Curating a large-scale regulatory network by evaluating its consistency with expression datasets

Carito Guziolowski, Jeremy Gruel, Ovidiu Radulescu, Anne Siegel
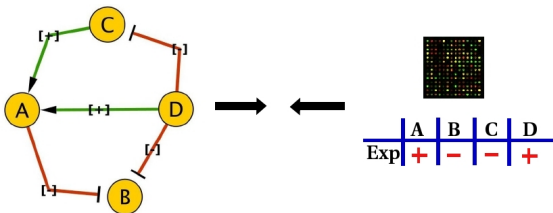
Project Symbiose - IRISA-INRIA - Rennes, France

CIBB 2008

IRISA   Symbiose

Objective

- Confront regulatory knowledge with experimental data
  - Regulatory knowledge → influence graph (discrete possitive and negative interactions)
  - Experimental Data: → qualitative shifts

Existing Approaches

- Local consistency approaches
  - Objective: correlation between regulator and targets expression (consistency)
  - Applied on: regulons, network modules
  - Experimental Data: genome scale experiments
 - Gutierrez *et al.*, *Genome Res* 2003, Herrgard *et al.*, *Genome Res* 2003

- Causal Models
  - Objective: detect/validate regulations in the network that explain observations
  - Causal rule: *"An observed change in expression of a gene, when another gene is altered (e.g. deleted), implies a regulation between the altered gene and the observed"*
  - Model optimization, path analysis
  - Applied on: precise pathways, average size
  - Experimental Data: Genetic perturbation experiments
 - Ideker *et al.*, *Science*, 2001

Our approach: Global Consistency

- Consistency rule: *"The variation of one molecule in the network must be explained by an influence received from at least one of its predecessors in the network"* (Siegel *et al.*, *Biosystems*, 2006)
- Rule encoded by an equation over qualitative variables: $\{+,-\}$ (signs in the graph, signs of gene-expression shifts)
- Global consistency in terms of solution of a qualitative system of equations
- Large scale networks and experimental data

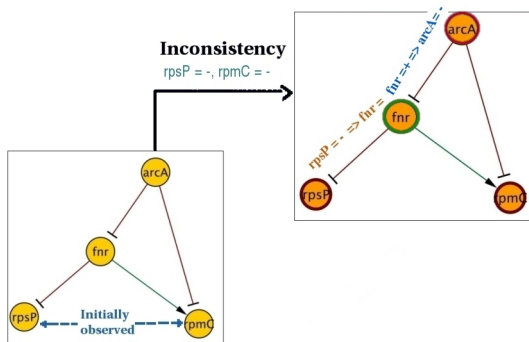## Consistency analysis: Examples
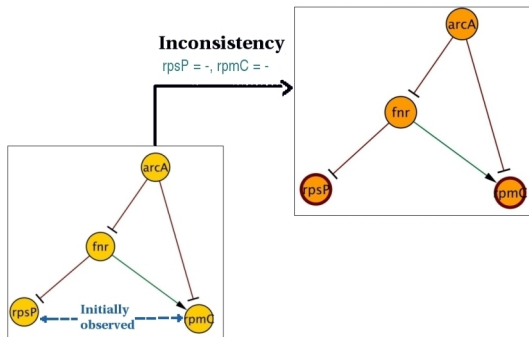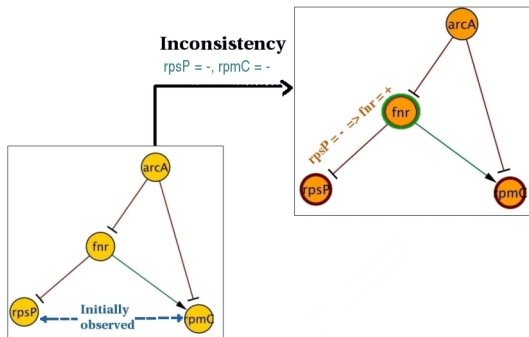
Inconsistency
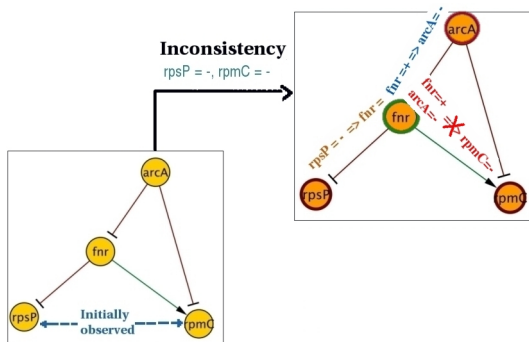rpsP = -, rpmC = -

