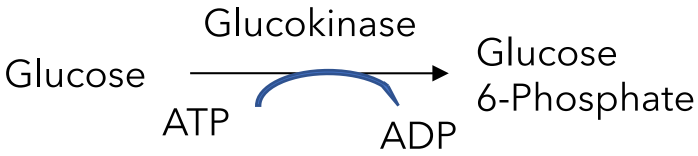
Learning boolean rules for the regulatory control of metabolism : a case study

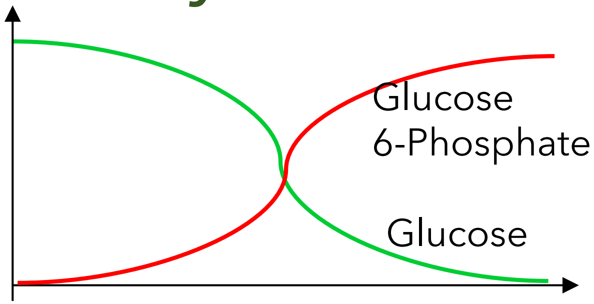
CNRS Rennes: Aurélien Cornet, Anne Siegel
Inra Toulouse: Caroline Barouck, Ludovic Cottret
CNRS Saclay: Loic Paulevé
Frei Berlin Univ: Alexander Bockmayr, Heike Siebert.

Modeling frameworks in systems biology

Metabolism



Glucose+ATP → Glucose 6-P+ADP

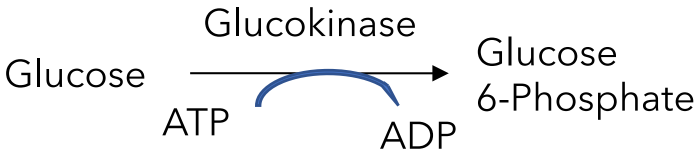


Differential equations

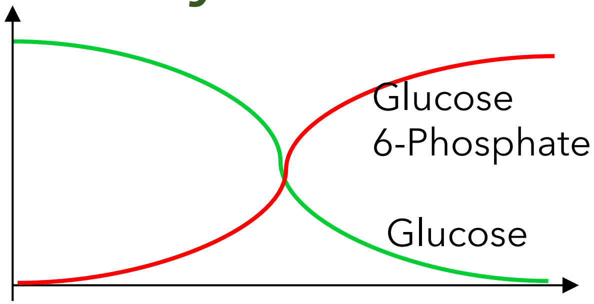
Entries are consumed by the reaction

Modeling frameworks in systems biology

Metabolism



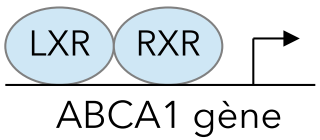
Glucose+ATP → Glucose 6-P+ADP



Differential equations

Entries are consumed by the reaction

Signaling and regulation



LXR (and) RXR → ABCA1

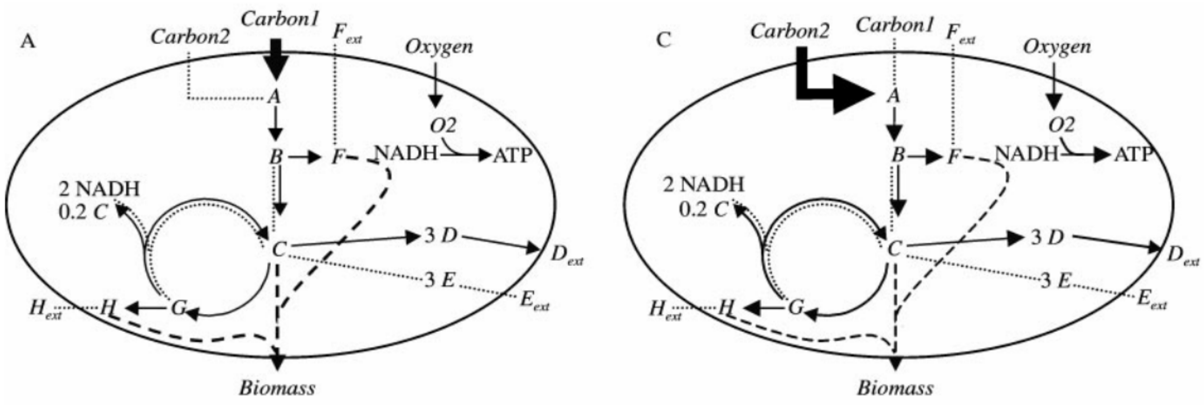
$$\begin{array}{c} \text{LXR} \\ \text{RXR} \\ \text{ABCA1} \end{array} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Logical rules

Entries are not modified by the interaction

General challenge: how can we couple metabolic and regulatory/signaling frameworks

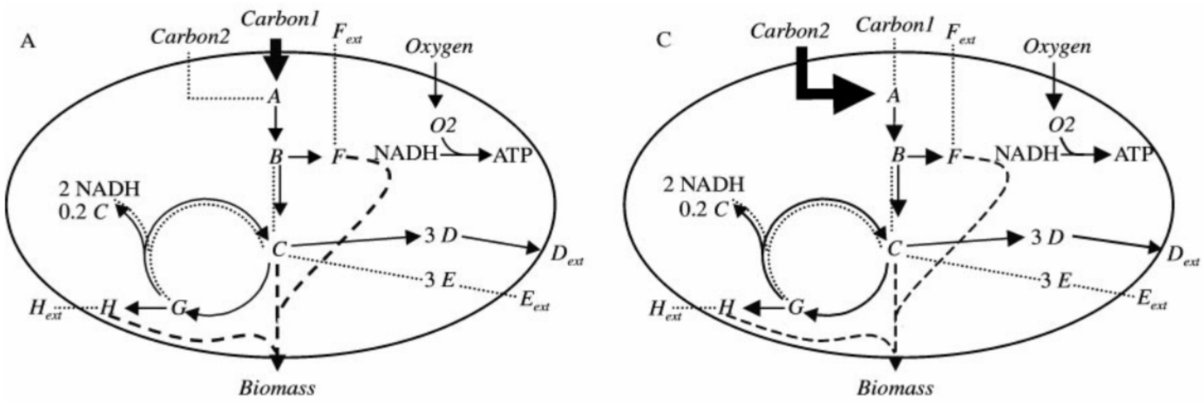
Carbon metabolism : diauxic shift

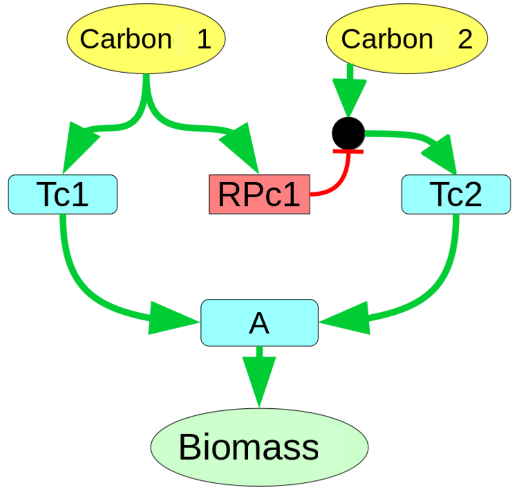


Mechanism (*Covert et al., 2001*)

