Regulatory controls of the metabolism and their inference with MERRIN

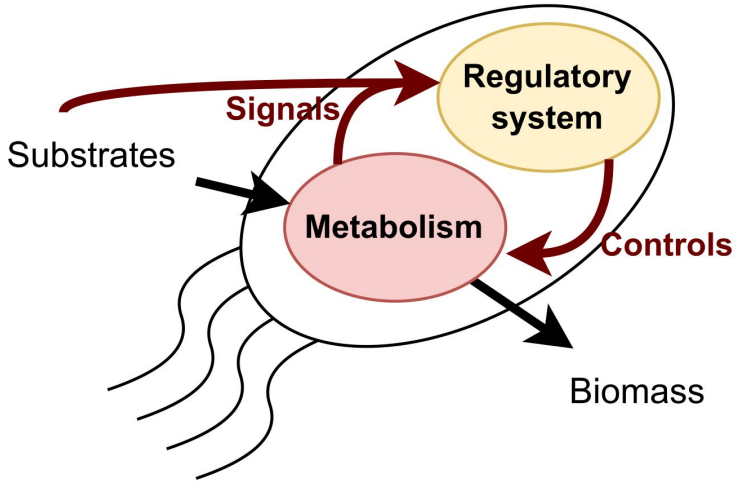
Kerian Thuillier
Université Paris-Saclay, CNRS, ENS Paris-Saclay, Laboratoire Méthodes Formelles, 91190 Gif-sur-Yvette, France

Anne Siegel[1], Loïc Paulevé[2], Caroline Baroukh[3], Ludovic Cottret[3], Alexander Bockmayr[4]

[1] Université de Rennes, Inria, CNRS, IRISA, 35000 Rennes, France
[2] Université de Bordeaux, Bordeaux INP, CNRS, LaBRI, 33400 Talence, France
[3] LIPME, INRAE, CNRS, Université de Toulouse, Castanet-Tolosan, France
[4] Freie Universität Berlin, Institute of Mathematics, D-14195 Berlin, Allemagne

Cells: hybrid multi-scale structures

*Composed of thousands of **interconnected** chemical processes*
*Occurring at different **scales***

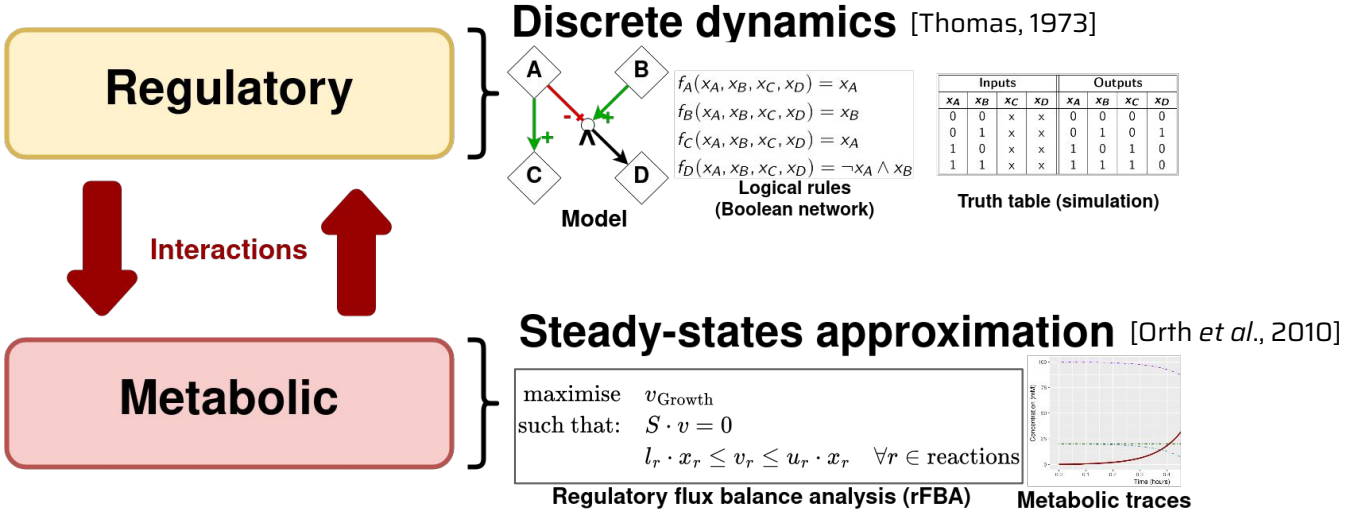


1. **Metabolic scale**
   Chemicals reactions converting substrates to energy and biomass