Initially observed

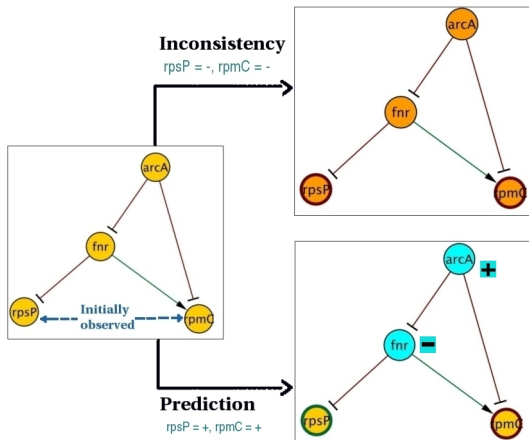# Consistency analysis: Examples

Consistency analysis: Examples

Consistency analysis: Examples

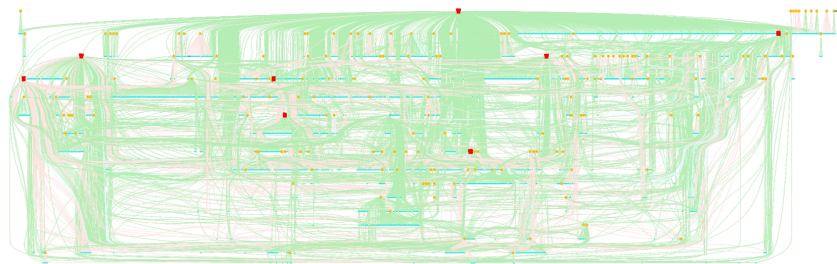- Inconsistency: No possible $\{+,-\}$ value for *arcA*, *fnr*

Consistency analysis: Examples

- Inconsistency: No possible $\{+, -\}$ value for *arcA*, *fnr*

- Prediction: Fixed values of *arcA* and *fnr* that explain the observations

Escherichia coli transcriptional network
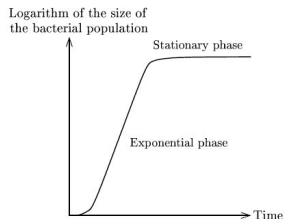

- Large-scale: 1763 nodes, 4491 edges
- Hierarchical topology
    - 8 global factors
    - 162 TF
    - 1593 genes
- Source: RegulonDB (Salgado *et al.* 2006)

- Regulations:
    - TF-gene
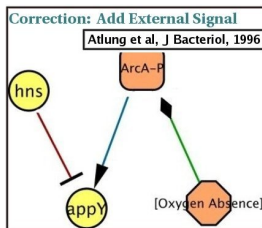   - sigma-gene
    - complex formation

Experimental Condition
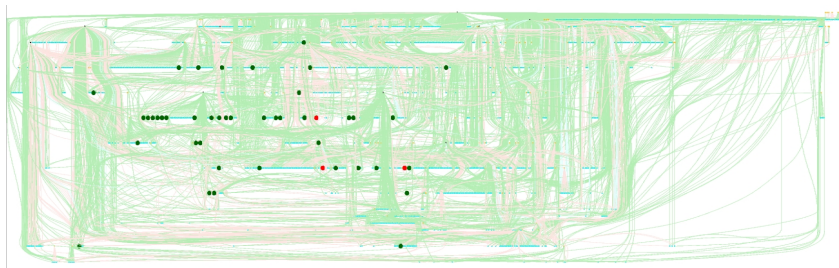
- Condition: Exponential-stationary growth shift
- 45 induced/repressed genes
- Heterogeneous data
- Source: RegulonDB (literature curated)

Ropers *et al.*, *Biosystems*, 2004.
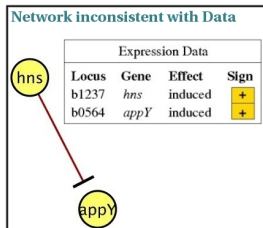
Results: inconsistency

- 1 inconsistent module

- We found the missing regulation that explains the heterogeneous experimental data in the large scale transcriptional *E. coli* network

