

# Checking Consistency Between Expression Data and Large Scale Regulatory Networks: A Case Study

Carito Guziolowski, Philippe Veber, Michel Le Borgne,  
Ovidiu Radulescu, Anne Siegel  
[Project Symbiose - IRISA](#)

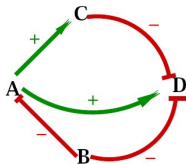
November 2006

# How to study a large scale system ?

- 1 Construct a model from interaction knowledge
  - 2 Test its self-consistency and correct it
  - 3 Validation with experimental data (variations of products)
  - 4 Infer variations of new products
    - Compare this inference with microarray results
- 
- Applied in E. Coli regulatory network ( $\approx$  1000 nodes)
  - Method: Qualitative modelling

# Qualitative modelling object: Interaction graph

- Signed and oriented graph
- Nodes: products of a regulatory network
- Arcs: influences over production



- $A \rightarrow C$  :

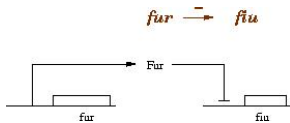
Increase of product  $A$  influences production of  $C$

- Arcs are labeled as:
  - + : activation
  - - : repression
  - ? : double-signed

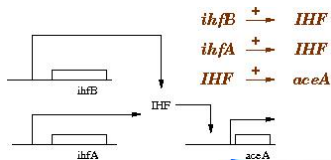
# Constructing interaction graph for E. Coli

- **Source: RegulonDB** <http://regulondb.ccg.unam.mx/index.html>
  - Synchronized with Ecocyc from 2005 (Salgado *et al.*, Bioinformatics, 2006.)
  - Genes, transcription factors, interactions, growth conditions
- **Nodes**
  - Transcription factors (proteins)
  - genes
  - 4 Protein-complexes: *IHF*, *HU*, *RcsB* and *GatR*
- **Interactions**

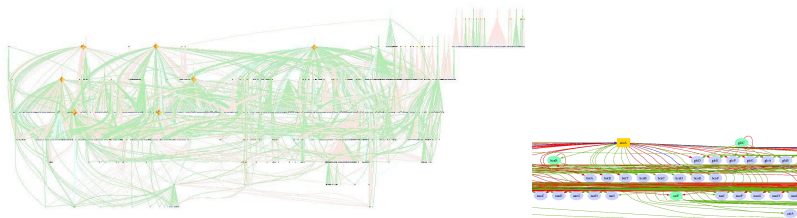
- *genA* → *genB*: *genA* produces protein *A* that regulates *genB*



- *genA* → *PC*: *genA* produces protein *A* that is part of protein-complex *PC*.
- *PC* → *genB*: *PC* is a protein-complex that regulates *genB*



# Interaction graph for E. Coli

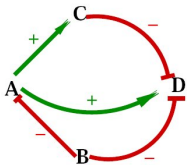


- Large scale network
- Hierarchical structure (87% of genes are regulated by 13%)  
Ma *et al.*, Nucleic Acid Res., 32. 2004.
- 160 doubled-signed interactions
- 7 global factors (> 80 successors):  
CRP, FNR, IHF, FIS, ArcA, NarL and Lrp

Number of nodes	1258
Number of interactions	2526
Nodes without successor	1101
Nodes with more than 80 successors	7
protein complex	4

# Qualitative equations from interaction graph

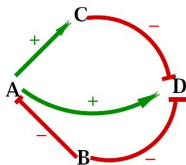
- Main interest: **variation of quantities of products between 2 equilibrium states**
- Remark: Not interested in trajectories
- Example with 4 products:



- $sign(\Delta C) \approx sign(\Delta A)$
- $sign(\Delta A) \approx -sign(\Delta B)$
- $sign(\Delta D) \approx sign(\Delta A) - sign(\Delta B) - sign(\Delta C)$
- $sign(\Delta A) : +, -, ?$
- Natural additions and multiplications among signs:
- Qualitative equality:  $+ \approx ?, - \approx ?$

# Qualitative equations from interaction graph

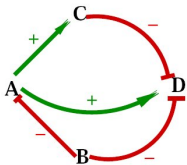
- Main interest: **variation of quantities of products between 2 equilibrium states**
- Remark: Not interested in trajectories
- Example with 4 products:



- $sign(\Delta C) \approx sign(\Delta A)$
  - $sign(\Delta A) \approx -sign(\Delta B)$
  - $sign(\Delta D) \approx sign(\Delta A) - sign(\Delta B) - sign(\Delta C)$
- 
- $sign(\Delta A) : +, -, ?$
  - Natural additions and multiplications among signs:
  - Qualitative equality:  $+ \approx ?, - \approx ?$

# Qualitative equations from interaction graph

- Main interest: **variation of quantities of products between 2 equilibrium states**
- Remark: Not interested in trajectories
- Example with 4 products:

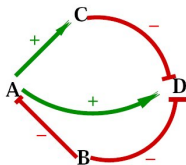


- $sign(\Delta C) \approx sign(\Delta A)$
  - $sign(\Delta A) \approx -sign(\Delta B)$
  - $sign(\Delta D) \approx sign(\Delta A) - sign(\Delta B) - sign(\Delta C)$
- 
- $sign(\Delta A) : +, -, ?$
  - Natural additions and multiplications among signs:
  - Qualitative equality:  $+ \approx ?$ ,  $- \approx ?$



# Qualitative equations from interaction graph

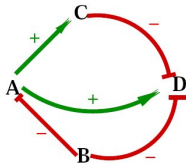
- Main interest: **variation of quantities of products between 2 equilibrium states**
- Remark: Not interested in trajectories
- Example with 4 products:



- $sign(\Delta C) \approx sign(\Delta A)$
  - $sign(\Delta A) \approx -sign(\Delta B)$
  - $sign(\Delta D) \approx sign(\Delta A) - sign(\Delta B) - sign(\Delta C)$
- $sign(\Delta A) : +, -, ?$ 
    - Natural additions and multiplications among signs:
    - Qualitative equality:  $+ \approx ?$ ,  $- \approx ?$

# Qualitative equations from interaction graph

- Main interest: **variation of quantities of products between 2 equilibrium states**
- **Remark:** Not interested in trajectories
- Example with 4 products:



- $sign(\Delta C) \approx sign(\Delta A)$
- $sign(\Delta A) \approx -sign(\Delta B)$
- $sign(\Delta D) \approx sign(\Delta A) - sign(\Delta B) - sign(\Delta C)$

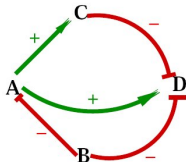
- $sign(\Delta A) : +, -, ?$
- Natural additions and multiplications among signs:

$$\begin{array}{cccccc}
 + + - = ? & + + + = + & - + - = - & + \times - = - & + \times + = + & - \times - = + \\
 ? + - = ? & ? + + = ? & ? + ? = ? & ? \times - = ? & ? \times + = ? & ? \times ? = ?
 \end{array}$$

- Qualitative equality:  $+ \approx ?, - \approx ?$

# Qualitative equations from interaction graph

- Main interest: **variation of quantities of products between 2 equilibrium states**
- **Remark:** Not interested in trajectories
- Example with 4 products:



- $sign(\Delta C) \approx sign(\Delta A)$
- $sign(\Delta A) \approx -sign(\Delta B)$
- $sign(\Delta D) \approx sign(\Delta A) - sign(\Delta B) - sign(\Delta C)$

- $sign(\Delta A) : +, -, ?$
- Natural additions and multiplications among signs:

$$\begin{array}{cccccc}
 + + - = ? & + + + = + & - + - = - & + \times - = - & + \times + = + & - \times - = + \\
 ? + - = ? & ? + + = ? & ? + ? = ? & ? \times - = ? & ? \times + = ? & ? \times ? = ?
 \end{array}$$

- Qualitative equality:  $+ \approx ?, - \approx ?$

# Solving a qualitative system

- Qualitative equation for node  $i$ :

$$\text{sign}(\Delta X_i) \approx \sum_{j \in \text{pred}(i)} \text{sign}(j \rightarrow i) \text{sign}(\Delta X_j)$$

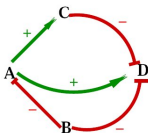
- All the influences over a given node arrive through its closest neighbors
- This equation is true *overknown mathematical hypothesis*
- **Solution** of the qualitative system:
  - Assignment of each variable to + or -
  - All equations are satisfied

# Solving a qualitative system

- Qualitative equation for node  $i$ :

$$\text{sign}(\Delta X_i) \approx \sum_{j \in \text{pred}(i)} \text{sign}(j \rightarrow i) \text{sign}(\Delta X_j)$$

- Solution** of the qualitative system:
  - Assignment of each variable to **+** or **-**
  - All equations are satisfied



$$\begin{aligned} C &\approx A \\ A &\approx -B \\ D &\approx A - B - C \end{aligned}$$

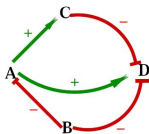
# Solving a qualitative system

- Qualitative equation for node  $i$ :

$$\text{sign}(\Delta X_i) \approx \sum_{j \in \text{pred}(i)} \text{sign}(j \rightarrow i) \text{sign}(\Delta X_j)$$

- Solution** of the qualitative system:

- Assignment of each variable to **+** or **-**
- All equations are satisfied



$$\begin{aligned} C &\approx A \\ A &\approx -B \\ D &\approx A - B - C \end{aligned}$$

$$A = +, B = -, C = +, D = +$$

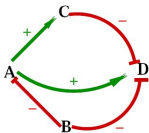
$$\begin{aligned} + &\approx + \\ + &\approx -(-) \\ + &\approx + - (-) - (+) \end{aligned}$$

