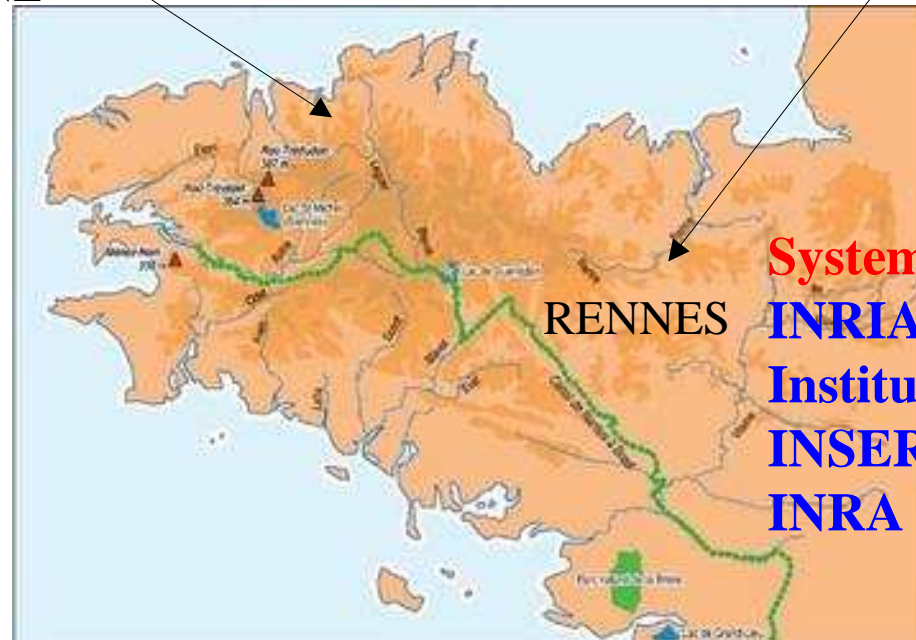


Improving model consistency by qualitative equations

Ovidiu Radulescu, IRMAR and
Symbiose project IRISA



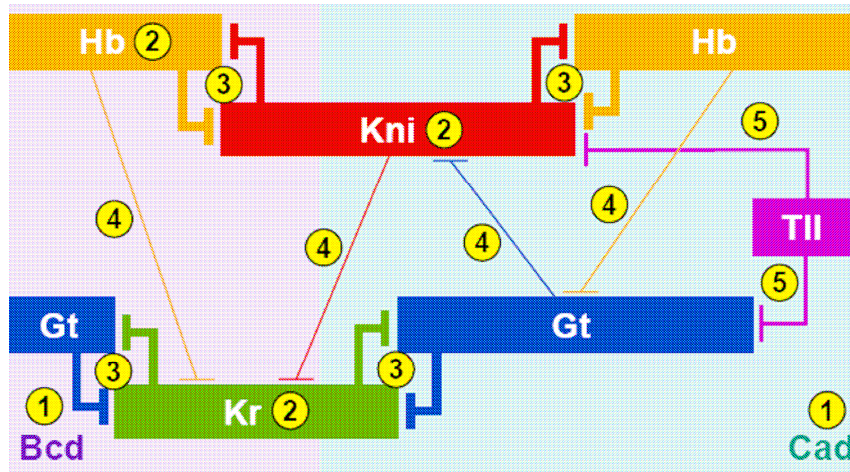
LA BRETAGNE



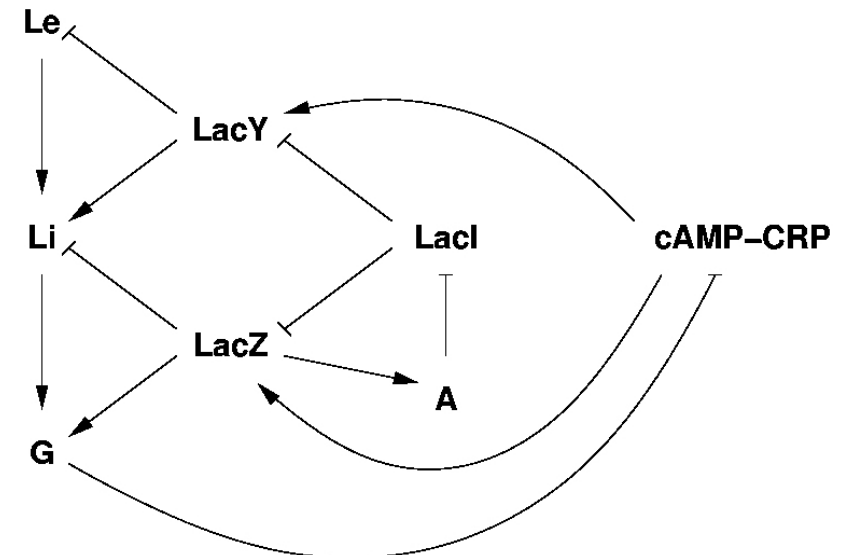
Systems Biology in Rennes:
INRIA (IRISA)
Institute of Mathematics
INSERM
INRA

Summary

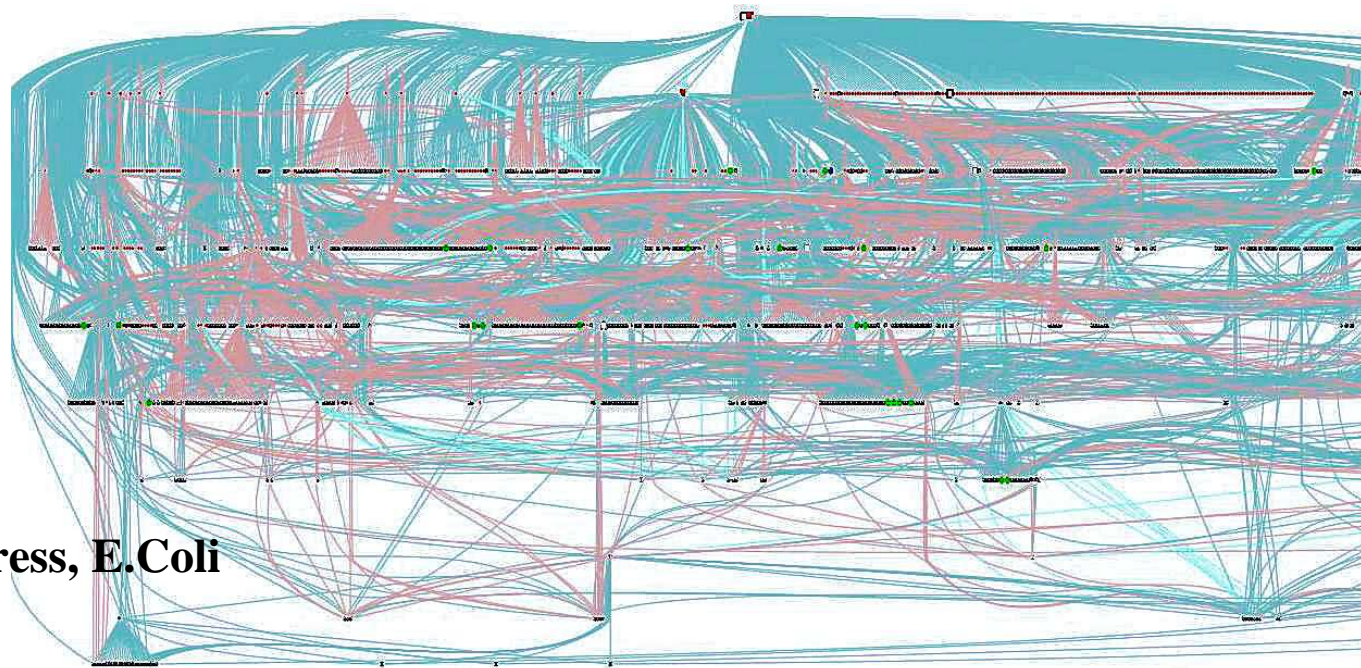
- Large scale models and data
- Qualitative equations
- Prediction and correction, a case study : nutritional stress in *Escherichia Coli*
- Reverse engineering : interaction sign inference
 - ✓ Feasability study : random solutions, *Escherichia Coli* network
 - ✓ Application : transcriptional regulatory network in *Saccharomyces cerevisiae*
- Conclusion



Gap genes, first 3 hours of Drosophila



Lactose operon, E.Coli



Nutritional stress, E.Coli

Network models : oriented graphs + signs

Knowledge databases : RegulonDB, KEGG, Ingenuity, BioBase, etc.