2. **Regulatory scale**
   Rules constraining the metabolism to adapt itself to its environment

Two scales of interest: metabolic and regulatory

Overview of modeling formalisms

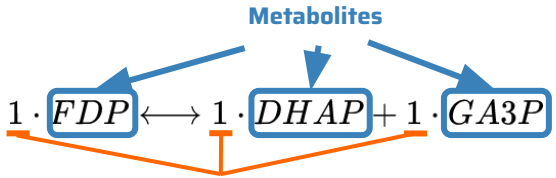


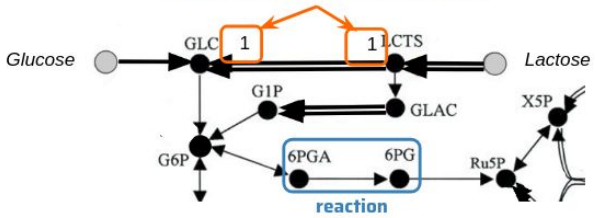
Two scales model based on different paradigms and formalisms

Structure

Metabolic scale

Set of chemical reactions

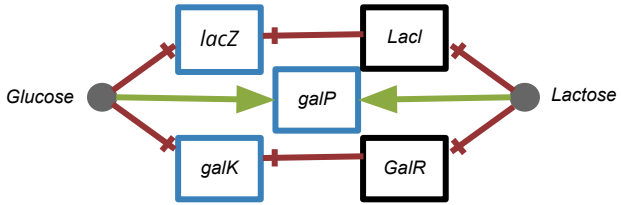




Metabolic networks

Regulatory scale

Interactions graph



Boolean network [Thomas, 1973]

Logical combination of interactions

$$f_{\text{lacZ}}(x) = \neg x_{\text{Glucose}} \wedge \neg x_{\text{LacI}} \qquad f_{\text{LacI}}(x) = \neg x_{\text{Lactose}}$$
$$f_{\text{galP}}(x) = x_{\text{Glucose}} \vee x_{\text{Lactose}} \qquad f_{\text{galK}}(x) = \neg x_{\text{Glucose}} \wedge \neg x_{\text{GalR}}$$
$$f_{\text{GalR}}(x) = \neg x_{\text{Lactose}}$$

Set of logical rules paired with an directed labeled graph

Inspired by [Covert and Palsson, 2002]

Dynamics

Metabolic scale

Flux balance analysis[1] (FBA) [Orth *et al.*, 2010]

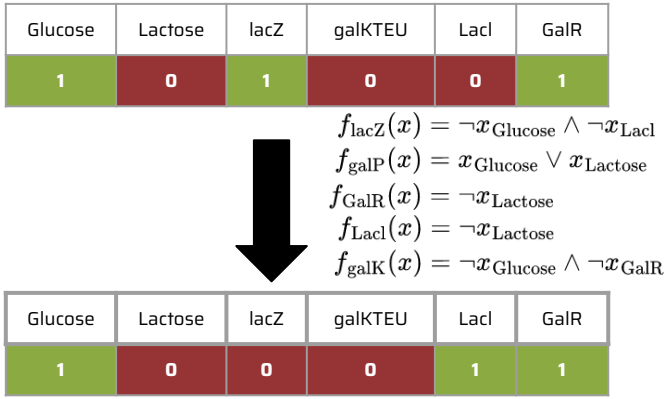
$$\begin{aligned} \text{maximize} \quad & v_{\text{Growth}} \\ \text{such that:} \quad & S \cdot v = 0 \\ & l_r \leq v_r \leq u_r \quad \forall r \in \text{reactions} \end{aligned}$$

Bacteria growth maximization [Feist and Palsson, 2010]

Based on heuristics: growth optimization + steady-state

Flux-based dynamics

Regulatory scale



Discrete dynamics [Thomas, 1973]