# Solving a qualitative system

- Qualitative equation for node  $i$ :

$$\text{sign}(\Delta X_i) \approx \sum_{j \in \text{pred}(i)} \text{sign}(j \rightarrow i) \text{sign}(\Delta X_j)$$

- Solution** of the qualitative system:
  - Assignment of each variable to **+** or **-**
  - All equations are satisfied



$$\begin{array}{l}
 C \approx A \quad + \approx + \\
 A \approx -B \quad + \approx -(-) \\
 D \approx A - B - C \quad + \approx + - (-) - (+)
 \end{array}$$

4 sets of variations  
(among 16) are solutions

A	B	C	D
+	-	+	+
+	-	+	-
-	+	-	+
-	+	-	-

# (In)Consistency of E.Coli qualitative system

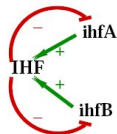
- 1100 qualitative equations
- Study of solutions
  - **PYQUALI**: Python library developed at IRISA
  - Equations represented by a Ternary Diagram of Decision (TDD) of  $3^{1100}$  nodes
  - Study of equivalent reduced system

The qualitative system for the transcriptional network of E.Coli does not have any solution.



# Corrections

- Problem (automatically detected by PYQUALI):



$$IHF \approx ihfA + ihfB \quad (1)$$

$$ihfA \approx -IHF \quad (2)$$

$$ihfB \approx -IHF \quad (3)$$

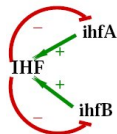
Inconsistent system (no valuation is solution)

ihfA	ihfB	IHF	Conflict
+	+	+	(2), (3)
+	+	-	(1)
+	-	+	(1)
+	-	-	(1)
-	+	+	(1)
-	+	-	(1)
-	-	+	(1)
-	-	-	(2), (3)

- Solution: Adding sigma-factor interactions

# Corrections

- Problem (automatically detected by PYQUALI):



$$IHF \approx ihfA + ihfB \quad (1)$$

$$ihfA \approx -IHF \quad (2)$$

$$ihfB \approx -IHF \quad (3)$$

Inconsistent system (no valuation is solution)

ihfA	ihfB	IHF	Conflict
+	+	+	(2), (3)
+	+	-	(1)
+	-	+	(1)
+	-	-	(1)
-	+	+	(1)
-	+	-	(1)
-	-	+	(1)
-	-	-	(2), (3)

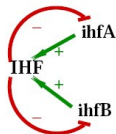
- Solution: Adding sigma-factor interactions

# Corrections

- Problem (automatically detected by PYQUALI):

Inconsistent system (no valuation is solution)

ihfA	ihfB	IHF	Conflict
+	+	+	(2), (3)
+	+	-	(1)
+	-	+	(1)
+	-	-	(1)
-	+	+	(1)
-	+	-	(1)
-	-	+	(1)
-	-	-	(2), (3)



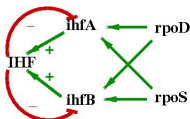
$$IHF \approx ihfA + ihfB(1)$$

$$ihfA \approx -IHF \quad (2)$$

$$ihfB \approx -IHF \quad (3)$$

- Solution: Adding sigma-factor interactions

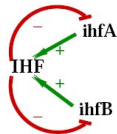
- A  $\sigma$  factor associates to RNAP allowing the transcription of different subset of genes



Protein	Gene	Function
$\sigma^{70}$	rpoD	Transcribes most genes in growing cells
$\sigma^{38}$	rpoS	The starvation/stationary phase sigma-factor
$\sigma^{28}$	rpoF	The flagellar sigma-factor
$\sigma^{32}$	rpoH	The heat shock sigma-factor
$\sigma^{24}$	rpoE	The extracytoplasmic stress sigma-factor
$\sigma^{54}$	rpoN	The nitrogen-limitation sigma-factor
$\sigma^{19}$	fecI	The ferric citrate sigma-factor

# Corrections

- Problem (automatically detected by PYQUALI):



$$IHF \approx ihfA + ihfB \quad (1)$$

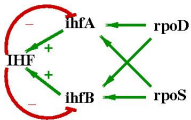
$$ihfA \approx -IHF \quad (2)$$

$$ihfB \approx -IHF \quad (3)$$

Inconsistent system (no valuation is solution)

ihfA	ihfB	IHF	Conflict
+	+	+	(2), (3)
+	+	-	(1)
+	-	+	(1)
+	-	-	(1)
-	+	+	(1)
-	+	-	(1)
-	-	+	(1)
-	-	-	(2), (3)