Transcriptional interactions : ChIP/chip, promoter analysis

Interaction signs : well controlled experiments

GARDON: our database (lipid metabolism regulation)

The image displays three screenshots of the GARDON database interface, illustrating the workflow for adding a new interaction from a literature article.

Left Screenshot: Literature main menu

- Navigation: File, Edit, View, Go, Bookmarks, Tools, Help.
- User: Test user (Change).
- Main menu: Litterature main menu.
- Queries: One product, Several products.
- Litterature: New litterature reference.
- Interactions: New interaction without article, Interactions list.
- Action types: New action types, Action types list.
- Products: New product, Products list, Update products.
- Product types: New product type, List product types.
- Footer: 239 articles, 197 used, 1884 interactions, 657 products.

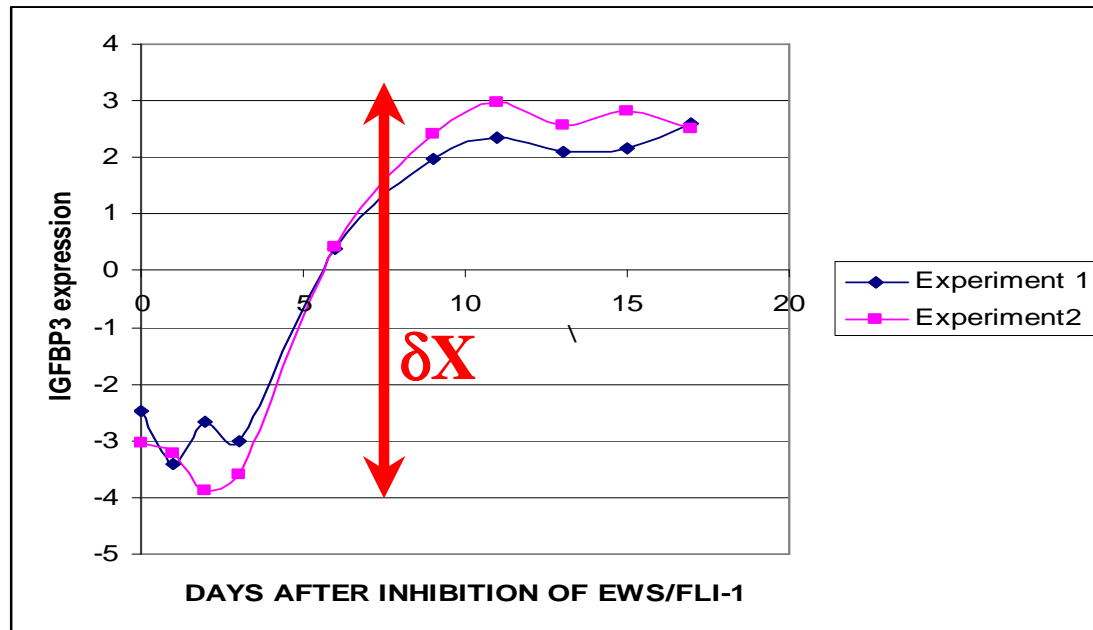
Middle Screenshot: New Article

- User: Test user.
- Main Menu > New Article.
- New Article: Please enter a PubMed article.
- PubMed ID: 14770367.
- Open a PubMed window.
- Next: Main Product >>.
- Cancel.
- Article details: PMID: 14770367, Med Sci (Paris), ISSN: 0767-0974, 20 Jan 2004, 73-7. Title: [Is the ileal bile acid-binding protein (I-BABP) homeostasis?]. Authors: Besnard Philippe P Landrier Jean-François JF Grober.

Right Screenshot: Interaction

- User: Test user.
- Main Menu > New Article > Main Product > Interaction.
- Interaction: Choose an interaction described in the article.
- Action Type: Action on interaction (behavioral, bio-chemical).
- Details: Article ID (14770367), Species, Input (Target, Factor), Output (Regulation mode, Observation Level, Confidence, Comment), Tissue, Cell type, Context, Location, Link.

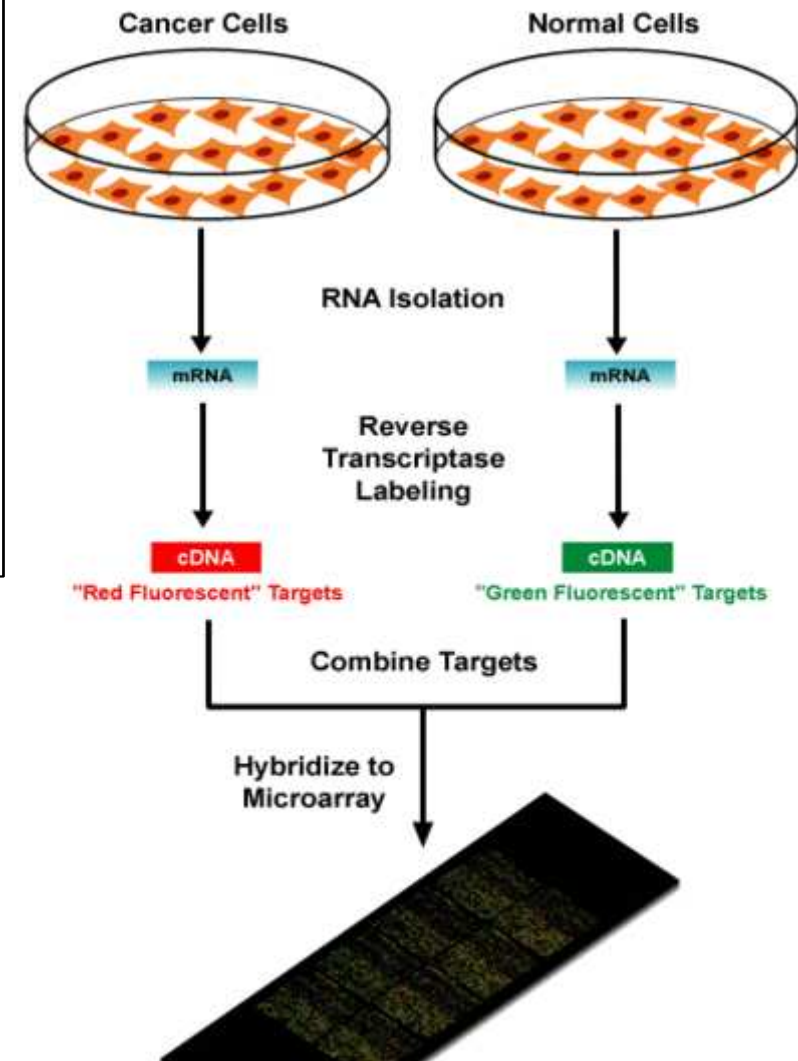
Time series, switch-like response



$|\delta \log X| < \theta, \text{ sign}=0$
 $\delta \log X > \theta, \text{ sign}=+$
 $\delta \log X < -\theta, \text{ sign}=-$
else sign=?

