COMPLEX NETWORKS OF BIOMOLECULAR INTERACTIONS

François Képès

Epigenomics Project

Genopole®, CNRS, University of Évry

Francois.Kepes@epigenomique.genopole.fr

ICTP, June 8th, 2009
www.mSSB.fr

Deadline for application
June 15th
COMPLEX NETWORKS OF BIOMOLECULAR INTERACTIONS

Complex Systems And Interdisciplinary Science - Vol. 3 BIOLOGICAL NETWORKS edited by François Képès
CONTENT

- Interacting partners and their interaction
- Methods to obtain a network of interactions
- Global structure of the network of interactions
- Local structure
- Dynamics and Modularity
Interacting partners and their interaction
PROTEIN / PROTEIN
Interactions between proteins
Proteomics

METABOLISM
Interactions between enzymes and metabolites
Metabolomics

PROTEIN / DNA
Interactions between regulatory proteins and DNA regulatory regions
Transcriptomics

OTHERS ...
Interactions between proteins

Protein 11
Protein 10
Protein 9
Protein 1
Protein 12
Protein 8
Protein 6
Protein 5
Protein 4
Protein 3
Protein 7

Interaction scores:
0.2, 0.1, 0.5, 0.2, 0.4, 0.1, 1.0, 0.9, 0.8, 0.5, 0.8, 0.3, 0.5, 0.8, 0.3, 0.2
Interactions between enzymes and metabolites

A > B
B > C
C > D

A → B → C → D

A → B (c1)
B → C (c2)
C → D (c3)
Interactions between enzymes and metabolites

Michaelis-Menten Law

\[
\frac{1}{V} = \frac{1}{V_{\text{max}}} + \frac{K_m}{V_{\text{max}}} \cdot \frac{1}{[S]_T}
\]

Lineweaver-Burk Equation
Interactions between enzymes and metabolites

Metabolome
Interactions between regulatory proteins and DNA regulatory regions

SEQUENCES

DNA

TRANSCRIPTION

mRNA

TRANSLATION

Protein

FOLDING

Active Protein

Interactions between regulatory proteins and DNA regulatory regions include transcriptions, translations, and foldings. The regulatory sequences can be either regulatory or coding.
Interactions between regulatory proteins and DNA regulatory regions
Interactions between regulatory proteins and DNA regulatory regions
PROTEIN / PROTEIN
Interactions between proteins
Proteomics

METABOLISM
Interactions between enzymes and metabolites
Metabolomics

PROTEIN / DNA
Interactions between regulatory proteins
and DNA regulatory regions
Transcriptomics
Investment of the Northern American Big Pharma industry for Systems Biology

An example: the discovery of new drugs by systematic screening of a combinatorial chemical bank

**Without**

- Activators
- Kinase for cell division
- Cell division

**Modelling**

- Activator 1
- Activator 2
- Kinase for cell division

**With**

- Activator 1
- Activator 2
- Kinase for cell division
- Cell division
Information Flux in Biology

Epigenesis

L’information et son flux en biologie

Génétique

ADN

ARN

Proteine

Fonction

Épigénétique

ADN

ARN

Proteine

Fonction

Réseaux épigénétiques

Function
Methods to obtain a network of interactions
(Post-genomics)
Proteomics

Separation, identification and quantification of proteins
Proteomics

Intracellular Localization
Proteomics

Double-hybrid approaches
Proteomics

Protein Complexes Assessment

Specific Antibody
Transcriptomics

Complementary DNA Microarrays
Transcriptomics

Oligonucleotide Biochip
Transcriptomics

Other techniques:

RT-PCR

SAGE
Transcriptomics

Chromatin Immuno-Precipitation (ChIP)

Pontage des protéines à l'ADN

Cassage de l'ADN

Immuno-précipitation d'une protéine

Déprotéinisation et marquage de l'ADN

Sondage d'une micromatrice d'ADN

ADN référence marqué
Transcriptomics

Bioinformatics

Information content expressed in bits
Global Structure of the Network of Interactions
SEVERAL GENES

Interactions between regulatory proteins and DNA regulatory regions

Gene A
RR ORF

Gene B

Gene C
+

Gene D
-

Gene E
-
A NETWORK OF GENES
Pleiotropic or multigenic regulation?
Regulated genes vs. Regulating proteins