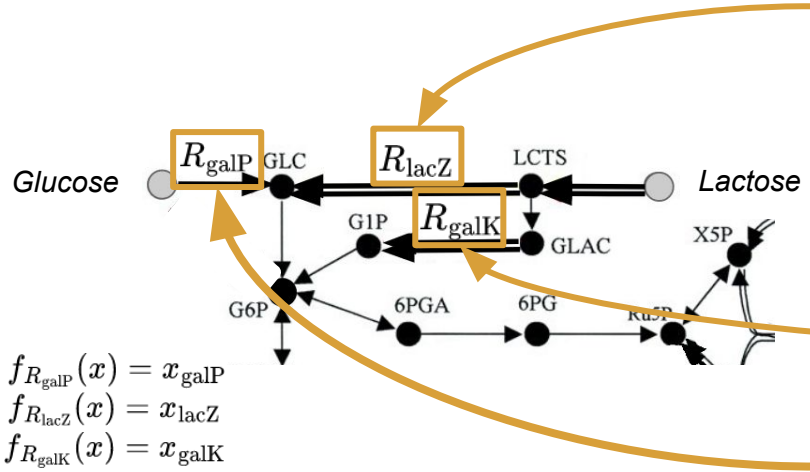
Various update semantics

Scale dynamics are based on different paradigms

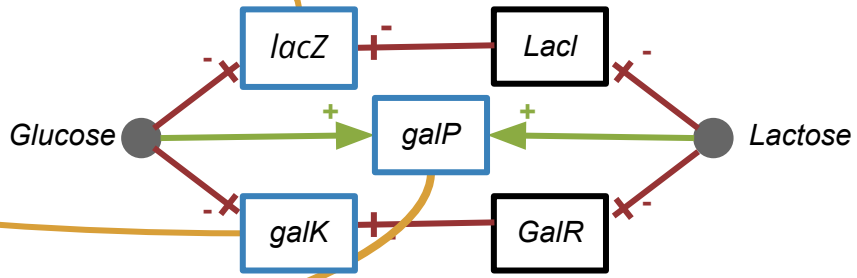
No straightforward formalism to encompass them