- Solution: Adding sigma-factor interactions



$$IHF \approx ihfA + ihfB$$

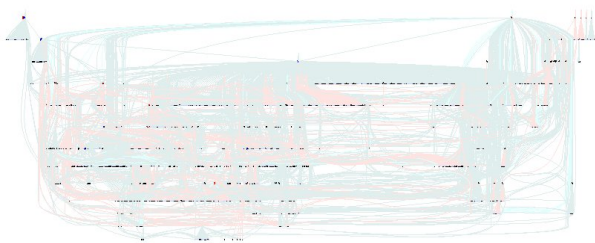
$$ihfA \approx -IHF + rpoD + rpoS$$

$$ihfB \approx -IHF + rpoD + rpoS$$

Consistent system (18 solutions among 32)

rpoD	rpoS	ihfA	ihfB	IHF
+	+	+	+	+
+	+	+	-	+
+	+	-	+	+
-	-	-	-	-
-	-	-	+	-
-	-	+	-	-
<hr/>				
+/-	-/+	+	+	+
+/-	-/+	+	-	+
+/-	-/+	+	-	-
+/-	-/+	-	+	+
+/-	-/+	-	+	-
+/-	-/+	-	-	-

# Self-consistent graph for E. Coli



Number of nodes	1529
Number of interactions	3883
Nodes without successor	1365
Nodes with more than 80 successors	10
sigma-factors	6
protein complex	4

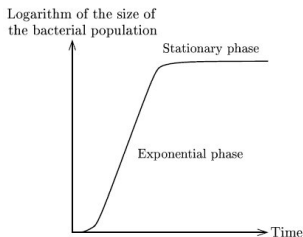
- Add interactions among sigma factors and its related genes
- Sigma factors are considered as external nodes: no equation

Protein	Gene	Induced Genes
$\sigma_{70}$	rpoD	1047
$\sigma_{38}$	rpoS	114
$\sigma_{54}$	rpoN	100
$\sigma_{24}$	rpoE	48
$\sigma_{32}$	rpoH	29
$\sigma_{19}$	fecI	7

Regulatory network of E.coli with sigma factors interactions is self-consistent

# Validation: Experimental data related to Nutritional Stress

- Population of E.coli grows exponentially until a nutritional stress
- Variation of products between 2 equilibrium states:
  - Exponential phase
  - Stationary phase
- 40 experimental data (from RegulonDB)



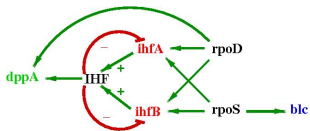
Ropers *et al.*, Biosystems, 2004.

gene	effect	gene	effect	gene	effect	gene	effect	gene	effect
acnA	+	csiE	+	gadC	+	osmB	+	recF	+
acrA	+	cspD	+	hmp	+	osmE	+	rob	+
adhE	+	dnaN	+	hns	+	osmY	+	sdaA	-
appB	+	dppA	+	hyaA	+	otsA	+	sohB	-
appC	+	fic	+	ihfA	-	otsB	+	treA	+
appY	+	gabP	+	ihfB	-	polA	+	yeiL	+
blc	+	gadA	+	lrp	+	proP	+	yfiD	+
bolA	+	gadB	+	mpl	+	proX	+	yihI	-

**Inconsistency:** There does not exist a solution of the qualitative system that agree with the observed values

# Solving inconsistency of the observations (Problem)

Inconsistent subgraph: (automatically detected by PYQUALI)

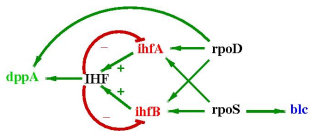


24 consistent solutions :

rpoD	rpoS	ihfA	ihfB	IHF	bhc	dppA
+	+	+	+	+	+	+
+	+	+	-	+	+	+
+	+	-	+	+	+	+
-	-	-	-	-	-	-
-	-	-	+	-	-	-
-	-	+	-	-	-	-
+	-	+	+	+	-	+
-	+	+	+	+	+	+
-	+	+	+	+	+	-
+/-	-/+	+	-	+	-/+	+/-
-	+	+	-	+	+	+
+	-	+	-	-	-	+
+/-	-/+	+	-	-	-/+	-
+/-	-/+	-	+	+	-/+	+/-
-	+	-	+	+	+	+
+	-	-	+	-	-	-
+/-	-/+	-	+	-	-/+	+/-
+	-	-	-	-	-	-
+	-	-	-	-	-	+
-	-	-	-	-	+	-

# Solving inconsistency of the observations (Problem)

Inconsistent subgraph: (automatically detected by PYQUALI)