- Growth based on 2 sources of carbon
- First, the carbon1 pathway (TC1 modulation) is activated
- When carbon 1 is over, the carbon2 pathway (TC2) is activated

Carbon metabolism : diauxic shift





Mechanism (*Covert et al., 2001*)

- Growth based on 2 sources of carbon
- First, the carbon1 pathway (TC1 modulation) is activated
- When carbon 1 is over, the carbon2 pathway (TC2) is activated
- **Regulatory mechanism: RPC1 regulates TC2**

Boolean rules

- RPc1 = Carbon1
- Tc2 = Carbon2 +!RPc1

The co-regulation of TC2 by carbon2, RCP1 and carbon1 is essential to the diauxic shift

Regulatory FBA: a way to simulate regulated metabolism

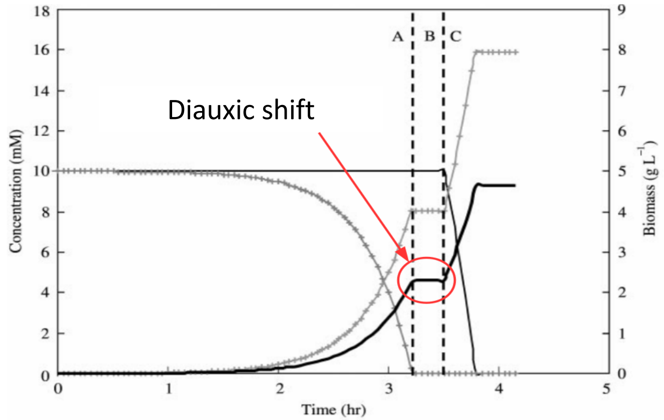
List of reactions

Reactions		List of regulations
Metabolic reactions		
$-1\ A - 1\ \text{ATP} + 1\ B$	R1	
$-1\ B + 2\ \text{ATP} + 2\ \text{NADH} + 1\ C$	R2a	IF NOT(RPb)
$-1\ C - 2\ \text{ATP} - 2\ \text{NADH} + 1\ B$	R2b	
$-1\ B + 1\ F$	R3	
$-1\ C + 1\ G$	R4	
$-1\ G + 0.8\ C + 2\ \text{NADH}$	R5a	IF NOT (RPO2)
$-1\ G + 0.8\ C + 2\ \text{NADH}$	R5b	IF RPO2
$-1\ C + 2\ \text{ATP} + 3\ D$	R6	
$-1\ C - 4\ \text{NADH} + 3\ E$	R7	IF NOT (RPb)
$-1\ G - 1\ \text{ATP} - 2\ \text{NADH} + 1\ H$	R8a	IF NOT (RPh)
$+1\ G + 1\ \text{ATP} + 2\ \text{NADH} - 1\ H$	R8b	
$-1\ \text{NADH} - 1\ O2 + 1\ \text{ATP}$	Rres	IF NOT (RPO2)
Transport processes		
$-1\ Carbon1 + 1\ A$	Tc1	
$-1\ Carbon2 + 1\ A$	Tc2	IF NOT(RPc1) AND IF(Carbon2)
$-1\ F_{ext} + 1\ F$	Tf	
$-1\ D + 1\ D_{ext}$	Td	
$-1\ E + 1\ E_{ext}$	Te	
$-1\ H_{ext} + 1\ H$	Th	
$-1\ Oxygen + 1\ O2$	To2	

List of regulations

→

FlexFlux,
(Cottret et al, 2015)



Covert et al., 2001

- **Entries** : reactions + regulations + logical rules

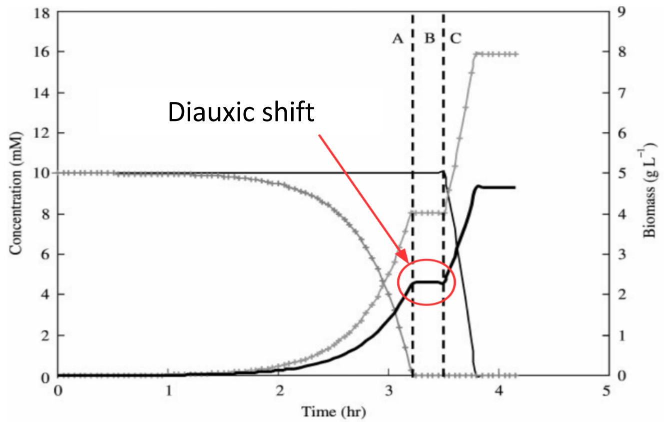
Regulatory FBA: a way to simulate regulated metabolism

List of reactions

Reactions		List of regulations
Metabolic reactions		
$-1\,A - 1\,\text{ATP} + 1\,B$	R1	
$-1\,B + 2\,\text{ATP} + 2\,\text{NADH} + 1\,C$	R2a	IF NOT(RPb)
$-1\,C - 2\,\text{ATP} - 2\,\text{NADH} + 1\,B$	R2b	
$-1\,B + 1\,F$	R3	
$-1\,C + 1\,G$	R4	
$-1\,G + 0.8\,C + 2\,\text{NADH}$	R5a	IF NOT (RPO2)
$-1\,G + 0.8\,C + 2\,\text{NADH}$	R5b	IF RPO2
$-1\,C + 2\,\text{ATP} + 3\,D$	R6	
$-1\,C - 4\,\text{NADH} + 3\,E$	R7	IF NOT (RPb)
$-1\,G - 1\,\text{ATP} - 2\,\text{NADH} + 1\,H$	R8a	IF NOT (RPh)
$+1\,G + 1\,\text{ATP} + 2\,\text{NADH} - 1\,H$	R8b	
$-1\,\text{NADH} - 1\,O2 + 1\,\text{ATP}$	Rres	IF NOT (RPO2)
Transport processes		
$-1\,Carbon1 + 1\,A$	Tc1	
$-1\,Carbon2 + 1\,A$	Tc2	IF NOT(RPc1) AND IF(Carbon2)
$-1\,F_{ext} + 1\,F$	Tf	
$-1\,D + 1\,D_{ext}$	Td	
$-1\,E + 1\,E_{ext}$	Te	
$-1\,H_{ext} + 1\,H$	Th	
$-1\,Oxygen + 1\,O2$	To2	