Coupling the scales

Metabolic scale



Regulatory scale

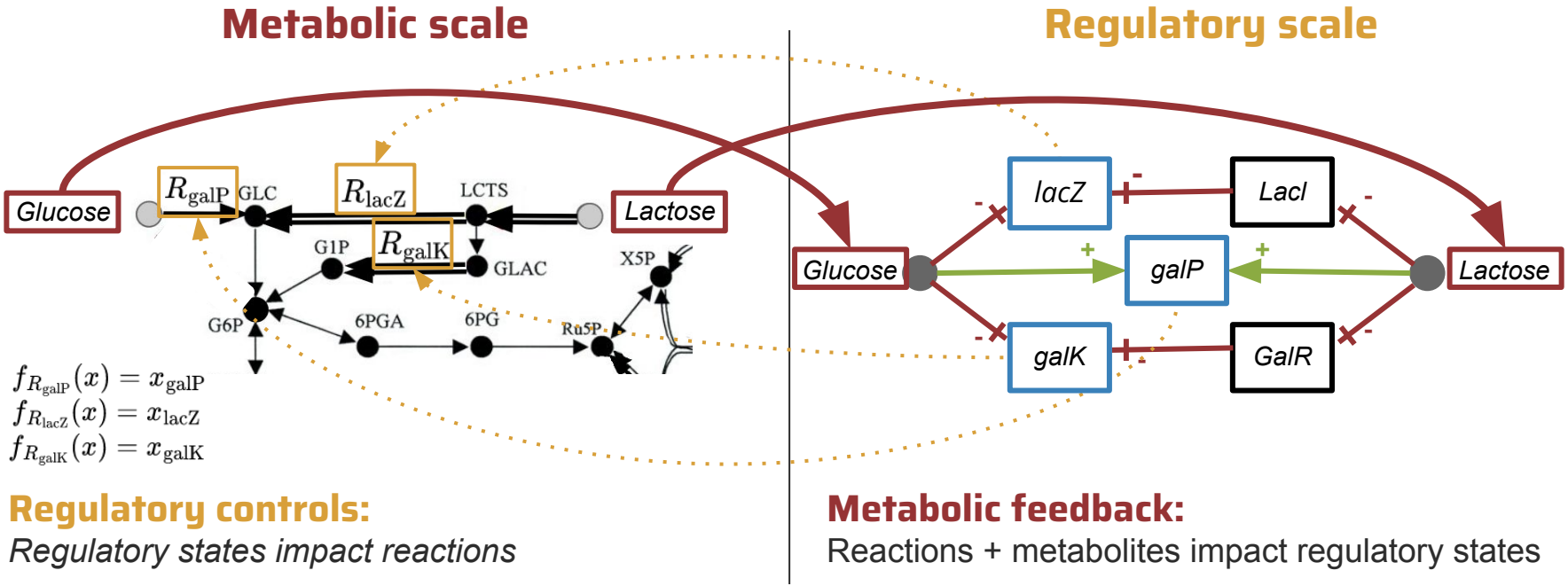


Regulatory controls:
Regulatory states impact reactions

Metabolic feedback:
Reactions + metabolites impact regulatory states

Interconnected scales through regulatory controls and metabolic feedback
Simulating the coupled dynamics through regulatory Flux Balance Analysis (rFBA) **[Covert *et al.*, 2001]**

Coupling the scales



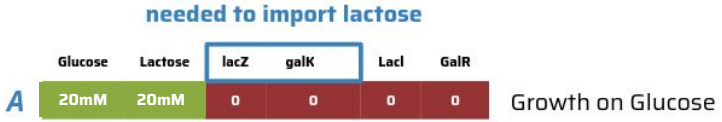
Regulatory controls:
Regulatory states impact reactions

Metabolic feedback:
Reactions + metabolites impact regulatory states

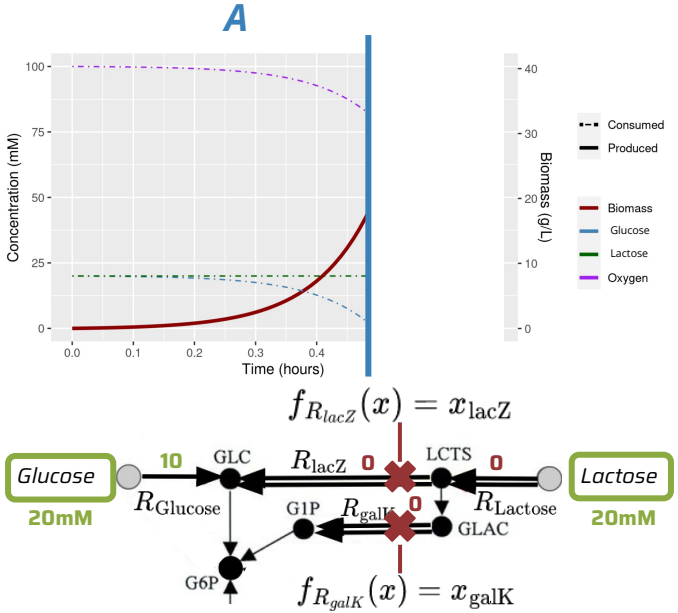
Interconnected scales through regulatory controls and metabolic feedback
Simulating the coupled dynamics through regulatory Flux Balance Analysis (rFBA) **[Covert *et al.*, 2001]**

Example of controlled induced behavior: diauxic shift[1]

Successives growth phases on different mediums
Controlled by the regulatory scale



rFBA simulation made with FlexFlux [Marmiesse *et al.*, 2015]

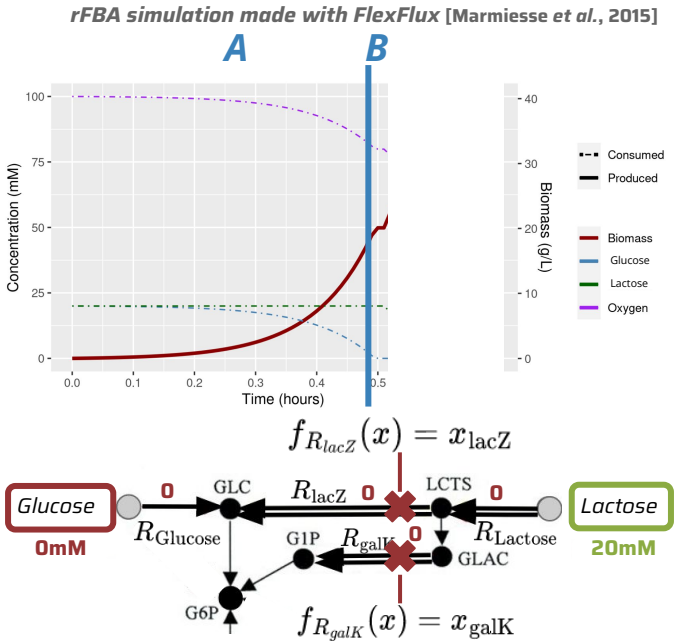


Phase A: **lactose** ***could not be imported due to regulatory rules***

[1] J. Monod, **Annales de l'Institut Pasteur**, 1942

Example of controlled induced behavior: diauxic shift[1]