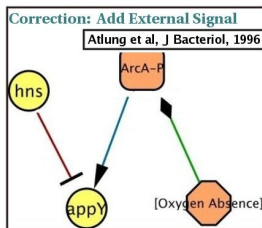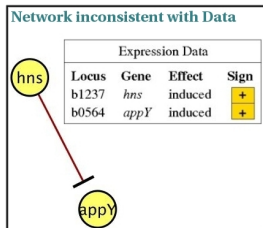- Consistency was obtained after adding a post-transcriptional effect

Results: inconsistency

- 1 inconsistent module

- We found the missing regulation that explains the heterogeneous experimental data in the large scale transcriptional *E. coli* network
- Consistency was obtained after adding a **post-transcriptional effect**

Results: post-transcriptional

- New rule (new equation) that models complex formation: *The concentration of a protein complex at the steady state follows the concentration of the limiting subunit* (Radulescu *et al.*, FOSBE 2007)

  - Applied to the IHF protein complex of *E. coli*

  - ihfA = +, ihfB =−, => IHF = −

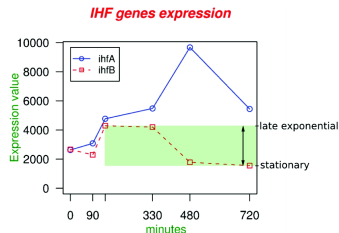## Results: post-transcriptional

- New rule (new equation) that models complex formation:*The concentration of a protein complex at the steady state follows the concentration of the limiting subunit* (Radulescu *et al.*, FOSBE 2007)

    - Applied to the IHF protein complex of *E. coli*

    - ihfA = **+**, ihfB =**−**, => IHF = **−**

Results: post-transcriptional

- New rule (new equation) that models post-transcriptional behaviours:*The concentration of a protein complex at the steady state follows the concentration of the limiting subunit* (Radulescu *et al.*, FOSBE 2007)

  - Applied to the IHF protein complex of *E. coli*
  - ihfA = **+**, ihfB =**−**, => IHF = **−**

IHF genes expression

- System predictions drastically changed: from predicting 0,01% to predicting 30% of the network products as {+, −}
- Small changes in the network may imply larger responses

Results: post-transcriptional

- New rule (new equation) that models post-transcriptional behaviours: *The concentration of a protein complex at the steady state follows the concentration of the limiting subunit* (Radulescu *et al.*, FOSBE 2007)

  - Applied to the IHF protein complex of *E. coli*
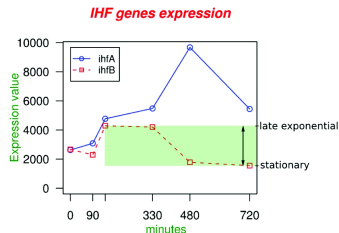 - ihfA = **+**, ihfB =**−**, => IHF = **−**

IHF genes expression

- System predictions drastically changed: from predicting 0,01% to predicting 30% of the network products as {**+**, **−**}
- Small changes in the network may imply larger responses

Results: post-transcriptional

- New rule (new equation) that models post-transcriptional behaviours: *The concentration of a protein complex at the steady state follows the concentration of the limiting subunit* (Radulescu *et al.*, FOSBE 2007)

IHF genes expression

- Applied to the IHF protein complex of *E. coli*
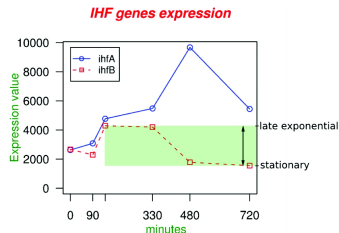- ihfA = **+**, ihfB = **−**, => IHF = **−**

- System predictions drastically changed: from predicting 0,01% to predicting 30% of the network products as {**+**, **−**}
- Small changes in the network may imply larger responses

Results: Validation of the approach

- Corrected transcriptional *E. coli* network was consistent with 45 litterature curated exponentially-stationary gene shifts

- The consistency of the network predicted (fixed) 30% of the network products to explain the experimental data

- Validation of predictions: comparision with microarray outputs (Faith *et al.*, PLoS Biology, 2007)

  - 80% of the reported changes were in agreement
  - Comparable to other methods validation rate (Covert *et al.*, *Nature*, 2002, Edwards *et al.*, *PNAS*, 200)
  - Intersting because we only used a transcriptional model