4 consistent solutions :

rpoD	rpoS	ihfA	ihfB	IHF	blc	dppA
+	+	+	+	+	+	+
+	+	+	-	+	+	+
+	+	-	+	+	+	+
-	-	-	-	-	-	-
-	-	-	+	-	-	-
-	-	+	-	-	-	-
+	-	+	+	+	-	+
-	+	+	+	+	+	+
-	+	+	+	+	+	-
+/-	-/+	+	-	+	-/+	+/-
-	+	+	-	+	+	+
+	-	+	-	-	-	+
+/-	-/+	+	-	-	-/+	-
+/-	-/+	-	+	+	-/+	+/-
-	+	-	+	+	+	+
+	-	-	+	-	-	-
+/-	-/+	-	+	-	-/+	+/-
+	-	-	-	-	-	-
+	-	-	-	-	-	+
-	+	-	-	-	+	-

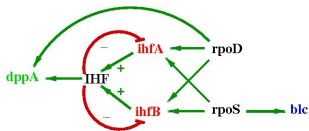
Subset of observations:

- *ihfA* = -
- *ihfB* = -
- *blc* = +
- *dppA* = +



# Solving inconsistency of the observations (Problem)

Inconsistent subgraph: (automatically detected by PYQUALI)



1 consistent solutions :

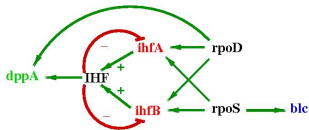
rpoD	rpoS	ihfA	ihfB	IHF	blc	dppA
+	+	+	+	+	+	+
+	+	+	-	+	+	+
+	+	-	+	+	+	+
-	-	-	-	-	-	-
-	-	-	+	-	-	-
+	-	+	+	+	-	+
-	+	+	+	+	+	+
-	+	+	+	+	+	-
+/-	-/+	+	-	+	-/+	+/-
-	+	+	-	+	+	+
+	-	+	-	-	-	+
+/-	-/+	+	-	-	-/+	-
+/-	-/+	-	+	+	-/+	+/-
-	+	-	+	+	+	+
+	-	-	+	-	-	-
+/-	-/+	-	+	-	-/+	+/-
+	-	-	-	-	-	-
+	-	-	-	-	-	+
-	+	-	-	-	+	-

Subset of observations:

- *ihfA* = -
- *ihfB* = -
- *blc* = +
- *dppA* = +

# Solving inconsistency of the observations (Problem)

Inconsistent subgraph: (automatically detected by PYQUALI)



0 consistent solutions :

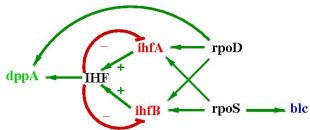
rpoD	rpoS	ihfA	ihfB	IHF	blc	dppA
+	+	+	+	+	+	+
+	+	+	-	+	+	+
+	+	-	+	+	+	+
-	-	-	-	-	-	-
-	-	-	+	-	-	-
-	-	+	-	-	-	-
+	-	+	+	+	-	+
-	+	+	+	+	+	+
-	+	+	+	+	+	-
+/-	-/+	+	-	+	-/+	+/-
-	+	+	-	+	+	+
+	-	+	-	-	-	+
+/-	-/+	+	-	-	-/+	-
+/-	-/+	-	+	+	-/+	+/-
-	+	-	+	+	+	+
+	-	-	+	-	-	-
+/-	-/+	-	+	-	-/+	+/-
+	-	-	-	-	-	-
+	-	-	-	-	-	+
-	+	-	-	-	+	-

Subset of observations:

- *ihfA* = -
- *ihfB* = -
- *blc* = +
- *dppA* = +

# Changing values of observations ?

Inconsistent subgraph: (automatically detected by PYQUALI)



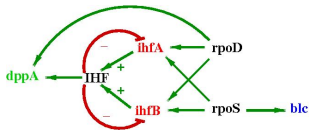
24 consistent solutions :

rpoD	rpoS	ihfA	ihfB	IHF	blc	dppA
+	+	+	+	+	+	+
+	+	+	-	+	+	+
+	+	-	+	+	+	+
-	-	-	-	-	-	-
-	-	-	+	-	-	-
-	-	+	-	-	-	-
+	-	+	+	+	-	+
-	+	+	+	+	+	+
-	+	+	+	+	+	-
+/-	-/+	+	-	+	-/+	+/-
-	+	+	-	+	+	+
+	-	+	-	-	-	+
+/-	-/+	+	-	-	-/+	-
+/-	-/+	-	+	+	-/+	+/-
-	+	-	+	+	+	+
+	-	-	+	-	-	-
+/-	-/+	-	+	-	-/+	+/-
+	-	-	-	-	-	-
+	-	-	-	-	-	+
-	-	-	-	-	+	-

- Corrections over the values of *ihfA* and *ihfB*

# Changing values of observations ?

Inconsistent subgraph: (automatically detected by PYQUALI)



4 consistent solutions :