Successives growth phases on different mediums
Controlled by the regulatory scale

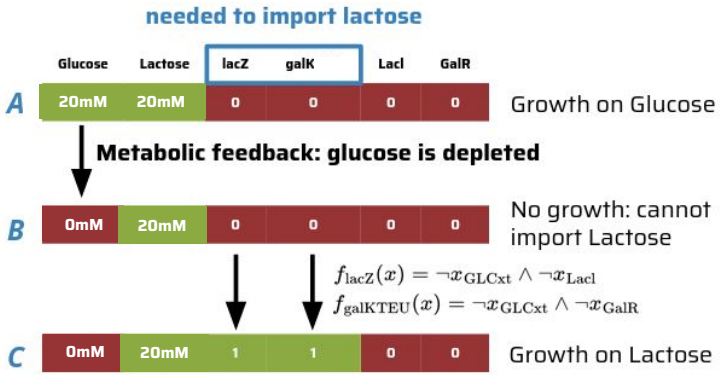


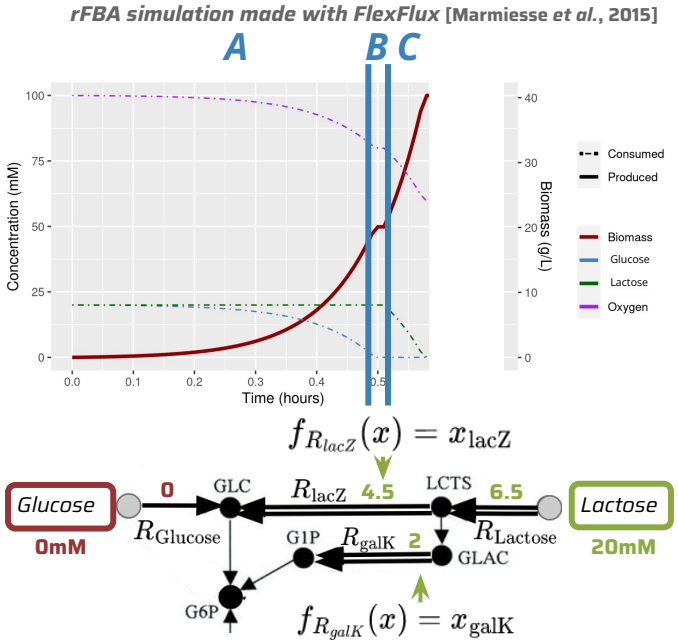
Phase B: ***regulatory mechanisms are slow and need time to react to* glucose *depletion***

[1] J. Monod, **Annales de l'Institut Pasteur**, 1942

Example of controlled induced behavior: diauxic shift[1]

Successives growth phases on different mediums
Controlled by the regulatory scale