Qualitative transcriptional data: signs of differences of concentrations between two states

Differential data



Exploit model-data consistency

Data prediction:

- Propose gene variation signs consistent with a model and with relatively reliable data

Data correction:

- Detect and correct false positives and false negatives

Model reverse engineering:

- Do not start from scratch
- Infer signs of interactions
- Infer new interactions

Experiment design:

- Guide data collection

Qualitative equations: define consistency

Biological problem:

Following a perturbation (stress, signal) the state of the cell changes. Variations of hundreds or thousands of variables can be monitored. How to use this information?

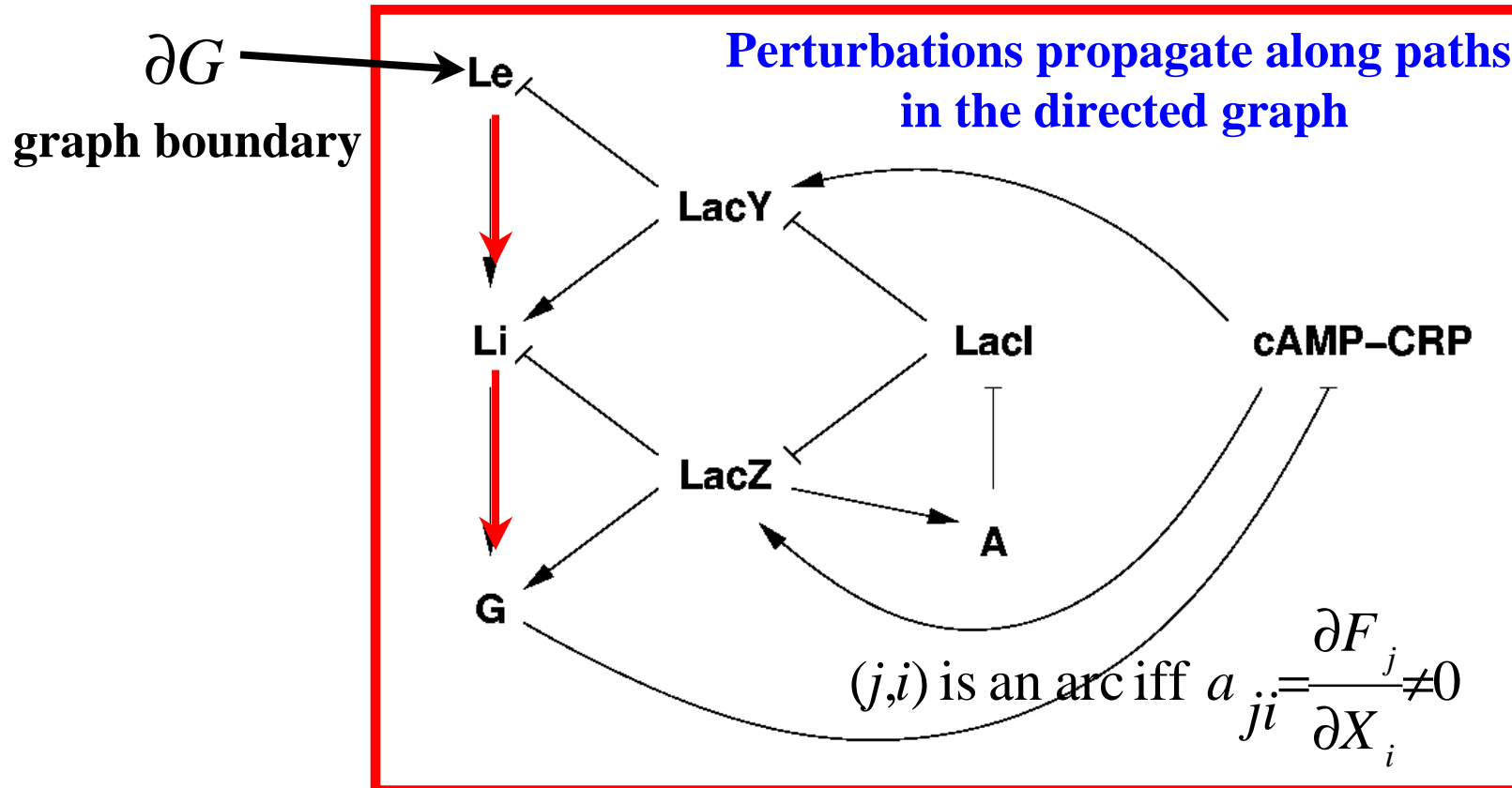
Steps:

- develop an “elasticity” theory of graphs (O.Radulescu et al. J.R.Soc.Interface 2006)
- translate this theory into qualitative equations (with A.Siegel et al. Biosystems 2006)
- polynomial algorithms for solving systems of qualitative equations (with Ph.Veber, M.leBorgne et al. Complexus 2006)
- application to large networks (E.Coli with C.Guziolowski et al., proc. RIAMS 2006, S. Cerevisiae with Ph.Veber, C.Guziolowski in work)

Elasticity of graphs

$$\frac{dX}{dt} = F(X, P) \quad \text{dynamics} \quad F(X, P) = 0 \quad \text{Steady state equation}$$

Steady state is perturbed $\delta P \rightarrow \delta X$



Dirichlet solution:

$$\delta X_i = \sum_{j \in \partial G} \sum_{j \rightarrow i} \frac{a_{j \rightarrow i}}{C_{j \rightarrow i}} \delta X_j$$

(O.Radulescu et al. J.R.Soc.Interface 2006)

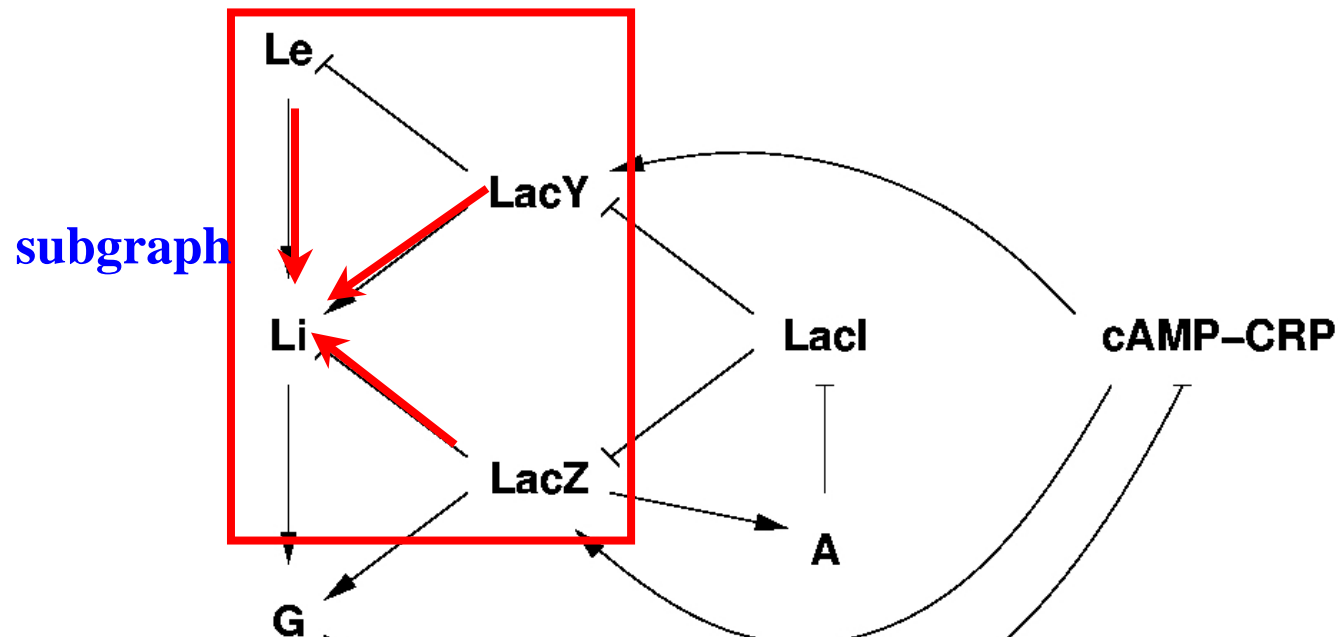
Qualitative equations

$$\delta X_i = - \left(\frac{\partial F_i}{\partial X_i} \right)^{-1} \sum_{j \in \text{pred}(i)} a_{ji} \delta X_j \quad \text{Dirichlet solution for subgraph}$$