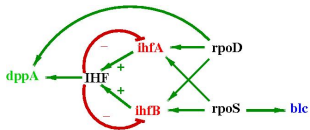
rpoD	rpoS	ihfA	ihfB	IHF	blc	dppA
+	+	+	+	+	+	+
+	+	-	+	+	+	+
-	-	-	+	-	-	-
-	-	+	-	-	-	-
+	-	+	+	+	-	+
-	+	+	+	+	+	+
-	+	+	+	+	+	-
+/-	-/+	+	-	+	-/+	+/-
-	+	+	-	+	+	+
+	-	+	-	-	-	+
+/-	-/+	+	-	-	-/+	-
+/-	-/+	-	+	+	-/+	+/-
-	+	-	+	+	+	+
+	-	-	+	-	-	-
+/-	-/+	-	+	-	-/+	+/-
+	-	-	-	-	-	-
+	-	-	-	-	-	+
-	+	-	-	-	+	-

Corrected observations:

- *ihfA* = +
- *ihfB* = +
- *blc* = +
- *dppA* = +

# Changing values of observations ?

Inconsistent subgraph: (automatically detected by PYQUALI)



3 consistent solutions :

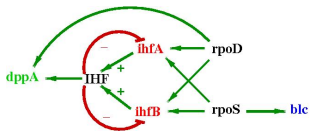
rpoD	rpoS	ihfA	ihfB	IHF	blc	dppA
+	+	+	+	+	+	+
+	+	-	+	+	+	+
-	-	-	+	-	-	-
-	-	+	-	-	-	-
+	-	+	+	+	-	+
-	+	+	+	+	+	+
-	+	+	+	+	+	-
+/-	-/+	+	-	+	-/+	+/-
-	+	+	-	+	+	+
+	-	+	-	-	-	+
+/-	-/+	+	-	-	-/+	-
+/-	-/+	-	+	+	-/+	+/-
-	+	-	+	+	+	+
+	-	-	+	-	-	-
+/-	-/+	-	+	-	-/+	+/-
+	-	-	-	-	-	-
+	-	-	-	-	-	+
-	+	-	-	-	+	-

Corrected observations:

- *ihfA* = +
- *ihfB* = +
- *blc* = +
- *dppA* = +

# Changing values of observations ?

Inconsistent subgraph: (automatically detected by PYQUALI)



## 2 consistent solutions :

rpoD	rpoS	ihfA	ihfB	IHF	blc	dppA
+	+	+	+	+	+	+
+	+	-	+	+	+	+
-	-	-	+	-	-	-
-	-	+	-	-	-	-
+	-	+	+	+	-	+
-	+	+	+	+	+	+
-	+	+	+	+	+	-
+/-	-/+	+	-	+	-/+	+/-
-	+	+	-	+	+	+
+	-	+	-	-	-	+
+/-	-/+	+	-	-	-/+	-
+/-	-/+	-	+	+	-/+	+/-
-	+	-	+	+	+	+
+	-	-	+	-	-	-
+/-	-/+	-	+	-	-/+	+/-
+	-	-	-	-	-	-
+	-	-	-	-	-	+
-	+	-	-	-	+	-

Corrected observations:

- *ihfA* = +
- *ihfB* = +
- *blc* = +
- *dppA* = +

# Solution

- Proposed corrections: *ihfA* = +, *ihfB* = +
- References related to *ihfA* and *ihfB*
- Data for stationary phase provided by RegulonDB data was incorrect

# Solution

- Proposed corrections: *ihfA* = +, *ihfB* = +
- References related to *ihfA* and *ihfB*

Expression of the genes coding for the Escherichia coli integration host factor are controlled by growth phase, rpoS, ppGpp and by autoregulation.

[Aviv M.](#), [Giladi H.](#), [Schreiber G.](#), [Oppenheim AB.](#), [Glaser G.](#)

Department of Cellular Biochemistry, Hebrew University–Hadassah Medical School, Jerusalem, Israel.

Transcriptional control of the *himA* and the *himD/hip* genes coding for the two subunits of the integration host factor (IHF) was investigated. The promoters for the two genes were identified by the use of primer extension and S1 analysis. Expression from both promoters was found to increase as the cells enter stationary phase. Mutation in *rpoS*, known to be induced upon entry to stationary phase, dramatically reduced the growth-phase response of the *himA* P4 promoter but had only a small effect on the induction of the *himD/hip* promoter. The increased activity of both promoters required the presence of the *relA* and *spoT* genes, suggesting that ppGpp plays a major role in the response to stationary phase. An artificial increase in ppGpp in exponentially growing cells induced a rapid increase in *himA* P4 and *himD/hip* mRNA levels. Experiments with a mutant defective in *rpoS* showed that the response of the *himA* P4 promoter to high ppGpp levels was greatly reduced while that of *himD/hip* was only slightly affected. Therefore, it seems that different mechanisms involving RpoS and ppGpp regulate the growth-phase response of the two promoters. We propose that the effect of ppGpp on *himA* P4 is mediated via RpoS whereas the *himD/hip* promoter is affected by ppGpp independently of RpoS. Expression of the *himD/hip* and *himA* genes was found to be subject to negative autoregulation. IHF-binding sites, implicated in autoregulation, were found to overlap both the *himD/hip* and *himA* P4 promoters. An additional IHF-binding site was found upstream of the *himD/hip* promoter. All three sites show low binding affinity to IHF suggesting that autoregulation can take place only after sufficiently high levels of IHF accumulate in the cell.