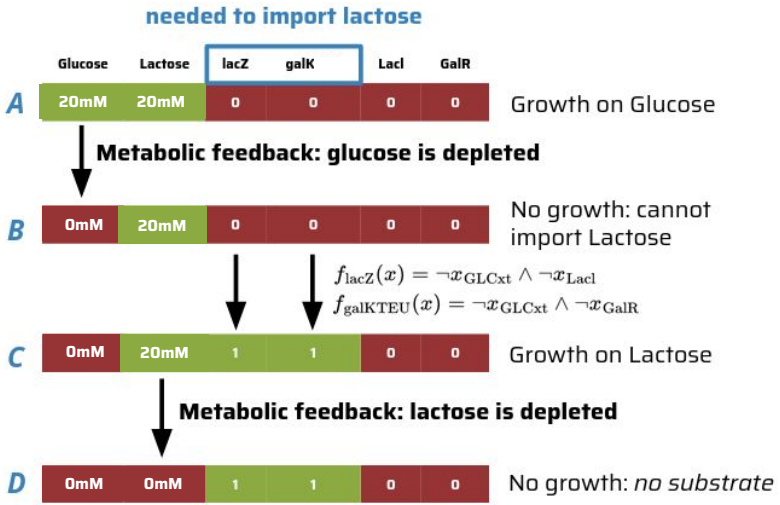


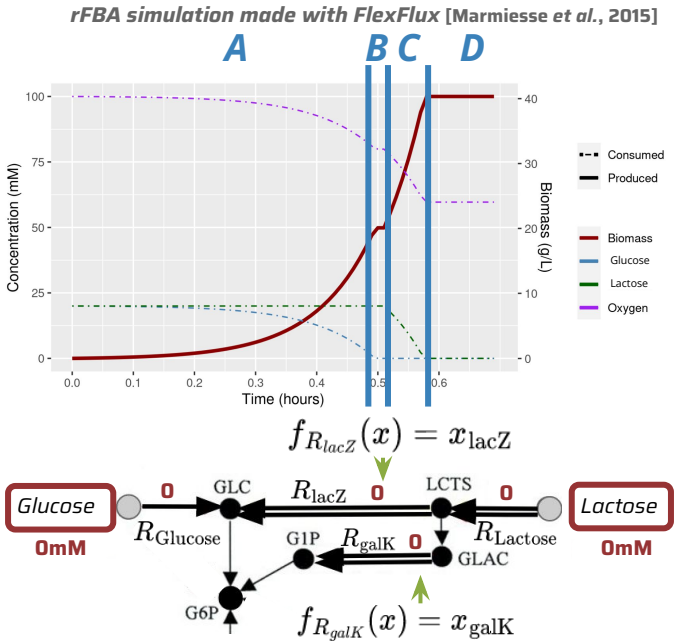
Phase C: **lacZ** ***and*** **galKTEU** ***states are updated allowing to import lactose***

[1] J. Monod, **Annales de l'Institut Pasteur**, 1942

Example of controlled induced behavior: diauxic shift[1]

Successives growth phases on different mediums
Controlled by the regulatory scale





Phase D: ***no carbon sources to allow growth***

[1] J. Monod, **Annales de l'Institut Pasteur**, 1942

MERRIN

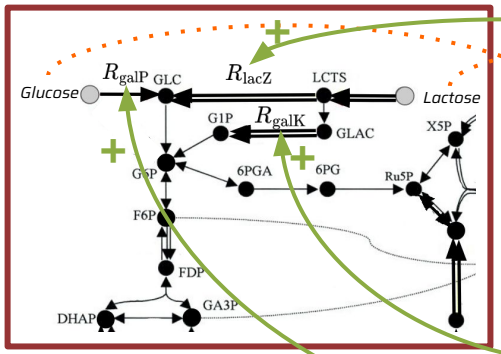
Infer Boolean regulatory rules from time series of kinetics, fluxomics, and transcriptomics observations

Thuillier, K., Baroukh, C., Bockmayr, A., Cottret, L., Paulevé, L., and Siegel, A. (2022). MERRIN: MEtabolic Regulation Rule INference from time series data. Bioinformatics.