$$\text{sign}(\delta X_i) = \sum_{j \in \text{pred}(i)} \text{sign}(a_{ji}) \text{sign}(\delta X_j) \quad \text{Qualitative equation}$$

$$\text{sign} \in \{-, +, ?\} \quad \text{Sign algebra}$$

$$\text{Li} = \text{Le} + \text{LacY} - \text{LacZ}$$

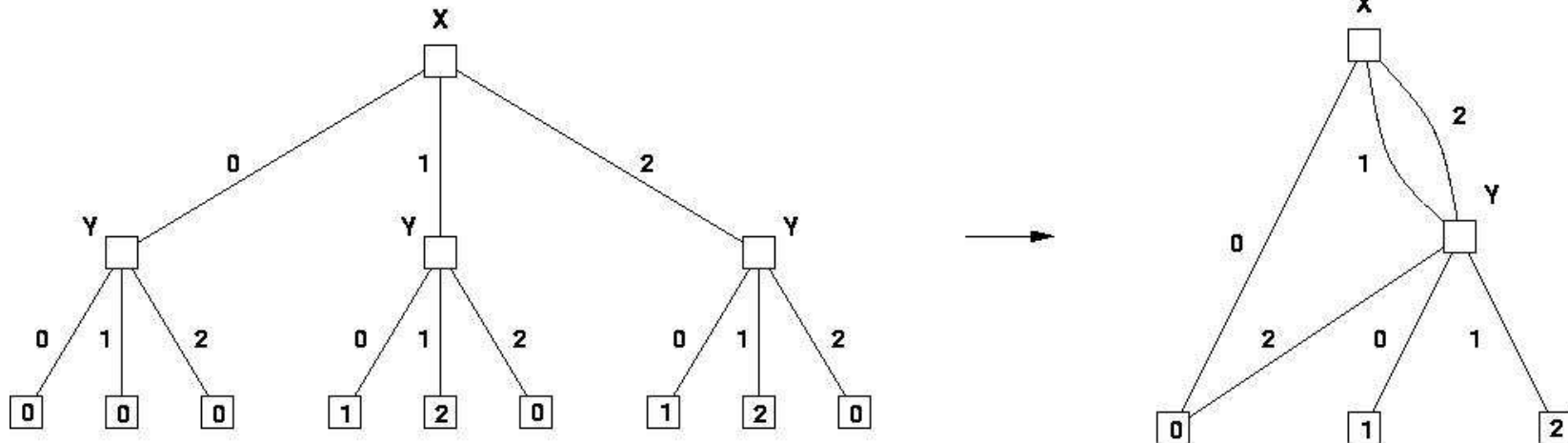


+	+	-	=	?	+	+	+	=	+
+	x	-	=	-	+	x	+	=	+
?	+	-	=	?	?	+	+	=	?
?	x	-	=	?	?	x	+	=	?

\approx	+	-	?
+	T	F	T
-	F	T	T
?	T	T	T

Algorithm for solving qualitative equations

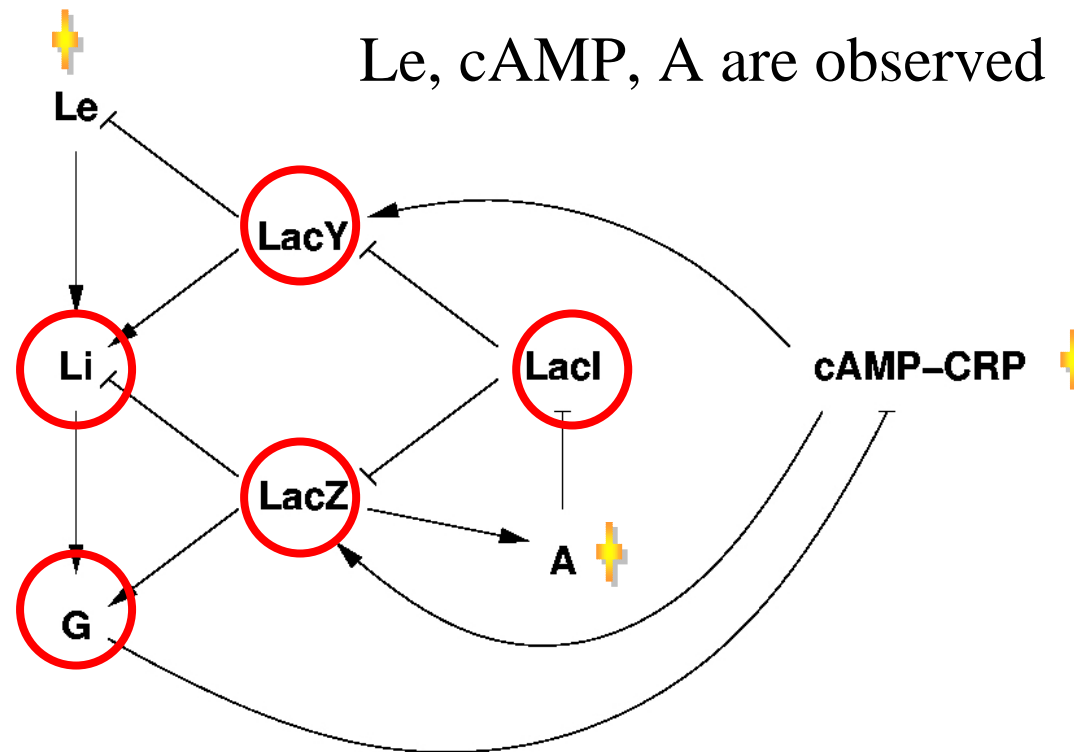
- Map signs to elements of the finite field $\mathbb{Z}/3\mathbb{Z}$
- Map qualitative equations to polynomial equations over $\mathbb{Z}/3\mathbb{Z}$
- NP complete problem
- Ternary Decision Trees contracted to directed acyclic graphs and systematic use of cache memory for non-redundant computation
- Obtain exhaustive lists of solutions within minutes for 1000 nodes