Results: Validation of the approach

- Corrected transcriptional *E. coli* network was consistent with 45 litterature curated exponentially-stationary gene shifts
- The consistency of the network predicted (fixed) 30% of the network products to explain the experimental data
- Validation of predictions: comparision with microarray outputs (Faith *et al.*, PLoS Biology, 2007)
  - 80% of the reported changes were in agreement
  - Comparable to other methods validation rate (Covert *et al.*, Nature, 2002, Edwards *et al.*, PNAS, 200)
 - Intersting because we only used a transcriptional model

Results: Validation of the approach

- Corrected transcriptional *E. coli* network was consistent with 45 litterature curated exponentially-stationary gene shifts

- The consistency of the network predicted (fixed) 30% of the network products to explain the experimental data

- Validation of predictions: comparision with microarray outputs (Faith *et al.*, PLoS Biology, 2007)
    - 80% of the reported changes were in agreement
    - Comparable to other methods validation rate (Covert *et al., Nature*, 2002, Edwards *et al., PNAS*, 200)
    - Intersting because we only used a transcriptional model

Results: Validation of the approach

- Corrected transcriptional *E. coli* network was consistent with 45 litterature curated exponentially-stationary gene shifts
- The consistency of the network predicted (fixed) 30% of the network products to explain the experimental data
- Validation of predictions: comparision with microarray outputs (Faith *et al.*, PLoS Biology, 2007)
  - 80% of the reported changes were in agreement
  - Comparable to other methods validation rate (Covert *et al.*, Nature, 2002, Edwards *et al., PNAS*, 200)
  - Intersting because we only used a transcriptional model

Consistency check process
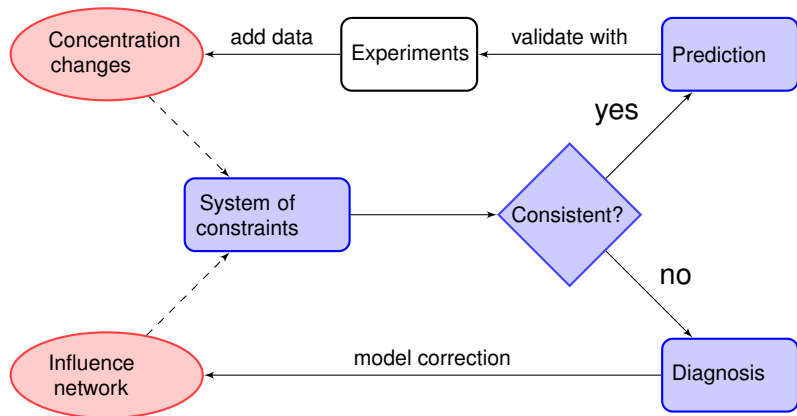
- Bioquali → www.irisa.fr/symbiose/bioquali/
- Cytoscape Plugin →
 www.irisa.fr/symbiose/projects/bioqualiCytoscapePlugin/

Conclusion

- Contront large scale transcriptional (incomplete) model with heterogeneous observations
- Formalize reasoning by a system of qualitative equations
- Global approach
- Diagnostic: Reasoning from a general rule allow us to retrieve missing post-transcriptional mechanisms that explain the observations
- Post-transcriptional effects: small changes in the network may cause big differences on predictions
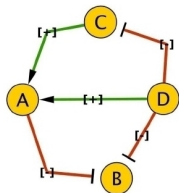- Validated approach on *E. coli*

Acknowledgements

- Symbiose Project at IRISA-INRIA Rennes
  http://www.irisa.fr/symbiose
  - Anne Siegel
  - Michel Le Borgne
  - Jeremy Gruel
  - Ovidiu Radulescu
  - Philippe Veber
- SITCON project
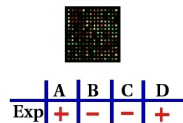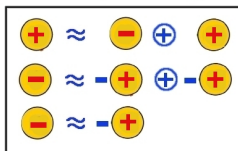 bioinfo-out.curie.fr/projects/sitcon/
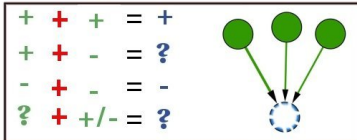
Thank you!
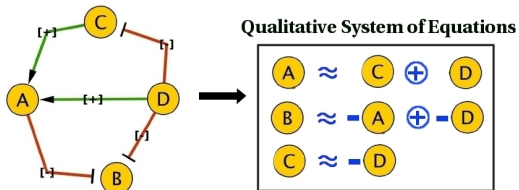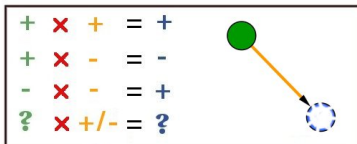
Qualitative System