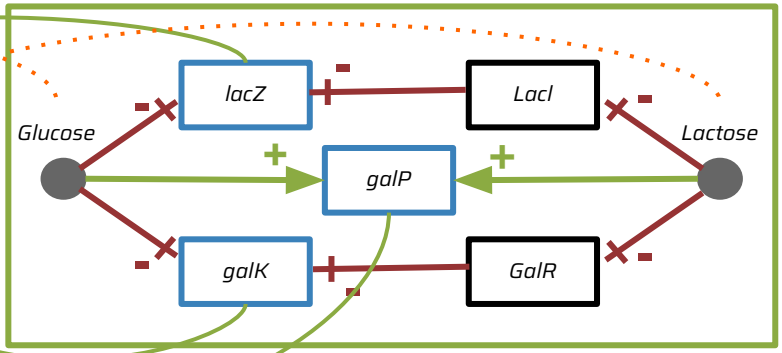
https://github.com/bioasp/merrin

MERRIN's inputs

Metabolic network



Interaction graph: define a search space

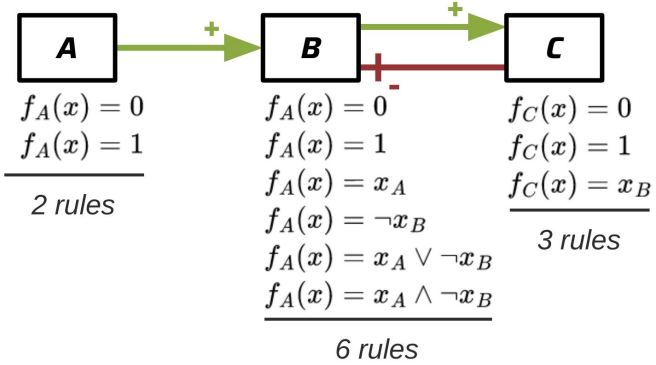


***Time series* observations**

Direct observations:
- transcriptomics

Indirect observations:
- kinetics
- fluxomics

SBML v3 file format
- with ***fbc*** data
- not required ***gene association rules***



Time series:
- **noise tolerant**
- Handle **mutant strains**

MERRIN's outputs

Sets of valid regulatory networks

Compatible with the interaction graph
Reproducing the input time series
*according to the **rFBA framework***

Enumeration modes:

***All** models or **subset minimal** models only*

Export as BNET *files:*

Easy to read for humans
Easy to convert into SBML-qual files

```
ArcA <- ![O2xt_b > 0]          substrates --> genes
SurplusFDP <- [FBP > 0]        reactions --> genes
Cra <- !SurplusFDP
galK <-(![GLCxt_b > 0] & !GalR)
LACUP <-(![GLCxt_b > 0] & ![LCTSxt_b > 0])
dcuC <- Fnr                    genes/substrates –> reactions
mdh <- !ArcA                   genes –> genes
ptsGHI <- ([ACxt_b > 0] | !Mlc)
```

Example of rules in BNET format

MERRIN *in practice*

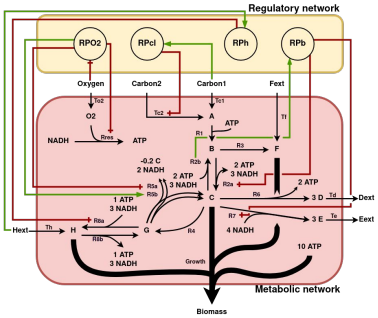
Core-carbon metabolism [Covert *et al.*, 2001]

Core-carbon model
20 reactions / 11 regulatory rules

From 5 time series
5-8 observations per time series

1 subset minimal model (~7s)
Infer 7 / 11 rules (in ~7s)
Explaining perfectly the data

Noise up to 20%



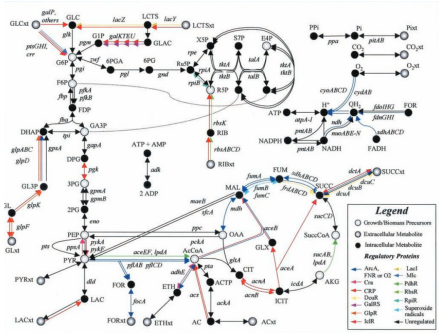
Transcriptomics + kinetics are enough
1 regulatory rule could never be inferred without kinetics

E. coli core-metabolism [Covert *et al.*, 2002]

Medium-scale model
113 reactions / 151 regulatory rules

From 3 time series

838 860 800 subset minimal models (< 8h)
Explaining perfectly the data
All subset minimal models



Rules may not be inferred due to data incompleteness
Precision: ~0.87 / Recall: ~0.11

What's next

Inferring Boolean regulatory rules from *Biolog data*

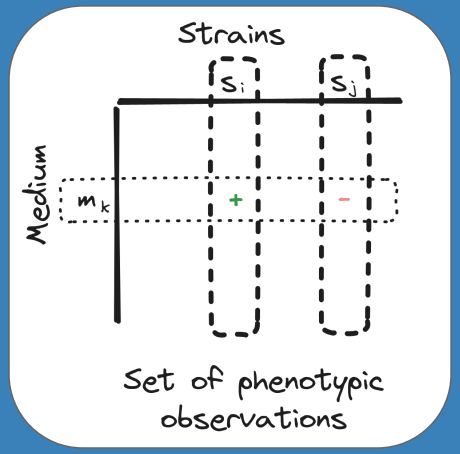
Set of phenotypic observations:
Does a mutant strain grow in a given medium?