List of regulations

→ FlexFlux, (Cottret et al, 2015)

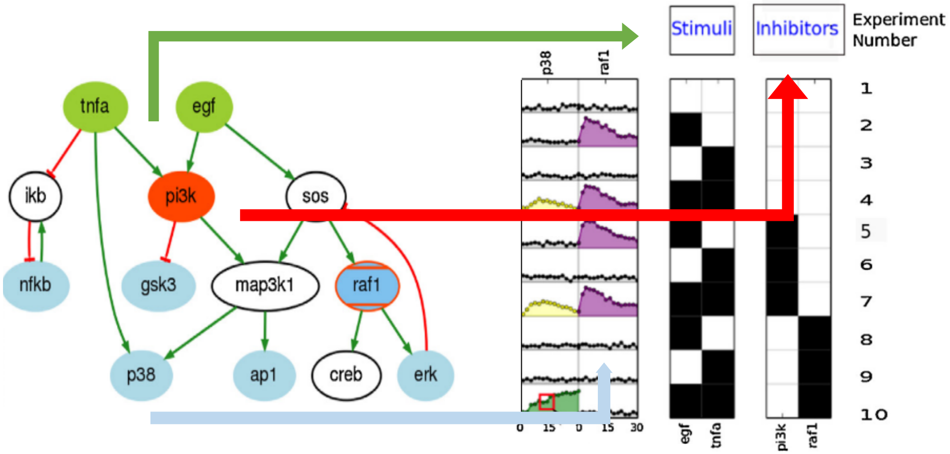


Covert et al., 2001

- **Entries** : reactions + regulations + logical rules
- **Metabolism** ← annotated genomes
- **Regulations** ← transcriptomic data
- **Logical rules** ← litterature, very hard to get

Main issue to perform dynamic FBA : a priori knowledge on logical rules

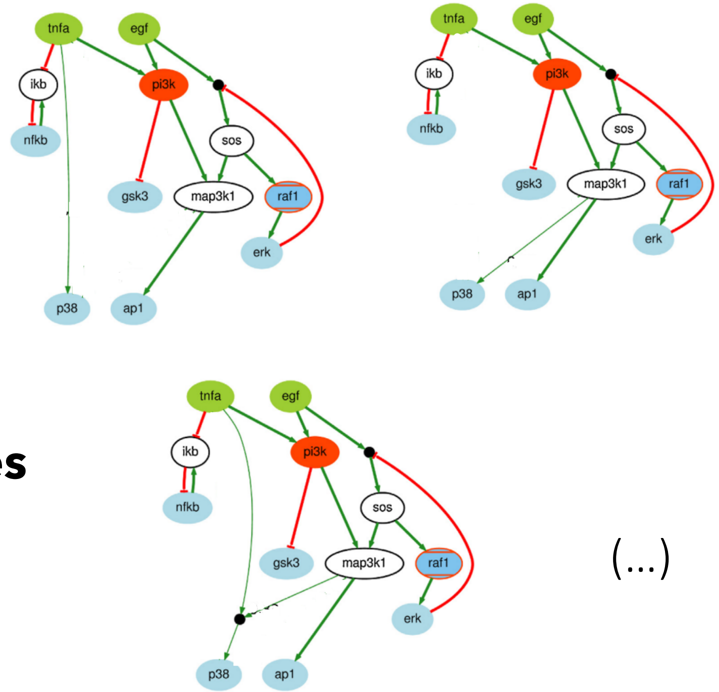
Learning logical rules: the caspo(ts) approach



Caspo time series

→

Ostrowski et al., 2016



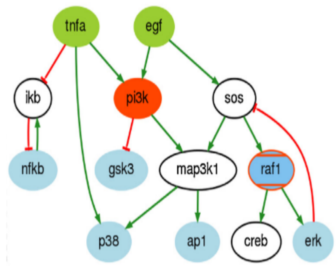
(...)

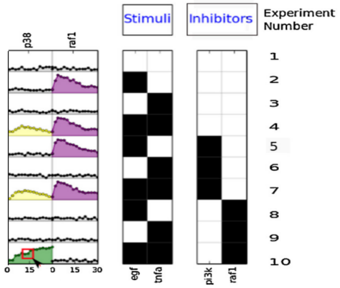
Several logical networks

- **PKN** : interaction graph derived from literature and data
  - stimuli, inhibitors, readouts
- **Data**: measures of graph entities in different experimental conditions

Caspots bottleneck: the value of activators is fixed all along each experimentation

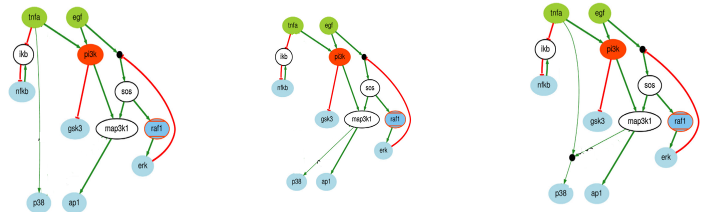
caspo(ts): underlying optimation problem





→

Ostrowski et al., 2016



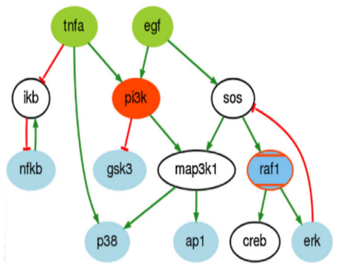
Modeling the network and experimentations

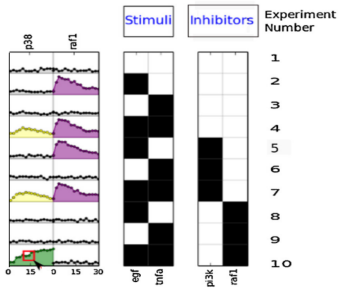
- V : set of vertices
- ϕ is the BN : maps any $v \in V$ to a propositional formula $\phi(v)$.
- Clamping ϕ_P according to an exp. P:
  fixing the values of logical formula $\phi(v)$ for stimuli and inhibitors.