Global Topology

Exponential
Average 2.3

INDEGREE DISTRIBUTION

\[ Y = M_0 e^{M_1 x} \]

- \( M_0 = 156.59 \)
- \( M_1 = -0.44525 \)
- \( R = 0.98574 \)
Regulating proteins vs. Regulated genes

Outdegree

Outdegree distribution

Global Topology

Power
Average 8.3

\[ Y = M_0 \times X^{M_1} \]

- \( M_0 \approx 22,765 \)
- \( M_1 \approx -0.86409 \)
- \( R \approx 0.94621 \)
Global Topology

Joint In / Out Distribution

Joint probability

Regulating proteins

Regulated genes
## ESSENTIALITY AND DESCENT SIZE

6 essential inter-regulatory genes

<table>
<thead>
<tr>
<th>ORF</th>
<th>Direct descent</th>
<th>Overall descent</th>
</tr>
</thead>
<tbody>
<tr>
<td>YBR049C</td>
<td>18</td>
<td>86</td>
</tr>
<tr>
<td>YMR043W</td>
<td>29</td>
<td>52</td>
</tr>
<tr>
<td>YNL216W</td>
<td>9</td>
<td>49</td>
</tr>
<tr>
<td>YGL207W</td>
<td>11</td>
<td>22</td>
</tr>
<tr>
<td>YGL073W</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>YML010W</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
ESSENTIAL INTER-REGULATORY GENES

Lethal gene inactivation
Local Structure of the Network of Interactions
LOCAL APPORTIONMENT OF EDGES PER VERTEX

Local topology

UNIFORM

SMALL WORLD

HIGHLY CLUSTERED FRAGMENTED
CLIQUISHNESS (clustering coefficient)

Local topology

- CLIQUISHNESS (clustering coefficient)
- Local topology
- CLIQUISHNESS (clustering coefficient)
Dynamics and Modularity
Homogeneous networks:

Why partition at all?

• Modularity

• Dynamical explanation

• Biological relevance

• Compositionnality
<table>
<thead>
<tr>
<th>Topology</th>
<th>Physical properties</th>
<th>Qualitative dynamics</th>
<th>Numerical simulation</th>
<th>In vivo simulation</th>
</tr>
</thead>
</table>

**MOTIFS**
MOTIFS

MODULES

SIGN
# negative interactions

Dynamic property

Biological property

Topology

Qualitative dynamics
(in vivo and numerical simulations carried out)

FEEDBACK CIRCUITS

POSITIVE
Even

Multistationarity
Differentiation

NEGATIVE
Odd

Homeostasis
Stable regulation

Out (A)

A
B

A

(in vivo and numerical simulations carried out)
FEEDBACK CIRCUITS AMONG INTERREGULATORY GENES

Positive circuits

Sporulation

MDR

Pseudo-hyph
FEEDBACK CIRCUITS AMONG INTERREGULATORY GENES

1. DNA damage
2. Glc absence
3. oxygen

Negative circuits
### MOTIFS

#### MODULES

**TYPE**
- # negative interactions
- Dynamic property
- Biological property

**Topology**

**Qualitative dynamics**

*(in vivo and numerical simulations carried out)*

#### FEEDBACK CIRCUITS

**NEGATIVE**
- Odd
- Oscillator
- Stable regulation or oscillation

![Diagram](image)
TRIANGLES (Feed-forward loops)

**MODULES**

- Dynamic property
- Biological property
- Topology
- Qualitative dynamics

**TRIANGLES**

**COHERENT**
- Filter out pulses
- Respond to persistent stimulations
- Rapidly shutdown
- Decide from fluctuating signal

**INCOHERENT**
- Initially reacts strongly
- Later comes back to intermediate levels
- Easily reverse

(Coherent triangle was numerically simulated)
YGL209W: repressor involved in glucose repression.
YGL035C: repressor involved in glucose repression.
YDR146C: controls cell cycle-specific transcription.
YKL109W: glucose-repressed subunit of the HAP transcriptional complex involved in the fermentation-respiration shift.
YMR280C: required for derepression of gluconeogenic enzymes.
YPL248C: involved in expression of galactose-induced genes.
MOTIFS