Pair of medium / strains with growth state

Qualitative observations

Only initial medium concentrations are known

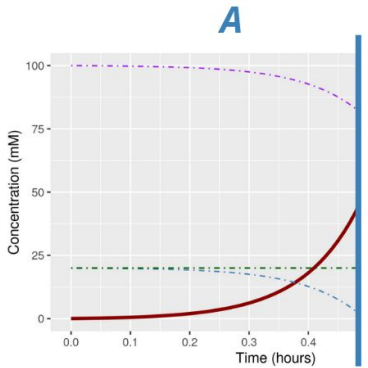
Easier to obtain than time series kinetics and transcriptomics

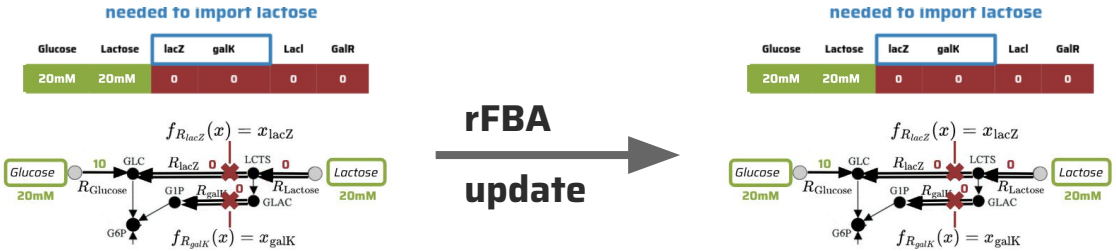


Regulated metabolic fixpoint

Heuristics: a biological system will converge toward a steady-state during growth phase

Growth condition:
exists 2 successives rFBA states are **identical** and **allow for growth**





Ensure growth until a substrate is added or depleted

Sufficient condition, but not necessary condition

Model Checking

Identify if a model is compatible with Biolog data

Model: **E.coli iJR904 + *i* MC1010**
1 076 reactions / 1 010 genes / 601 regulatory rules

Data: **13 640** Biolog observations
124 mediums / 110 strains

Model checking:

Synthetic data		Prediction	
		Growth	No Growth
Biolog	Growth	58.28%	2.81%
	No Growth	0.00%	38.91%

Experimental data		Prediction	
		Growth	No Growth
Biolog	Growth	65.23%	15.92%
	No Growth	7.77%	11.07%

***[Covert et al., 2004]* results are not reproducible from the paper model**

Identifying spurious rules

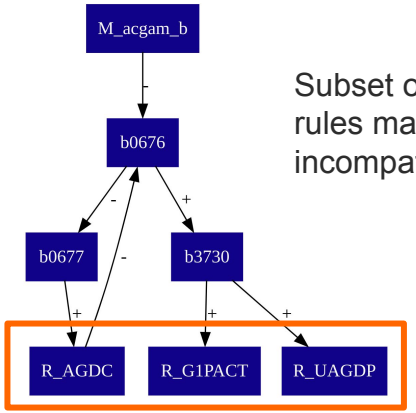
Identify candidate spurious rules for “predicted no-growth” but “observed growth” False negative

Example:

Medium: *M_acgam_b* | Strain: *b3172*

Covert: *Growth*

Our prediction: *No growth*



Subset of nodes whose rules may explain this incompatibility

No growth if these reactions are inhibited

Finding candidate for *False Positive* errors requires solving another class of theoretical problem

Conclusion

Merrin: inference from time series data

- **Metabolic network (SBML)**
- **Interaction graph**
- **Time series data**
  Kinetics + transcriptomics
  Mutant strain, < 20% noise

- **Regulatory networks (BNETs)**
- **rFBA compatible with input data**
- **Scale to *medium-scale model***

Inference from biolog data

Currently

- Compatibility between model and Biolog data
- Identify candidate rules leading to false negatives

Future works

- Identify candidate rules leading to false positive
  New class of theoretical problem to solve
- Divide and conquer methods to improve scalability

université PARIS-SACLAY
cnrs
école normale supérieure paris—saclay
LMF Laboratoire Méthodes Formelles