Weighted multi-objective optimization

$$\underset{\phi : BNs}{\arg\min}(\underbrace{score_{rss}((V,\phi),(P_1,\ldots,P_n))}_{residual\ sum\ of\ squares}, \underbrace{score_{size}((V,\phi))}_{complexity})$$

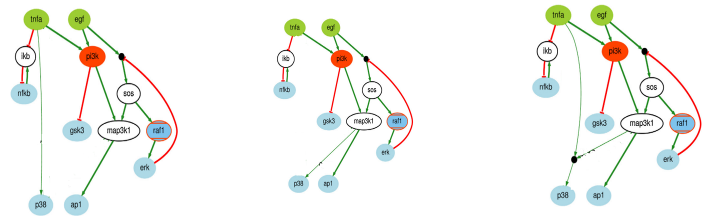
caspo(ts): underlying optimation problem





→

Ostrowski et al., 2016



Weighted multi-objective optimization

$$\underset{\phi\,:\,BNs}{\arg\min}\big(\underbrace{\text{score}_{rss}((V,\phi),(P_1,\ldots,P_n))}_{residual\ sum\ of\ squares},\underbrace{\text{score}_{size}((V,\phi))}_{complexity}\big)$$

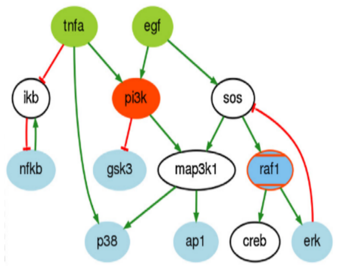
Modeling the network and experimentations

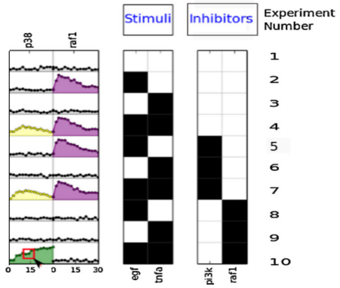
- V : set of vertices
- ϕ is the BN : maps any $v \in V$ to a propositional formula $\phi(v)$.
- Clamping ϕ_P according to an exp. P:
  fixing the values of logical formula $\phi(v)$ for stimuli and inhibitors.

2nd objective ($\in \mathbb{N}$). model complexity (length of logic formulas):

$$\text{score}_{size}((V,\phi)) = \sum_{v:variable} |\phi(v)|$$

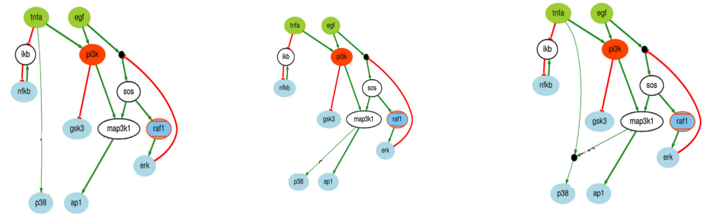
caspo(ts): underlying optimation problem





$\rightarrow$

Ostrowski et al., 2016



Modeling the network and experimentations

- V : set of vertices
- ϕ is the BN : maps any $v \in V$ to a propositional formula $\phi(v)$.
- Clamping ϕ_P according to an exp. P:
  fixing the values of logical formula $\phi(v)$ for stimuli and inhibitors.

Weighted multi-objective optimization

$$\arg\min_{\phi\,:\,BNs}(\underbrace{\text{score}_{rss}((V,\phi),(P_1,\ldots,P_n))}_{\textit{residual sum of squares}}, \underbrace{\text{score}_{size}((V,\phi))}_{\textit{complexity}})$$

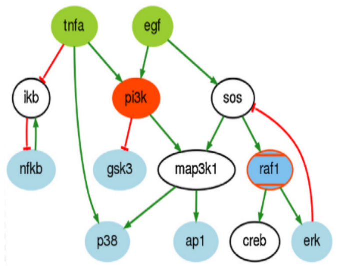
2nd objective ($\in \mathbb{N}$). model complexity (length of logic formulas):

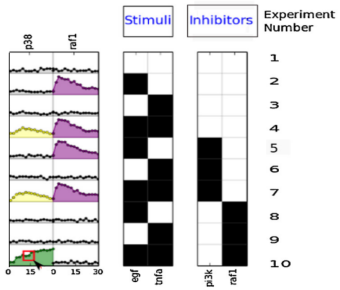
$$\text{score}_{size}((V,\phi)) = \sum_{v:variable} |\phi(v)|$$

1st objective ($\in \mathbb{R}$). difference between observations and predictions:

$$\text{score}_{rss}((V,\phi),\underbrace{(P_1,\ldots,P_n)}_{n\ experiments}) = \sum_{\substack{t\ :\ timepoints\\ P\ :\ experimentations\\ v\ :\ variable}} \left(\underbrace{\text{measure}(v,P,t)}_{observations\in[0,1]} - \underbrace{\phi_P(v,t)}_{predictions\in\{0,1\}}\right)^2$$

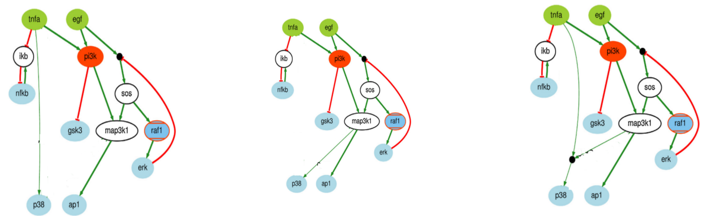
caspo(ts): underlying optimation problem





→

Ostrowski et al., 2016



Weighted multi-objective optimization

$$\underset{\phi : BNs}{\arg\min}(\underbrace{\text{score}_{rss}((V,\phi),(P_1,\dots,P_n))}_{residual\ \ sum\ \ of\ \ squares}, \underbrace{\text{score}_{size}((V,\phi))}_{complexity})$$

All optimal BNs equivalently explain the data

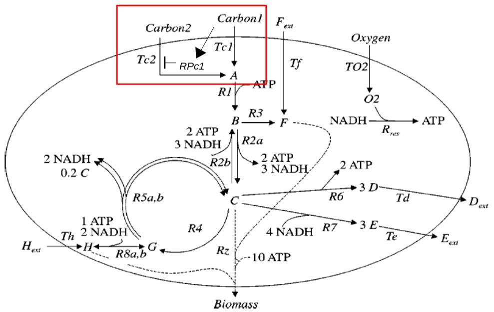
Introduce a threashold in data at 0.5

1st objective **($\in \mathbb{R}$).** difference between observations and predictions:

$$\text{score}_{rss}((V,\phi),\underbrace{(P_1,\dots,P_n)}_{n\ experiments}) = \sum_{\substack{t\ :\ timepoints\\ P\ :\ experimentations\\ v\ :\ variable}} \left(\underbrace{\text{measure}(v,P,t)}_{observations\in[0,1]} - \underbrace{\phi_P(v,t)}_{predictions\in\{0,1\}}\right)^2$$

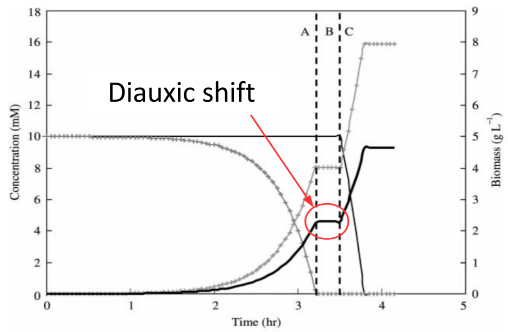
Stimuli and inhibitors have fixed values

Issue: how to recover the logical gates of the diauxic shift ?



