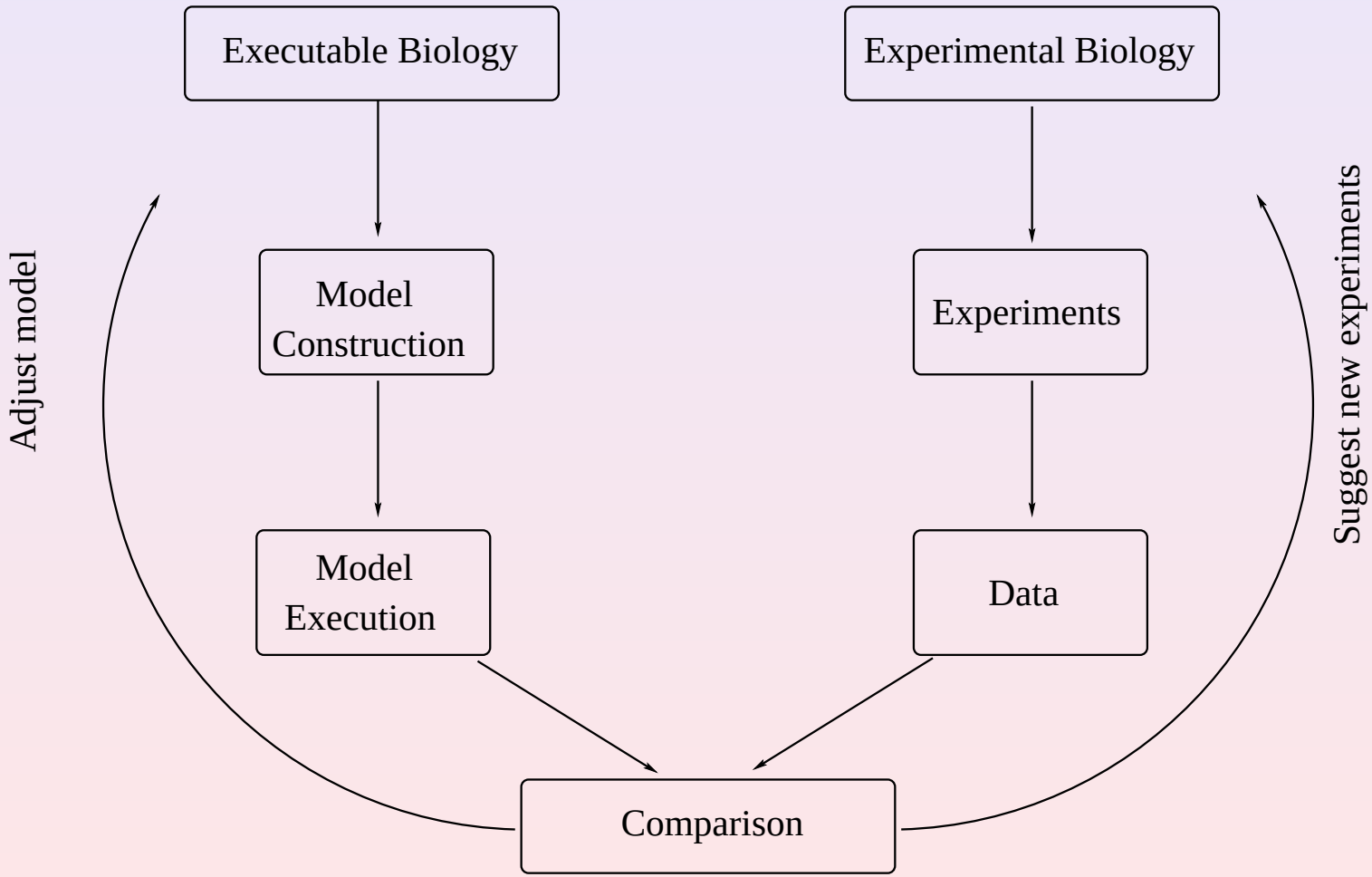
Quantitative Models

Quantitative models are difficult to obtain and analyze if the number of interdependent variables grows and if the relationships depend on qualitative events, such as a concentration reaching a threshold value.

Qualitative Models

A significant advantage of qualitative models is that different models can be used to describe the same system at different levels of detail and that the various levels can be related formally.

Model Tuning



Differential Equations

Differential Equation and Dynamical Systems have been largely used as modeling language in physics, chemistry, biology, engineering, economics.

Definition

A differential equation is a mathematical equation for an unknown function of one or several variables that relates the values of the function itself and of its derivatives of various orders.

- Numerical Analysis methods allow to approximate/simulate solutions.
- Dynamical Systems methods study the qualitative behaviors.