PMID: 7715442 [PubMed - indexed for MEDLINE]

- Data for stationary phase provided by RegulonDB data was incorrect



# Solution

- Proposed corrections: *ihfA* = +, *ihfB* = +
- References related to *ihfA* and *ihfB*

Expression of the genes coding for the Escherichia coli integration host factor are controlled by growth phase, rpoS, ppGpp and by autoregulation.

[Aviv M. Giladi H. Schreiber G. Oppenheim AB. Glaser G.](#)

Department of Cellular Biochemistry, Hebrew University–Hadassah Medical School, Jerusalem, Israel.

Transcriptional control of the *himA* and the *himD/hip* genes coding for the two subunits of the integration host factor (IHF) was investigated. The promoters for the two genes were identified by the use of primer extension and S1 analysis. Expression from both promoters was found to increase as the cells enter stationary phase. Mutation in *rpoS*, known to be induced upon entry to stationary phase, dramatically reduced the growth-phase response of the *himA* P4 promoter but had only a small effect on the induction of the *himD/hip* promoter. The increased activity of both promoters required the presence of the *relA* and *spoT* genes, suggesting that ppGpp plays a major role in the response to stationary phase. An artificial increase in ppGpp in exponentially growing cells induced a rapid increase in *himA* P4 and *himD/hip* mRNA levels. Experiments with a mutant defective in *rpoS* showed that the response of the *himA* P4 promoter to high ppGpp levels was greatly reduced while that of *himD/hip* was only slightly affected. Therefore, it seems that different mechanisms involving RpoS and ppGpp regulate the growth-phase response of the two promoters. We propose that the effect of ppGpp on *himA* P4 is mediated via RpoS whereas the *himD/hip* promoter is affected by ppGpp independently of RpoS. Expression of the *himD/hip* and *himA* genes was found to be subject to negative autoregulation. IHF-binding sites, implicated in autoregulation, were found to overlap both the *himD/hip* and *himA* P4 promoters. An additional IHF-binding site was found upstream of the *himD/hip* promoter. All three sites show low binding affinity to IHF suggesting that autoregulation can take place only after sufficiently high levels of IHF accumulate in the cell.

PMID: 7715442 [PubMed - indexed for MEDLINE]

- Data for stationary phase provided by RegulonDB data was incorrect

# Solution

- Proposed corrections: *ihfA* = +, *ihfB* = +
- References related to *ihfA* and *ihfB*

**Expression of the genes coding for the Escherichia coli integration host factor are controlled by growth phase, rpoS, ppGpp and by autoregulation.**

[Aviv M.](#), [Giladi H.](#), [Schreiber G.](#), [Oppenheim AB.](#), [Glaser G.](#)

Department of Cellular Biochemistry, Hebrew University–Hadassah Medical School, Jerusalem, Israel.

Transcriptional control of the himA and the himD/hip genes coding for the two subunits of the integration host factor (IHF) was investigated. The promoters for the two genes were identified by the use of primer extension and S1 analysis. Expression from both promoters was found to increase as the cells enter stationary phase. Mutation in rpoS, known to be induced upon entry to stationary phase, dramatically reduced the growth-phase response of the himA P4 promoter but had only a small effect on the induction of the himD/hip promoter. The increased activity of both promoters required the presence of the relA and spoT genes, suggesting that ppGpp plays a major role in the response to stationary phase. An artificial increase in ppGpp in exponentially growing cells induced a rapid increase in himA P4 and himD/hip mRNA levels. Experiments with a mutant defective in rpoS showed that the response of the himA P4 promoter to high ppGpp levels was greatly reduced while that of himD/hip was only slightly affected. Therefore, it seems that different mechanisms involving RpoS and ppGpp regulate the growth-phase response of the two promoters. We propose that the effect of ppGpp on himA P4 is mediated via RpoS whereas the himD/hip promoter is affected by ppGpp independently of RpoS. Expression of the himD/hip and himA genes was found to be subject to negative autoregulation. IHF-binding sites, implicated in autoregulation, were found to overlap both the himD/hip and himA P4 promoters. An additional IHF-binding site was found upstream of the himD/hip promoter. All three sites show low binding affinity to IHF suggesting that autoregulation can take place only after sufficiently high levels of IHF accumulate in the cell.