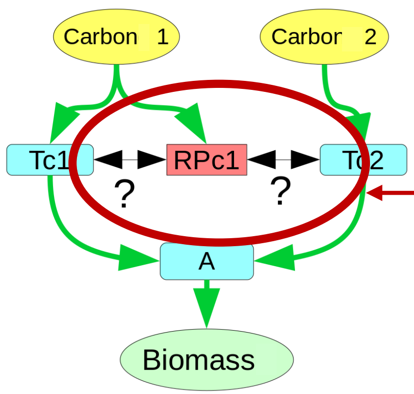
→

simulated data

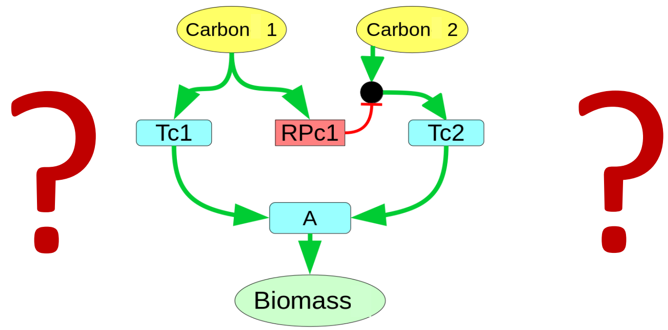


TR:Cell:CellLine	DA:RPcl	DV:RPcl	DA:Tc2	DV:Tc2	DA:Tc1	DV:Tc1	DA:Carbon1	DV:Carbon1	DA:Carbon2	DV:Carbon2	DA:Biomass	DV:Biomass
1	0	0	0	0	0	0	0	10	0	10	0	0
1	1	0	1	10.5	1	10.5	1	10	1	10	1	0
1	2	1	2	10.5	2	10.5	2	9.99	2	9.99	2	0.01
1	3	1	3	0	3	10.5	3	9.96	3	9.96	3	0.03
1	4	1	4	0	4	10.5	4	9.91	4	9.96	4	0.04
1	5	1	5	0	5	10.5	5	9.84	5	9.96	5	0.06
1	6	1	6	0	6	10.5	6	9.75	6	9.96	6	0.09
1	7	1	7	0	7	10.5	7	9.62	7	9.96	7	0.14
1	8	1	8	0	8	10.5	8	9.44	8	9.96	8	0.19
1	9	1	9	0	9	10.5	9	9.18	9	9.96	9	0.28
1	10	1	10	0	10	10.5	10	8.83	10	9.96	10	0.39
1	11	1	11	0	11	10.5	11	8.33	11	9.96	11	0.55
1	12	1	12	0	12	10.5	12	7.63	12	9.96	12	0.77
1	13	1	13	0	13	10.5	13	6.66	13	9.96	13	1.08
1	14	1	14	0	14	10.5	14	5.3	14	9.96	14	1.51
1	15	1	15	0	15	10.5	15	3.4	15	9.96	15	2.12
1	16	1	16	0	16	2.49	16	0.74	16	9.96	16	2.97
1	17	1	17	0	17	0	17	0	17	9.96	17	3.22
1	18	0	18	0	18	0	18	0	18	9.96	18	3.22
1	19	0	19	10.5	19	0	19	0	19	9.96	19	3.22
1	20	0	20	10.5	20	0	20	0	20	5.93	20	4.5
1	21	0	21	0.49	21	0	21	0	21	0.31	21	6.29
1	22	0	22	0	22	0	22	0	22	0	22	6.39

↓ Prior knowledge network
On a very small toy-example

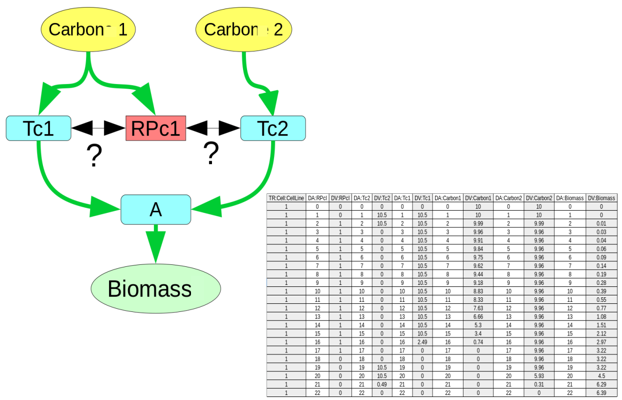


Search among all possible boolean interactions between RPc1, TC1 and TC2



TC2 = (Carbon2) AND (NOT RPc1)

Basic approach based on caspots

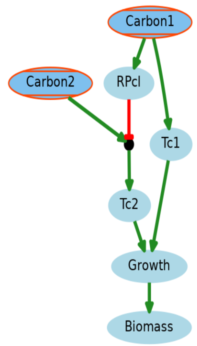


Feed caspots with simulation data

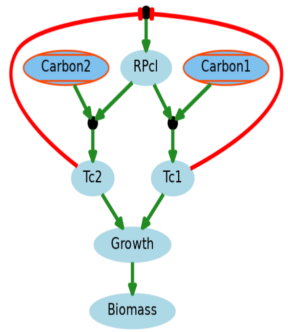
40 models …

→

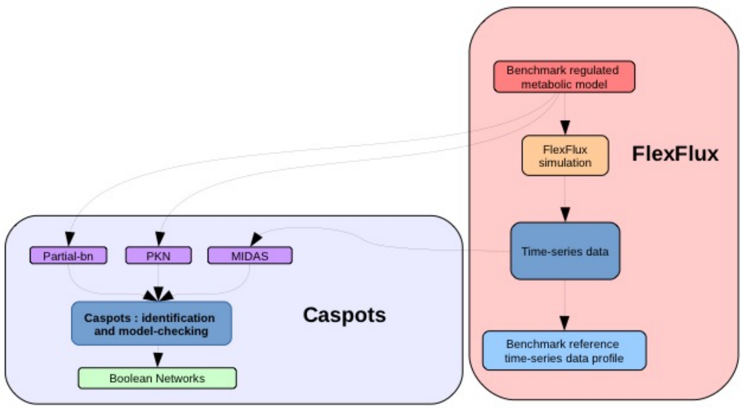
(….)