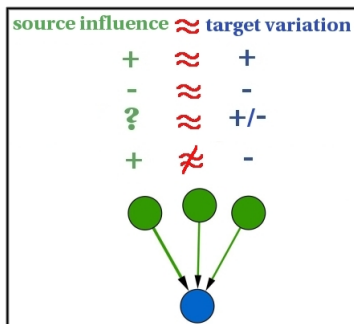**Qualitative System of Equations**

Qualitative System

Signs' Algebra
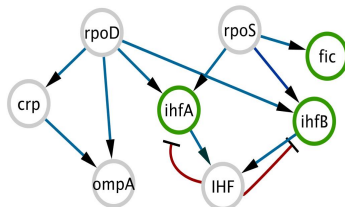
Sign Algebra

Consistency measure

Validation of the predictions

- Compared to microarray outputs (Faith *et al.* 2007 PLoS Biology)
- 80% of the reported changes were in agreement
- And the rest ?

Applications: E. coli (3)

- Validation of predictions
  - Compared to microarray outputs (Faith *et al.* 2007 PLoS Biology)
  - 80% of the reported changes were in agreement
- And the rest ?

Expression Data		
ihfA	=	**+**
ihfB	=	**–**
fic	=	**+**

Prediction		
crp	=	**–**
ompA	=	**–**
rpoD	=	**–**
rpoS	=	**+**
IHF	=	**–**

Applications: E. coli (3)

- Validation of predictions
    - Compared to microarray outputs (Faith *et al.* 2007 PLoS Biology)
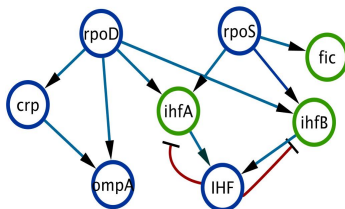   - 80% of the reported changes were in agreement
- And the rest ?

Expression Data		
ihfA	=	**+**
ihfB	=	**−**
fic	=	**+**

Prediction		
crp	=	**−**
ompA	=	**−**
rpoD	=	**−**
rpoS	=	**+**
IHF	=	**−**

- Rsd protein forms a complex with RpoD preventing it to bind RNA-polymerase (Jishage and Ishihama 1998)
- active RpoD predicted not rpoD mRNA

Applications: E. coli (3)

- Validation of predictions
  - Compared to microarray outputs (Faith *et al.* 2007 PLoS Biology)
  - 80% of the reported changes were in agreement
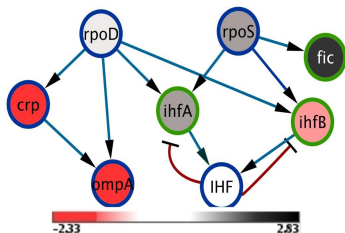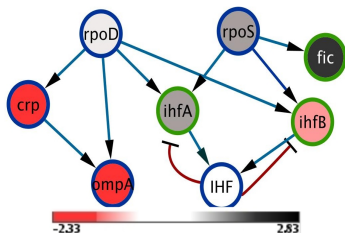- And the rest ?

Expression Data		
ihfA	=	**+**
ihfB	=	**–**
fic	=	**+**

Prediction		
crp	=	**–**
ompA	=	**–**
rpoD	=	**–**
rpoS	=	**+**
IHF	=	**–**

- Rsd protein forms a complex with RpoD preventing it to bind RNA-polymerase (Jishage and Ishihama 1998)
- active RpoD predicted not rpoD mRNA

Applications: E. coli (3)

- Validation of predictions
    - Compared to microarray outputs (Faith *et al.* 2007 PLoS Biology)
    - 80% of the reported changes were in agreement
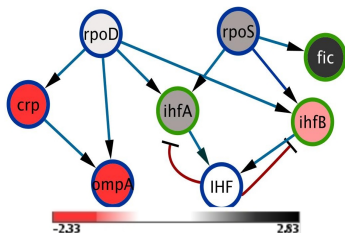- And the rest ?

Expression Data		
ihfA	=	+
ihfB	=	−
fic	=	+

Prediction		
crp	=	−
ompA	=	−
rpoD	=	−
rpoS	=	+
IHF	=	−

- Rsd protein forms a complex with RpoD preventing it to bind RNA-polymerase (Jishage and Ishihama 1998)
- active RpoD predicted not rpoD mRNA