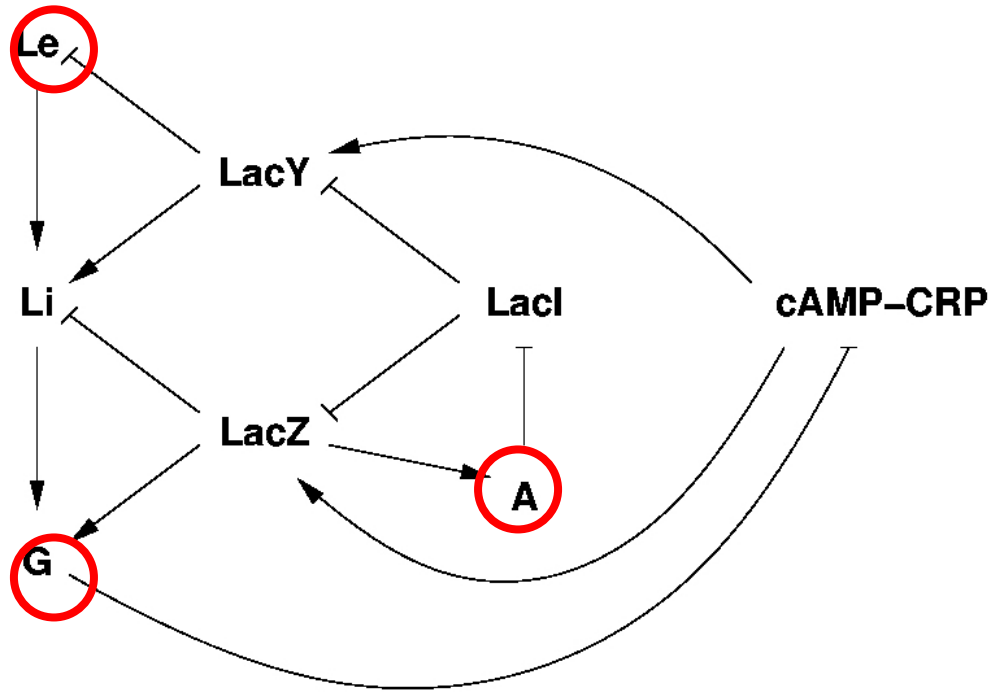
Predictions of a model

hard components: variables whose values are the same (+ or -) in any solution

the hard components are the predictions of the model



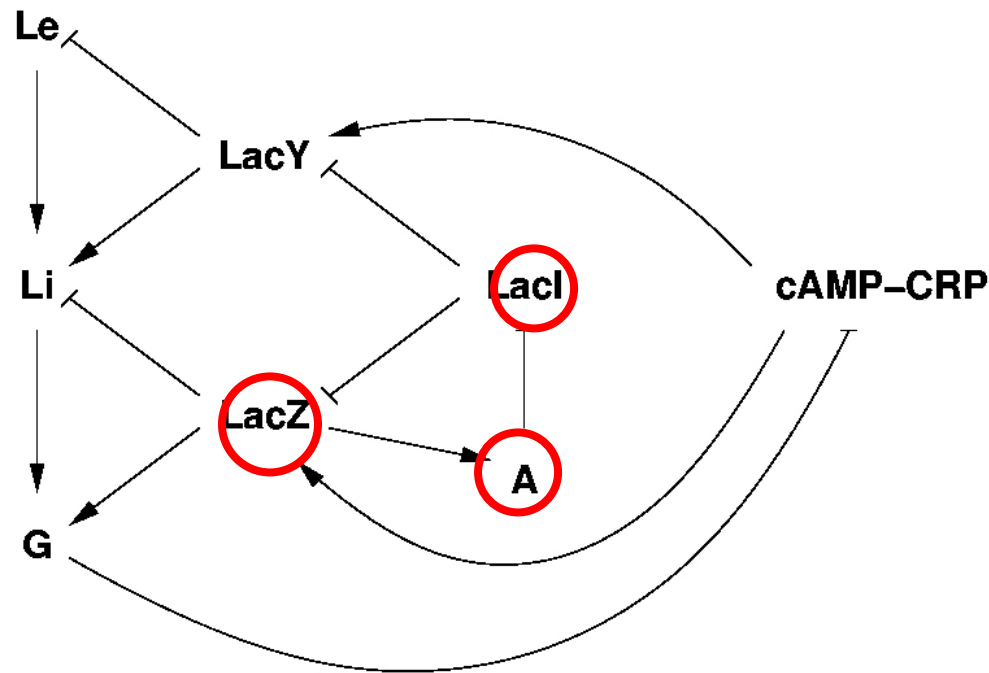
Experiment design



Any value of the triplet (Le,G,A) can be extended to a solution

These variables have no validation power

Use validation power for experiment design



Only 2 values (out of 8) of (LacI,A,LacZ), namely (+,-,-) (-,+,+) can be extended to a solution

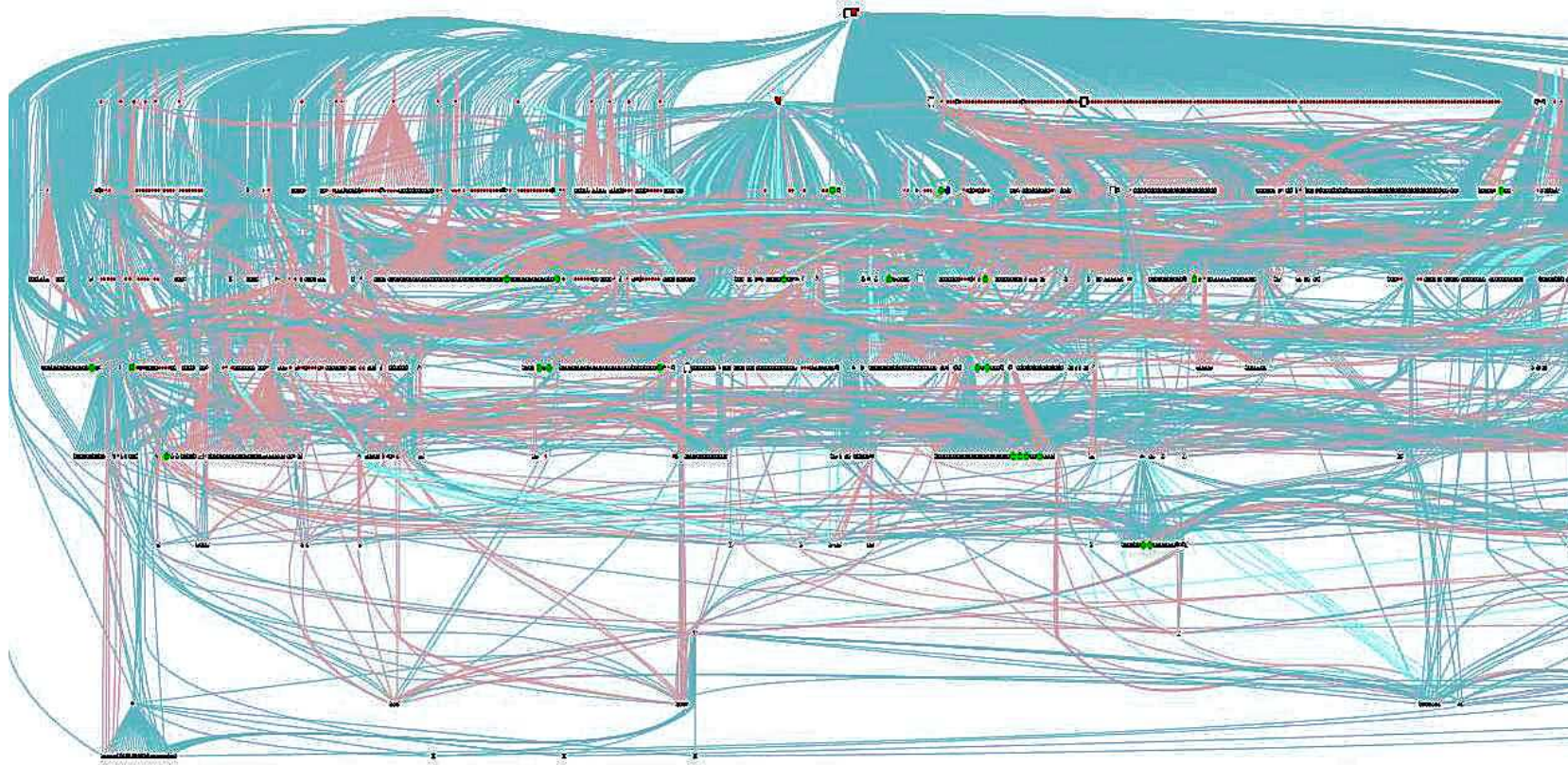
Define validation power as:

$$\tau(X_1, \dots, X_p) = 1 - \frac{\text{val}(X_1, \dots, X_p)}{2^p}$$

Choose high validation power sets for optimal design

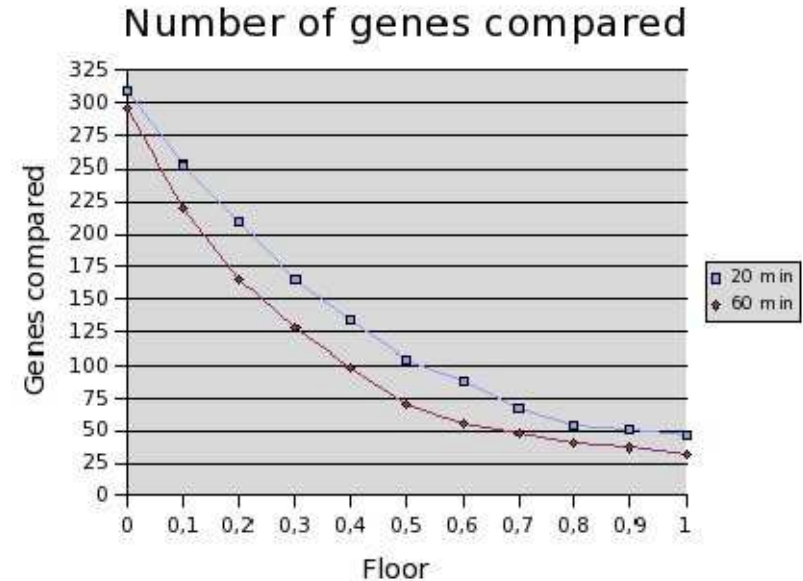
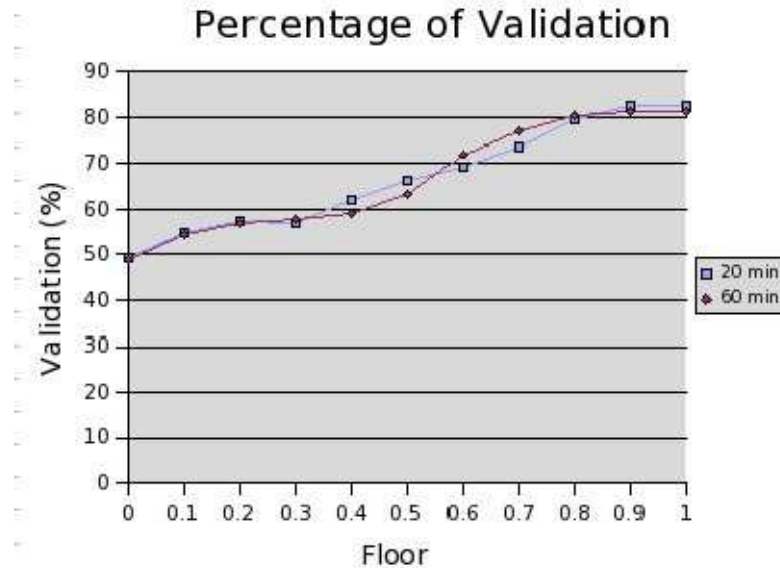
Large scale application: nutritional stress of E.Coli

1258 nodes, 2526 interactions, 10^{600} states, 10^{16} solutions
Source: RegulonDB (March 2006)

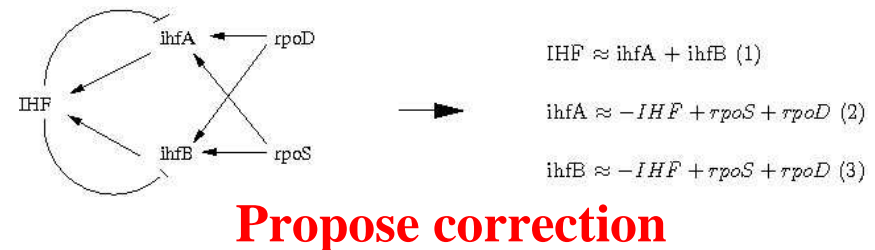
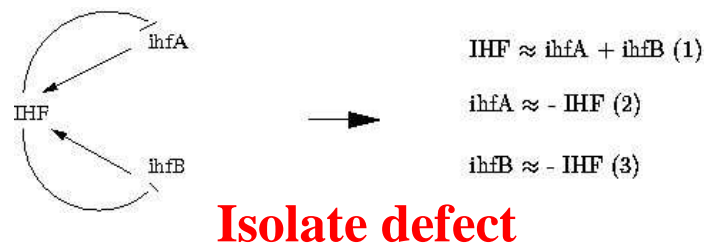


We have obtained:

- a set of predictions: from 40 observations in the stationary phase, 401 hard components, 26% of the network; these were compared to expression data



- a set of corrections to the model



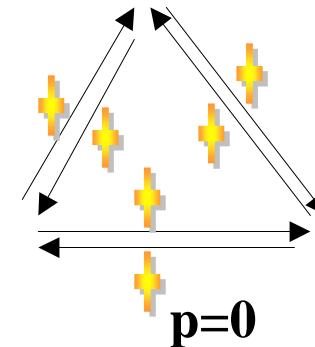
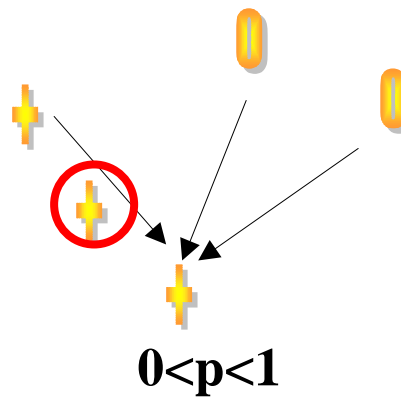
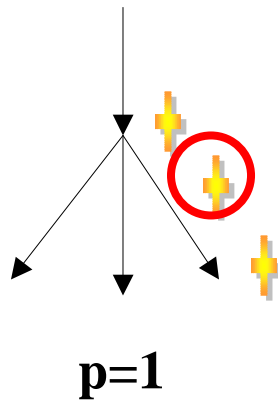
- a set of corrections to data: minimality with respect to Hamming distance

Reverse engineering : interaction sign inference

Problem: ChIP/chip data provide unsigned interactions. Which proportion of signs can be inferred from expression data? Use QEs with interaction signs as variables.

Numerical feasibility study: Generate random, consistent observations, from them compute hard components (predicted signs).