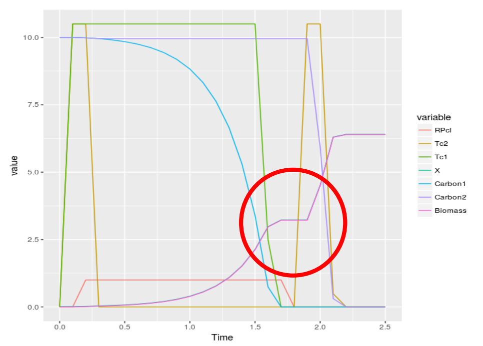
(….)

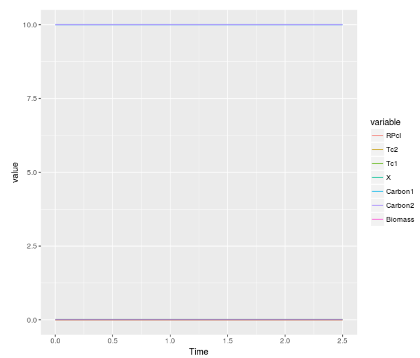


(….)



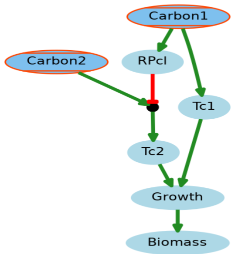
↓ Simulations …



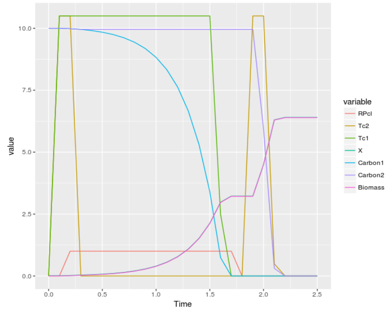


Only one model among the 40 solutions show a diauxic shift in the simulation

Flux-based post-validation is required !



A model
among others



Simulation

TR:Cell:CellLine	DA:RPcl	DV:RPcl	DA:Tc2	DV:Tc2	DA:Tc1	DV:Tc1	DA:Carbon1	DV:Carbon1	DA:Carbon2	DV:Carbon2	DA:Biomass	DV:Biomass
1	0	=	0	+	0	+	0	=	0	=	0	=
1	1	+	1	=	1	=	1	-	1	-	1	+
1	2	=	2	-	2	=	2	-	2	-	2	+
1	3	=	3	=	3	=	3	-	3	=	3	+
1	4	=	4	=	4	-	4	-	4	=	4	+
1	16	-	16	=	16	-	16	-	16	=	16	+
1	18	=	18	+	18	=	18	=	18	=	18	=
1	19	=	19	=	19	=	19	=	19	-	19	+
1	20	=	20	-	20	=	20	=	20	-	20	+

Extract the profile of the output variables
(changes of variations)

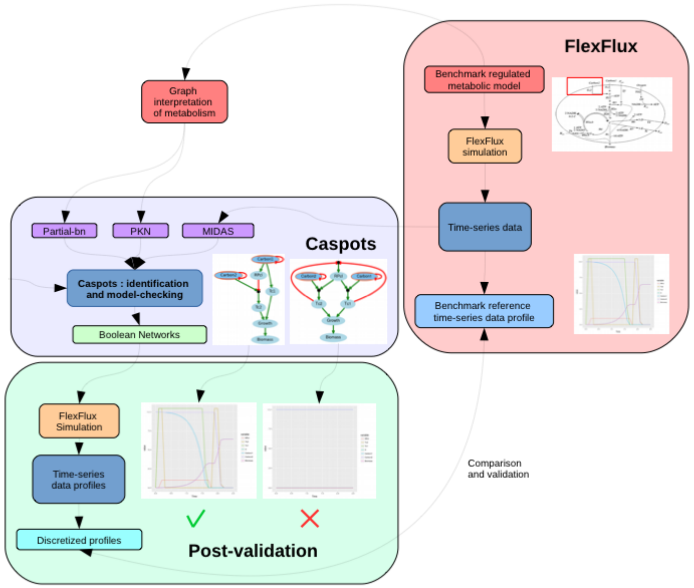
TR:Cell:CellLine	DA:RPcl	DV:RPcl	DA:Tc2	DV:Tc2	DA:Tc1	DV:Tc1	DA:Carbon1	DV:Carbon1	DA:Carbon2	DV:Carbon2	DA:Biomass	DV:Biomass
1	0	=	0	+	0	+	0	=	0	=	0	=
1	1	+	1	=	1	=	1	-	1	-	1	+
1	2	=	2	-	2	=	2	-	2	-	2	+
1	3	=	3	=	3	=	3	-	3	=	3	+
1	4	=	4	=	4	-	4	-	4	=	4	+
1	16	-	16	=	16	-	16	-	16	=	16	+
1	18	=	18	+	18	=	18	=	18	=	18	=
1	19	=	19	=	19	=	19	=	19	-	19	+
1	20	=	20	-	20	=	20	=	20	-	20	+

Compare the profiles of simulation with the profile of data

Simulation score =number of discrepancies between
the profile of the simulated network and the profile of experimental data

Revisiting the optimisation problem

Adding a post-validation to the pipeline



New optimization problem

Weighted multi-objective optimization

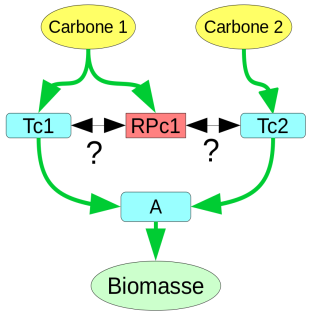
$$\arg\min_{\phi : BNs}(\underbrace{\text{score}_{rss}((V,\phi),(P_1,\ldots,P_n))}_{residual\ sum\ of\ squares}, \underbrace{\text{score}_{size}((V,\phi))}_{complexity})$$

$$\downarrow$$

$$\arg\min_{\phi : BNs}(\underbrace{\text{score}_{rss}((V,\phi),(P_1,\ldots,P_n))}_{residual\ sum\ of\ squares}, \underbrace{\text{score}_{size}((V,\phi))}_{network\ size}, \underbrace{\text{score}_{simu}(met\_network,\phi))}_{flux-based\ simulation}$$

The selection of the best-model involves a flux-based simulation which cannot be addressed with logical networks.