Law of Mass Action

Example

Consider the following set of chemical reactions:

$$\begin{aligned} A + B &\xrightarrow{k_1} C + C \\ A + C &\xrightarrow{k_2} D \end{aligned}$$

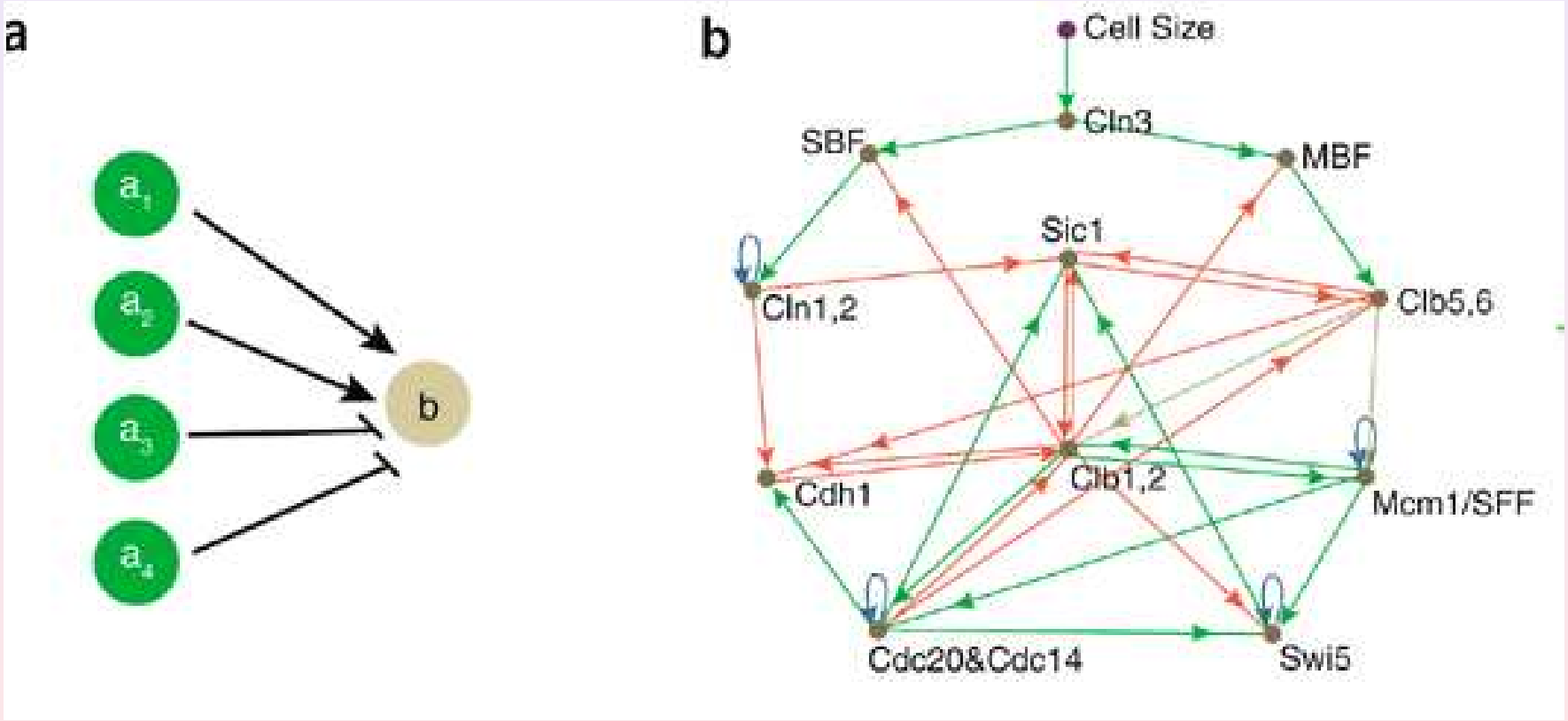
It is modeled by the following ODE:

$$\begin{cases} [\dot{A}] &= -k_1[A][B] - k_2[A][C] \\ [\dot{B}] &= -k_1[A][B] \\ [\dot{C}] &= 2k_1[A][B] - k_2[A][C] \\ [\dot{D}] &= k_2[A][C] \end{cases}$$

Boolean Networks – S. A. Kauffman 1969

- Boolean networks are qualitative, discrete, computational models.
- Each molecule (e.g., gene or protein) is either active or inactive.
- A molecule becomes active if the sum of its activations is larger than the sum of its inhibitions.
- The state of the system is the set of active molecules.
- The states are nodes of a graph. State changes are edges.
- Loops are used to deduce stable states. The number of loops is used to reason about robustness.