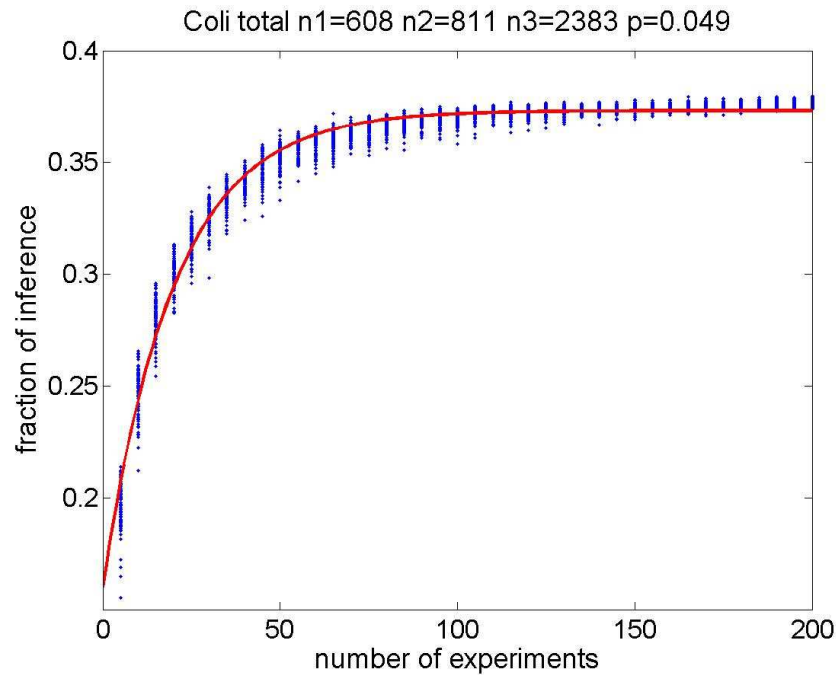
Analytical approximate guess: three classes of interactions.
 p =probability of inference in one experiment



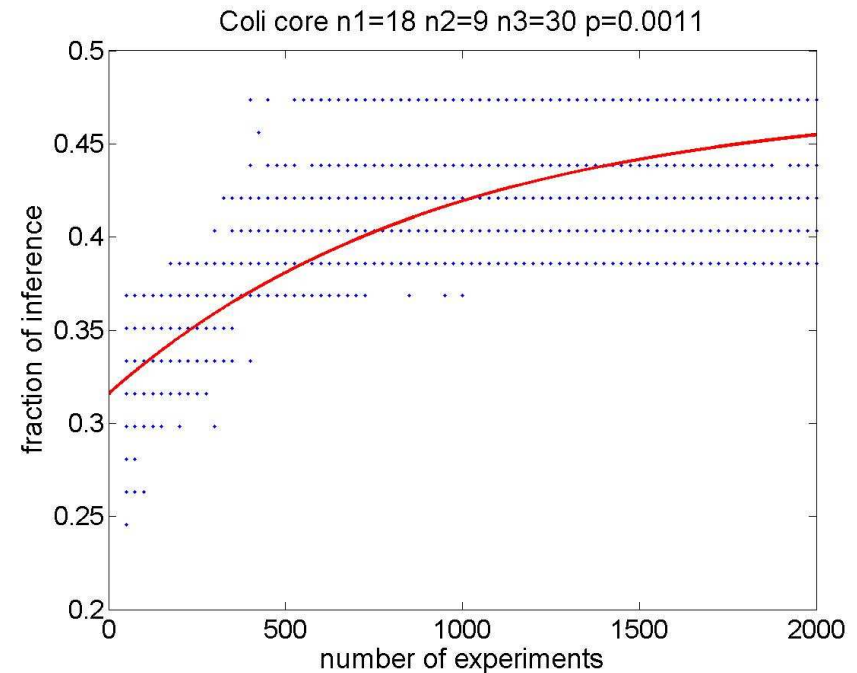
$$n = n_1 + n_2 \left[1 - (1-p)^n \exp \right]$$

Sign reverse engineering : numerical feasibility study

**type of experiments : generate gene and TF variations
compatible with the network constraints**



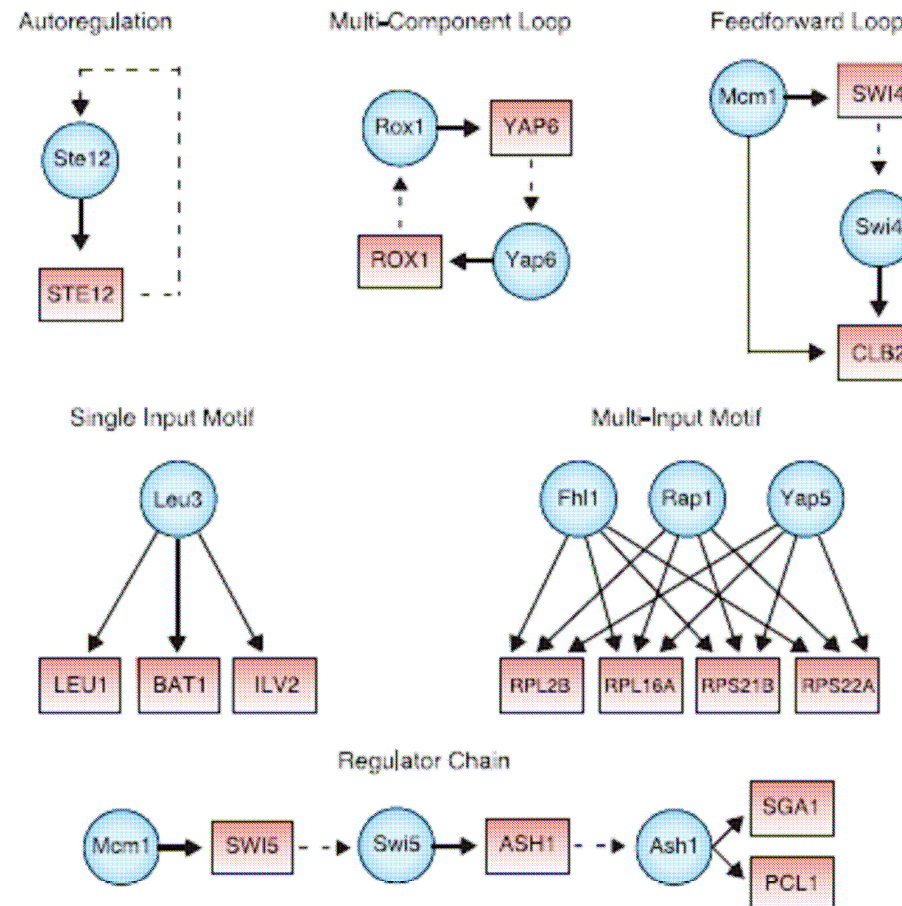
**Total network : need few
experiments, randomly chosen**



**Core network : need more
experiments, eventually well
chosen; experiment design useful**

Transcriptional regulation network in *S.Cerevisiae*

Sources: Lee et al. (2002), MacIsaac et al. (2005), Harbinson et al. (2004)

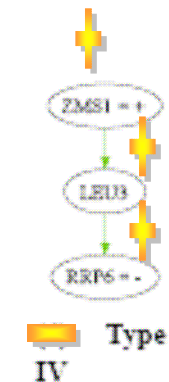
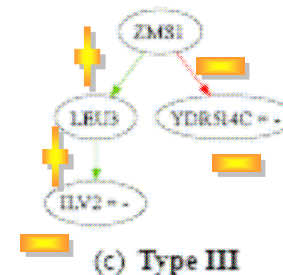
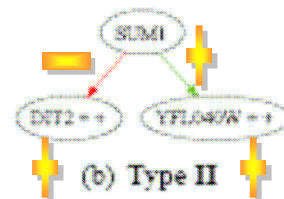
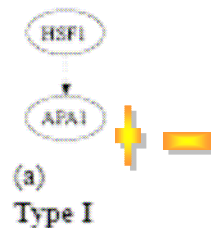


global network (at p-value 0.001, 2419 nodes, 4344 interactions) and **core network** (at p-value 0.001, 70 nodes, 96 interactions)

Inferred signs and defects

Interaction network	Nodes	Edges	Number of Experiments	Input/Output obs. simult.	Inferred signs	MBM Int. Type I	MBM Int. Type II, III, IV	Total Inf. rate
Core of Lee transcriptional network <i>Lee et al. 2002, Kauffman et al. 2003</i>	31	52	15	46	11 (21.1%)	3 (5.7%)	0	26.8%
Extended Lee transcriptional network <i>Lee et al. 2002</i>	70	96	15	70	29 (30.2%)	7 (7.2%)	0	37.4%
Global transcriptional network <i>Lee et al. 2002, p-value = 0.001</i>	2419	4344	14	2270	631 (14.5%)	281 (6.5%)	463 (11%)	32%
Inferred network <i>MacIsaac et al. 2005, Harbinson et al. 2004</i> threshold = 0.001 ; bindings=2	83	131	14	91	21 (16%)	4 (3%)	0	19%