Reducing the number of false-positive ?

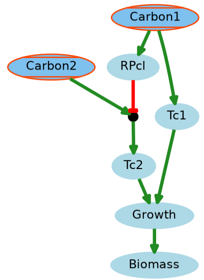


TR:Cell:CellLine	DA:RPcl	DV:RPcl	DA:Tc2	DV:Tc2	DA:Tc1	DV:Tc1	DA:Carbon1	DV:Carbon1	DA:Carbon2	DV:Carbon2	DA:Biomass	DV:Biomass
1	0	0	0	0	0	0	0	10	0	10	0	0
1	1	0	1	10.5	1	10.5	1	10	1	10	1	0
1	2	1	2	10.5	2	10.5	2	9.99	2	9.99	2	0.01
1	3	1	3	0	3	10.5	3	9.96	3	9.96	3	0.03
1	4	1	4	0	4	10.5	4	9.91	4	9.96	4	0.04
1	5	1	5	0	5	10.5	5	9.84	5	9.96	5	0.06
1	6	1	6	0	6	10.5	6	9.75	6	9.96	6	0.09
1	7	1	7	0	7	10.5	7	9.62	7	9.96	7	0.14
1	8	1	8	0	8	10.5	8	9.44	8	9.96	8	0.19
1	9	1	9	0	9	10.5	9	9.18	9	9.96	9	0.28
1	10	1	10	0	10	10.5	10	8.83	10	9.96	10	0.39
1	11	1	11	0	11	10.5	11	8.33	11	9.96	11	0.55
1	12	1	12	0	12	10.5	12	7.63	12	9.96	12	0.77
1	13	1	13	0	13	10.5	13	6.66	13	9.96	13	1.08
1	14	1	14	0	14	10.5	14	5.3	14	9.96	14	1.51
1	15	1	15	0	15	10.5	15	3.4	15	9.96	15	2.12
1	16	1	16	0	16	2.49	16	0.74	16	9.96	16	2.97
1	17	1	17	0	17	0	17	0	17	9.96	17	3.22
1	18	0	18	0	18	0	18	0	18	9.96	18	3.22
1	19	0	19	10.5	19	0	19	0	19	9.96	19	3.22
1	20	0	20	10.5	20	0	20	0	20	5.93	20	4.5
1	21	0	21	0.49	21	0	21	0	21	0.31	21	6.29
1	22	0	22	0	22	0	22	0	22	0	22	6.39

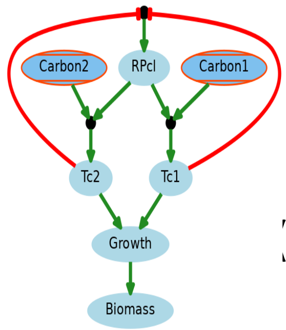
40 models …

→

(…)



(…)

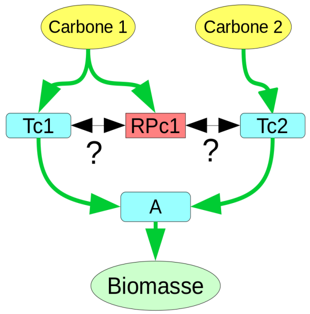


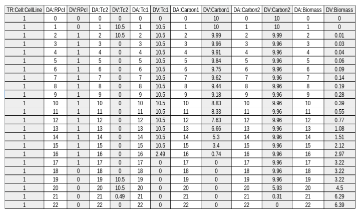
(…)

?

Most of the noise is generated by the fact that
the carbon1 and carbon2 decrease is not explained

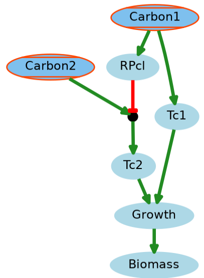
Reducing the number of false-positive ?



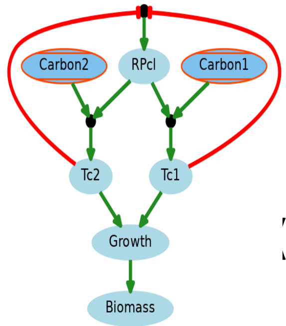


40 models …

→ (…)



(…)

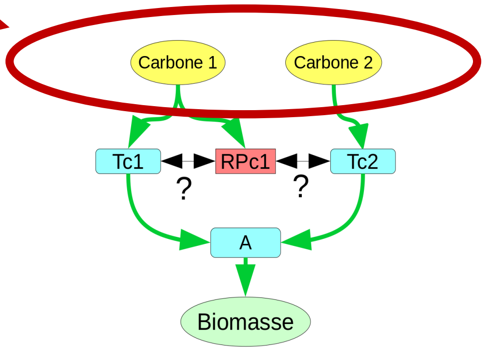


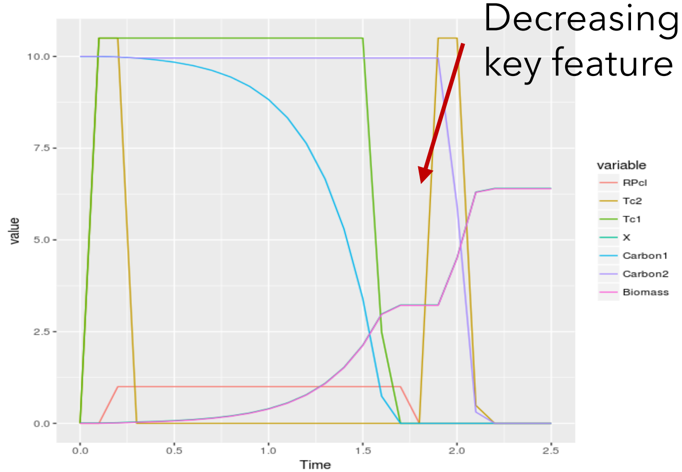
(…)

?