Boolean Network – Cell Cycle Budding Yeast



They are not hierarchical and hard to compose.

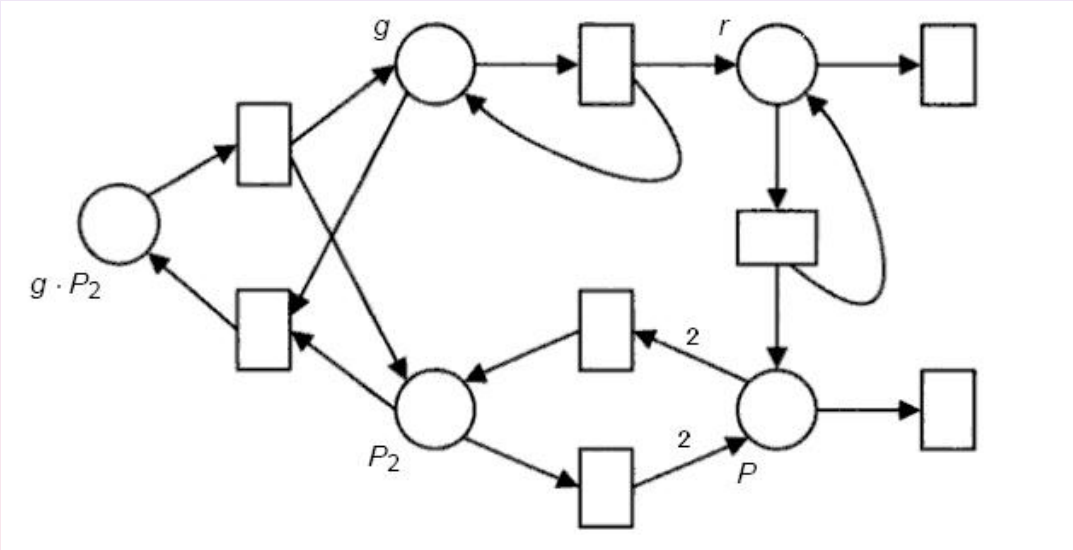
Petri Nets – C. A. Petri 1962, Talcott et al. 2005

- A Petri net is a graph with two types of nodes: places (resources) and transitions (state changes).
- The edges of the graph connect places to transitions and vice-versa.

- The state of the system is represented by places holding tokens.
- Transitions change the state of the system by moving tokens along edges.

- We can also find coloured and stochastic Petri nets.