MODULES
TYPE
Dynamic property

CASCADeS (Linear set of regulations)
SHORT
Rapidly shoots up

LONG
Long lag before it shoots up

Biological property
Fast response in microbe

Time counting in multicellular

Topology

A → Z

A → B → ... → Z

Qualitative dynamics
(Not simulated yet)

In (A)

Time ->

Out (Z)
MOTIFS

MODULES Combination of long cascade and positive circuits

TYPE
Dynamic property

LOCK-ON
Ratchet

Biological property

Succession of time lags and differentiation events in multicellular development

Topology

Qualitative dynamics
for a single-colored series
(Not simulated yet)

\[
\begin{align*}
\text{In (A)} & \quad \text{Out (C)} \\
A \rightarrow B \rightarrow C \rightarrow D \rightarrow E \rightarrow F \rightarrow \ldots \rightarrow X \rightarrow Y \rightarrow Z
\end{align*}
\]
**MOTIFS**

<table>
<thead>
<tr>
<th>MODULES</th>
<th>SIM / SOM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TYPE</strong></td>
<td><strong>Single-Input Module</strong></td>
</tr>
<tr>
<td>Dynamic property</td>
<td>Ordered temporal response</td>
</tr>
<tr>
<td>Biological property</td>
<td>Sequential firing based on differential thresholds</td>
</tr>
<tr>
<td>Topology</td>
<td></td>
</tr>
<tr>
<td>Qualitative dynamics</td>
<td></td>
</tr>
</tbody>
</table>

(Verified in vivo)

- Single-Input Module Diagram:
  - In (A) → Out (B)
  - In (A) → Out (C)

- Single-Output Module Diagram:
  - In (A) → D
  - In (B) → D
  - Out (D) → C

Diagram:
- A → B
- A → C
- B → D
- C → D
Is it possible to spot combinations of motifs?

What could be said about their representation?

What could be said about their topology?

What could be said about their dynamics?
Some of the present challenges

1) Recompose motifs in a useful way

2) Top-down partition into 'functional' modules

3) Go into more global dynamics

4) Analyze mixed networks

5) Realistically model evolution of networks

6) Unfold topologies in geometrical cellular space

7) Control

8) Exploit hybrid formalisms
Mathematical formalisms
A small boolean network. a) The wiring diagram in a boolean network with three genes (A, B and C), each an input to the other two. b) The boolean rules for the diagram shown in a), assuming that gene A represents an AND gate, while genes B and C each represent an OR gate. c) The state transition graph of the boolean network depicted in a) and b). Each triplet of digits correspond to a state for genes A-C, from left to right.
A small logical network "à la René Thomas". a) The regulatory interactions for mucus production in the opportunistic pathogen *Pseudomonas aeruginosa*. Two genes, encoding an activator (A) and an inhibitor (B) of mucus production are considered. Each edge in the graph is labeled with the rank number of the threshold, followed by the sign of its regulatory influence (-, inhibition; +, activation). Given parameters (not shown here), a dynamics may be deduced. b) The asynchronous state graph. This graph is one among several graphs that would fulfill the constraints based on biological knowledge or hypotheses. A can take any value among {0, 1, 2}, and B among {0, 1}. Thresholds are represented by dashed lines, and transitions by arrows. The graph shows two steady states, one for $A = 0$, and one cycle: $11 \rightarrow 10 \rightarrow 20 \rightarrow 21 \rightarrow 11$. 
A Petri net representation of a set of regulatory interactions. Circles denote places identified by a letter, black rectangles are transitions, and arrows are arcs. Only discrete elements are shown.
A bayesian network. a) In this directed acyclic graph, a vertex corresponds to a molecular entity such as a gene or a protein, and holds a random variable representing the gene expression level or the protein concentration. b) A conditional probability distribution is defined for the variable of each vertex, given the variables of its direct inputs in the directed graph. c) A joint probability distribution is finally defined from all conditional distributions.
Hill activation function $H^+$. $\theta_i$ is the threshold of the regulatory influence of the $i^{th}$ molecule onto its targets, for which $H = 0.5$. $S > 0$ is the steepness parameter. For $S > 1$, Hill curves show a sigmoidal shape, as shown on this graph. As $S$ increases, so does the sigmoidicity of the curve.
For which purpose do we need non-biologists in Biology?

a) To hold the pipettes
b) To cope with the massive amount of data
c) To cope with the lack of data
d) No opinion