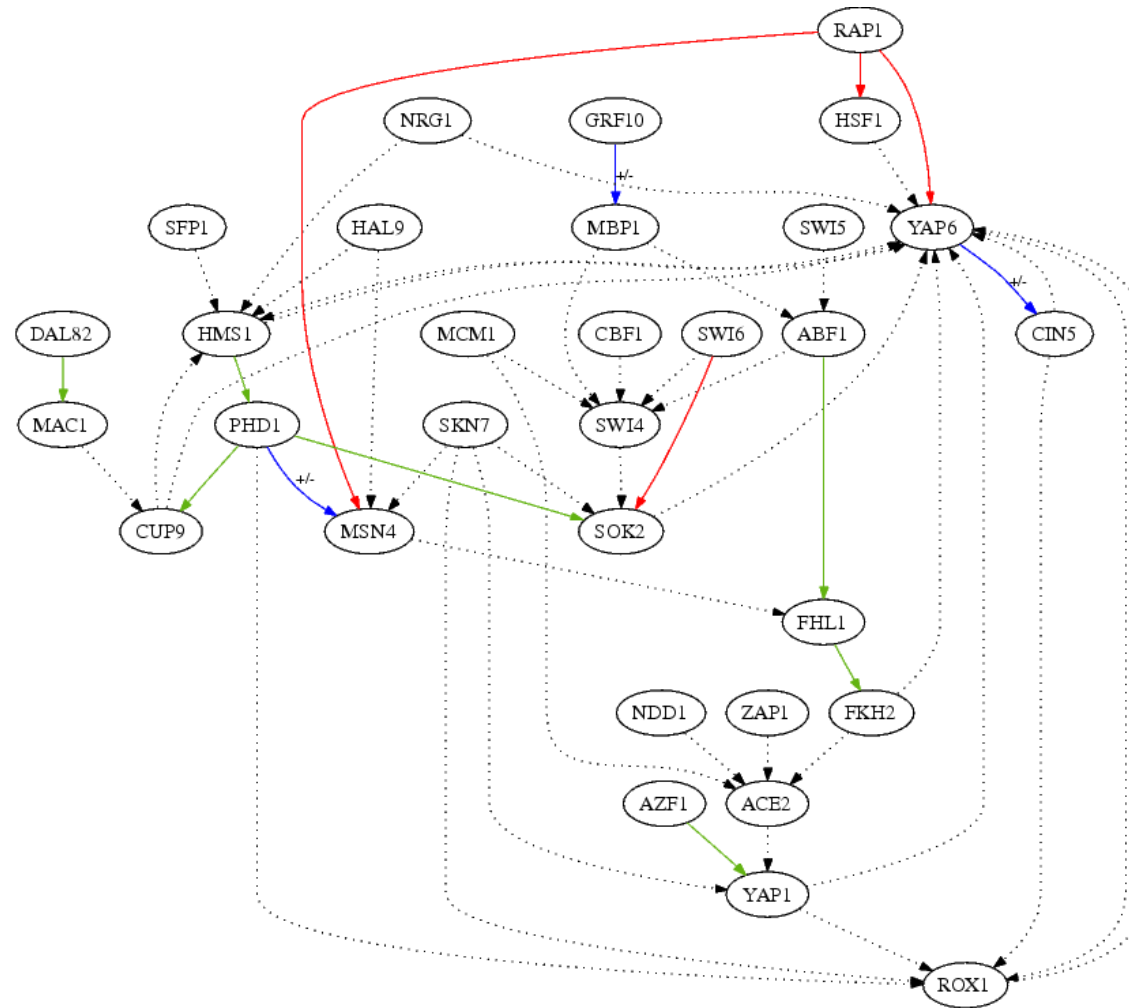
Types of defects



Iterations

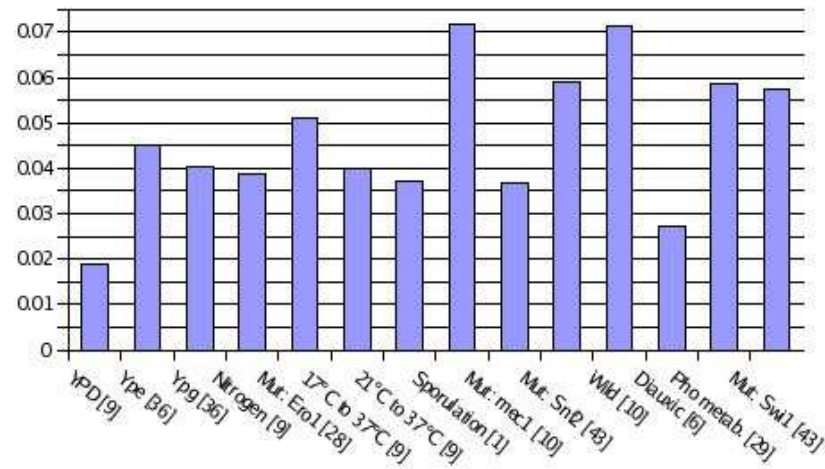
short range defects → inference → longer range defects

Inferred signs for core of Lee et al. (2002)

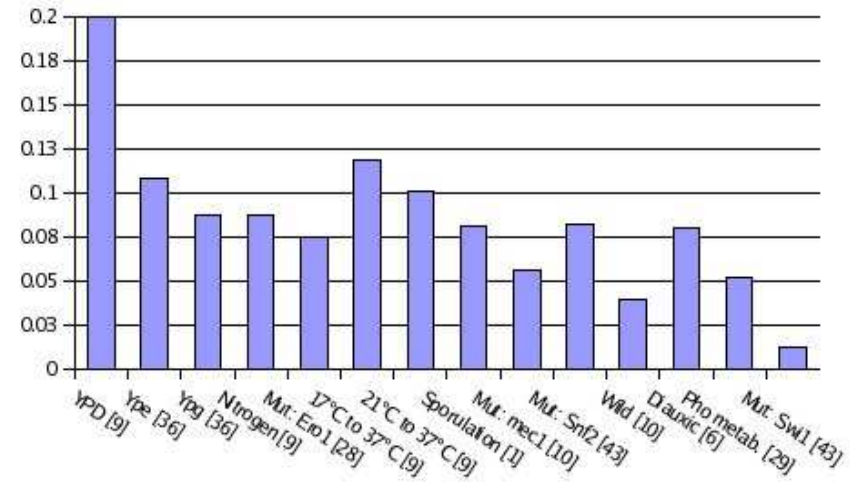


Inference per type of experiment

Eliminated roles per Experiment by observed nodes



Inferred roles per Experiment by observed node



YPD: Broth to stationary phase Gash et al 2000

Conclusion

- Qualitative equations suitably describe consistency between data and models
- Our rules are over-approximations of boolean rules: they produce robust predictions
- Sign reverse engineering is possible with a small number of experiments (10-15) ; nevertheless limits exist, specially for core network ; use experiment design (choice of experiments)
- Future work:
 - ✓ Diagnosis: better characterization of defects, propose corrections
 - ✓ Experiment design, useful for core network
 - ✓ Model other types of data : project in cancer genomics CGH, microRNA silencing, expression time series

Acknowledgements

Symbiose team, IRISA RENNES

Michel le Borgne

Anne Siegel

Carito Guziolowski

Philippe Veber