Petri Nets – Example

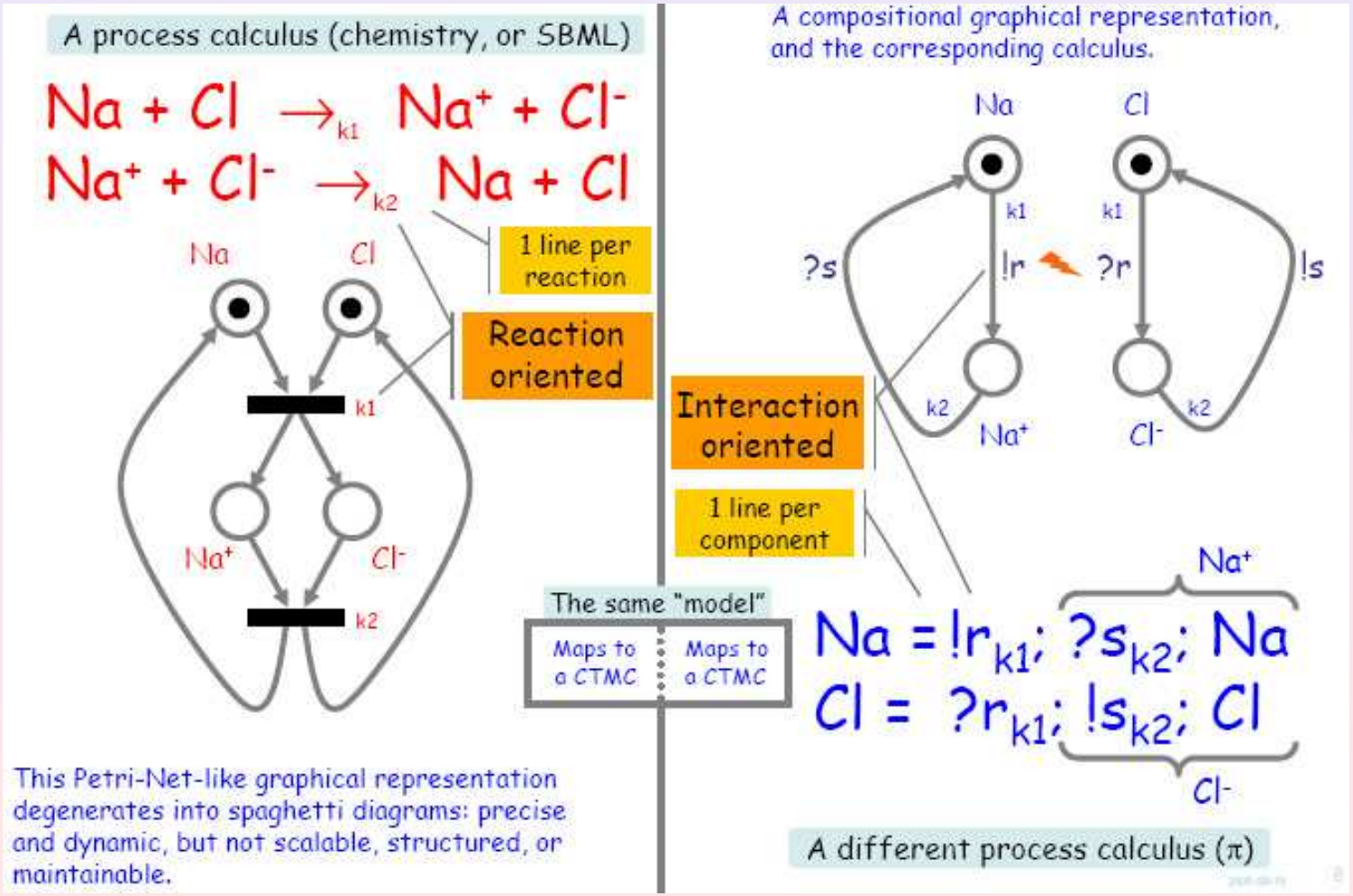


$$g + P_2 \leftrightarrow g \cdot P_2 \quad \text{Repression}$$
$$g \rightarrow g + r \quad \text{Transcription}$$
$$r \rightarrow r + P \quad \text{Translation}$$
$$2P \leftrightarrow P_2 \quad \text{Dimerisation}$$
$$r \rightarrow \emptyset \quad \text{mRNA degradation}$$
$$P \rightarrow \emptyset \quad \text{Protein degradation}$$

Process Calculi – R. Milner 1973, C. Priami 1995, L. Cardelli 2002, . . .

- We are moving toward quantitative, dense, stochastic, computational models.
- Processes model molecules. Many copies of the same process run in parallel to simulate the existence of many molecules.
- Communication between processes models interactions between molecules.
- It is applicable to molecular interactions that occur stochastically.
- Scalability is an issue.

Process Calculi – Example



Other Approaches

- Model Checking and Temporal Logics (see, e.g., Alur et al., Fages et al., Mishra et al.)
- Constraint Programming (see, e.g., WCB proceedings)
- Answer Set Programming (see, e.g., Schaub et al.)
- ...
- Hybrid Automata. Do not miss the second part of this tutorial.

Issues

There are so many . . .

- Compositionality
- Scalability
- Computational Complexity
- Comparisons
- Query Languages
- Robustness
- . . .

Issues

There are so many . . .

- Compositionality
- Scalability
- Computational Complexity
- Comparisons
- Query Languages
- Robustness
- . . .

Suggestions/Solutions are welcome.

Please, BOB (Bring Other Biologists).

Few Names

- Bisca Project: Bologna, Pisa, Siena, Trento, Udine
- Camerino, Milano

- Rajeev Alur, Calin Belta, Luca Cardelli, Edmund Clarke, François Fages, David Harel, Thomas Henzinger, Katsuhisa Horimoto, Hiroaki Kitano, Reinhard Laubenbacher, Bud Mishra, George Pappas, Torsten Schaub, Carolyn Talcott, Ashish Tiwari, Claire Tomlin, Adelinde Uhrmacher, ...

Conferences, Journals, Schools

- Algebraic Biology (AB)
- Computational Methods in Systems Biology (CMSB)
- International Conference on Systems Biology (ICSB)
- From Biology to Concurrency (FBTC)
- Membrane Computing and Biologically Inspired Process Calculi (MeCBIC)
- Workshop on Constraint Based Methods for Bioinformatics (WCB)

- Transactions on Computational Systems Biology
- IEEE/ACM Transactions on Computational Biology and Bioinformatics

- Biology, Computation, and Information (BCI)
- Lipari Summer School on BioInformatics and Computational Biology

Conclusions

- Biology provides us a bunch of interesting problems.
- We briefly talked together about “one” of them.
- There is space for models, logics, algorithms, . . .
- However, integration is fundamental:
  - integration with the biologists.
  - integration of modeling techniques.

Thank you!