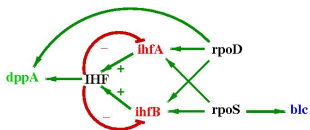
PMID: 7715442 [PubMed - indexed for MEDLINE]

- Data for stationary phase provided by RegulonDB data was incorrect

Consistency of E.coli regulatory network plus sigma-factors with the corrected set of 40 observations for Stationary Phase

# Inference

- Always from a set of known variations



2 consistent solutions:

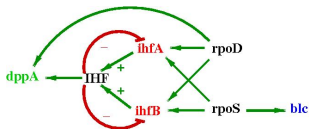
rpoD	rpoS	ihfA	ihfB	IHF	blc	dppA
+	+	+	+	+	+	+
-	+	+	+	+	+	+

Subset of observations:

- ihfA** = +
- ihfB** = +
- blc** = +
- dppA** = +

# Inference

- Always from a set of known variations



## 2 consistent solutions:

rpoD	rpoS	ihfA	ihfB	IHF	bcl	dppA
+	+	+	+	+	+	+
-	+	+	+	+	+	+

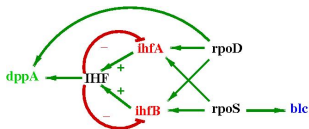
Subset of observations:

- ihfA** = +
- ihfB** = +
- bcl** = +
- dppA** = +

- Inference :
  - No inferred value for **rpoD**
  - rpoS** = +
  - IHF** = +

# Inference

- Always from a set of known variations



## 2 consistent solutions:

rpoD	rpoS	ihfA	ihfB	IHF	bcl	dppA
+	+	+	+	+	+	+
-	+	+	+	+	+	+

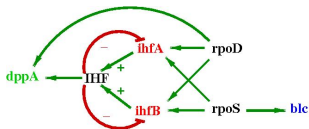
Subset of observations:

- $ihfA = +$
- $ihfB = +$
- $bcl = +$
- $dppA = +$

- Inference :
  - No inferred value for  $rpoD$
  - $rpoS = +$
  - $IHF = +$

# Inference

- Always from a set of known variations



2 consistent solutions:

<i>rpoD</i>	<i>rpoS</i>	<i>ihfA</i>	<i>ihfB</i>	<i>IHF</i>	<i>bcl</i>	<i>dppA</i>
+	+	+	+	+	+	+
-	+	+	+	+	+	+

Subset of observations:

- ihfA* = +
- ihfB* = +
- bcl* = +
- dppA* = +

- Inference :
  - No inferred value for *rpoD*
  - rpoS* = +
  - IHF* = +

# Inferring variations of new products of E.coli network



- With **40** observations of Stationary Phase were found  $5.2 \cdot 10^{16}$  solutions
- From these, **381** (26%) variables maintain always the same sign

# Inferring variations of new products of E.coli network



42 (of 381) products inferred under stationary phase condition

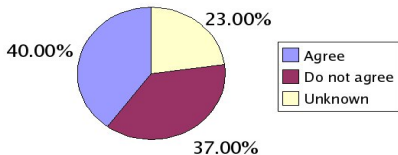
gene	value	gene	value	gene	value	gene	value
IHF	+	cpxR	+	fucR	+	lysR	+
ada	+	crp	+	fur	+	melR	+
agaR	+	cusR	+	galR	+	mngR	+
alsR	+	cynR	+	gcvA	+	oxyR	+
araC	+	cysB	+	glcC	+	phoB	+
argP	+	cytR	+	gntR	+	prpR	+
argR	+	dnaA	+	ilvY	+	rbsR	+
baeR	+	dsdC	+	iscR	+	rhaR	+
cadC	+	evgA	+	lexA	+	rpoD	+

gene	value
rpoS	+
soxR	+
soxS	+
srlR	+
trpR	+
tyrR	+



# Comparing the predictions with MA results

- Source of Microarray data: GEO (Gene Expression Omnibus)  
Barret *et al.*, Nucleic Acid Res, 33. 2005
- Comparison between 2 sets of variations of genes:
  - Microarray data for stationary phase after 20 minutes: 4025
  - Inferred variations: 381
- Number of common genes for both sets: 292
- Comparison results for the 381 inferred variations:



- Possible reasons of disagreement:
  - Microarray data related to small variations of certain genes
  - Missing interactions in our model

# Conclusion

- We were able to propose methods and algorithms that can handle automatically (PYQUALI) **large scale graphs**
- **What we are able to do**
  - Test the **self-consistency** of a network
  - Check its **consistency** with a set of **experimental data**
  - **Propose corrections** to the model or to the experimental data
  - **Infer new variations** of non-observed products
- **Future work**
  - Find the **best set of observations** to validate a model
  - Find the **best set of observations** to infer the most
  - Propose methods to **complete a network** (adding metabolites)

Thank you!