Most of the noise is generated by the fact that
the carbon1 and carbon2 decrease is not explained

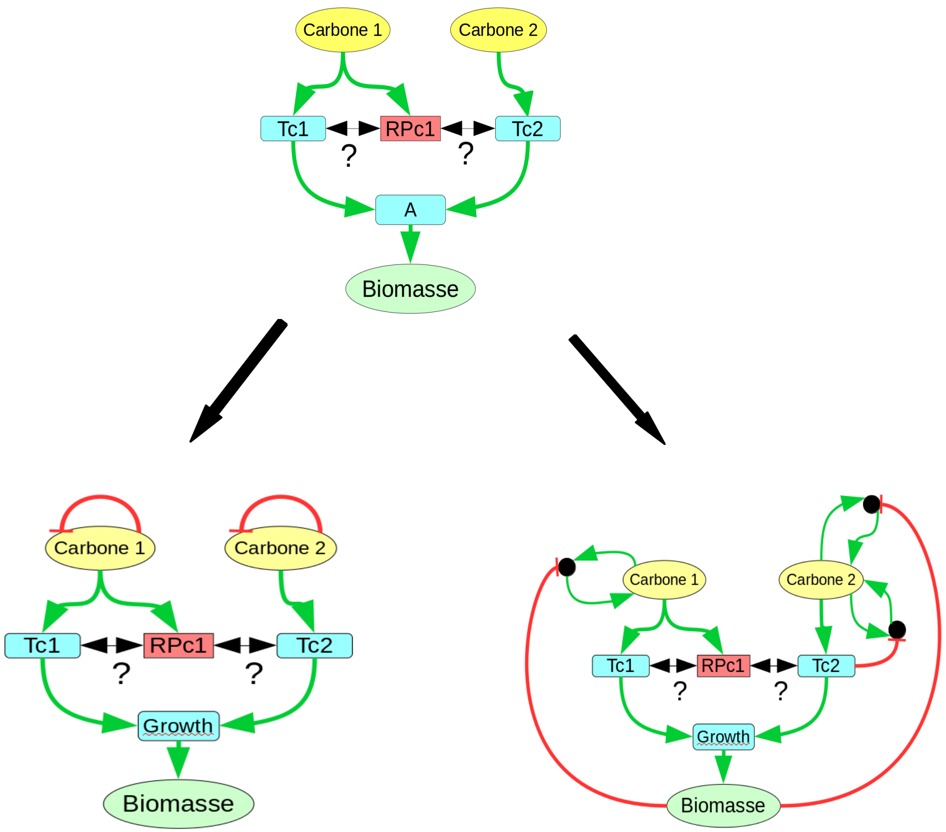
Fixed by the clamping
operation in caspots





Decreasing carbons in a
key feature of the process

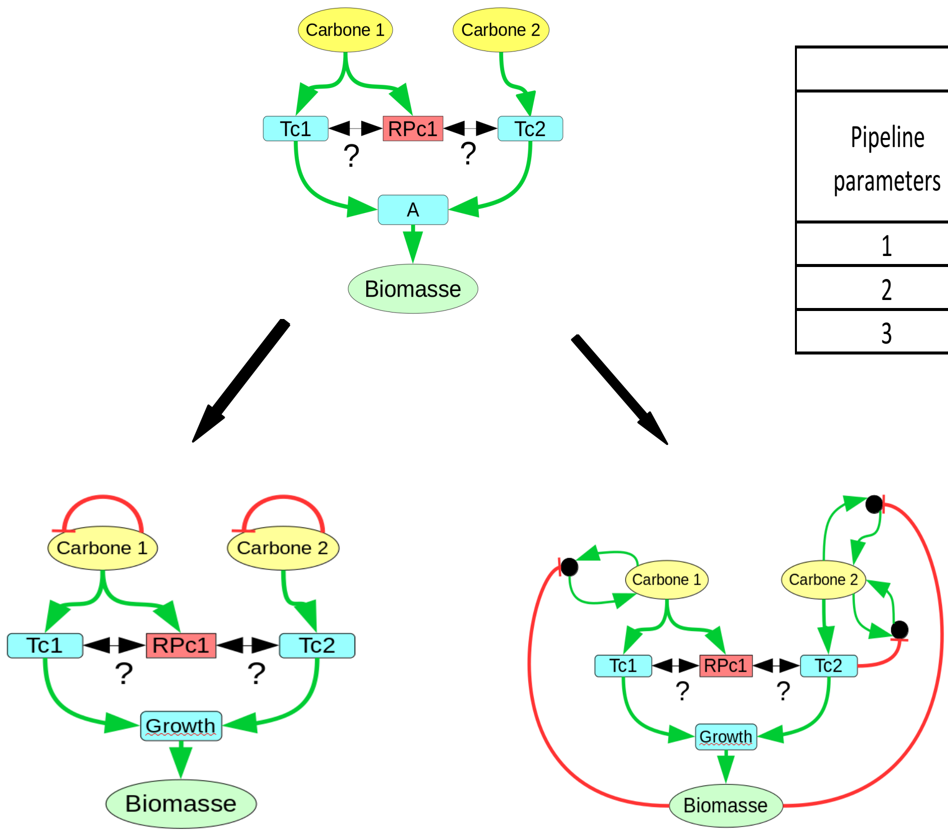
A better modeling for carbon inputs



Feed-forward loop

Artificial logical dependency

A better modeling for carbon inputs



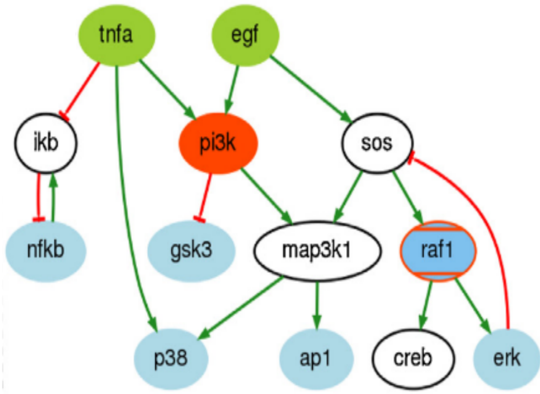
Feed-forward loop

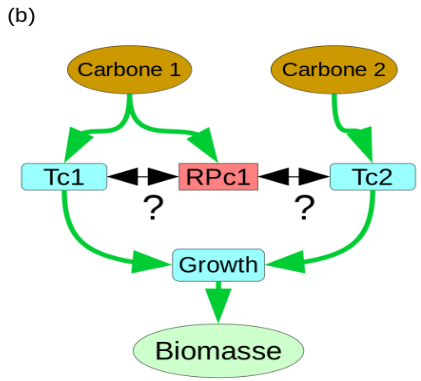
Artificial logical dependency

	PKN modelling			Number of output networks		Running time
Pipeline parameters	Integration Graph	Interaction graph with retro-controls	Interaction graph with artificial links	Learning	Comparison to benchmark dataset	
1	x			40	1	seconds
2		x		15	1	minutes
3			x	4	1	hours

Introducing artificial dependencies highly reduces the number of false-positive

Adding a fourth class of nodes to caspots

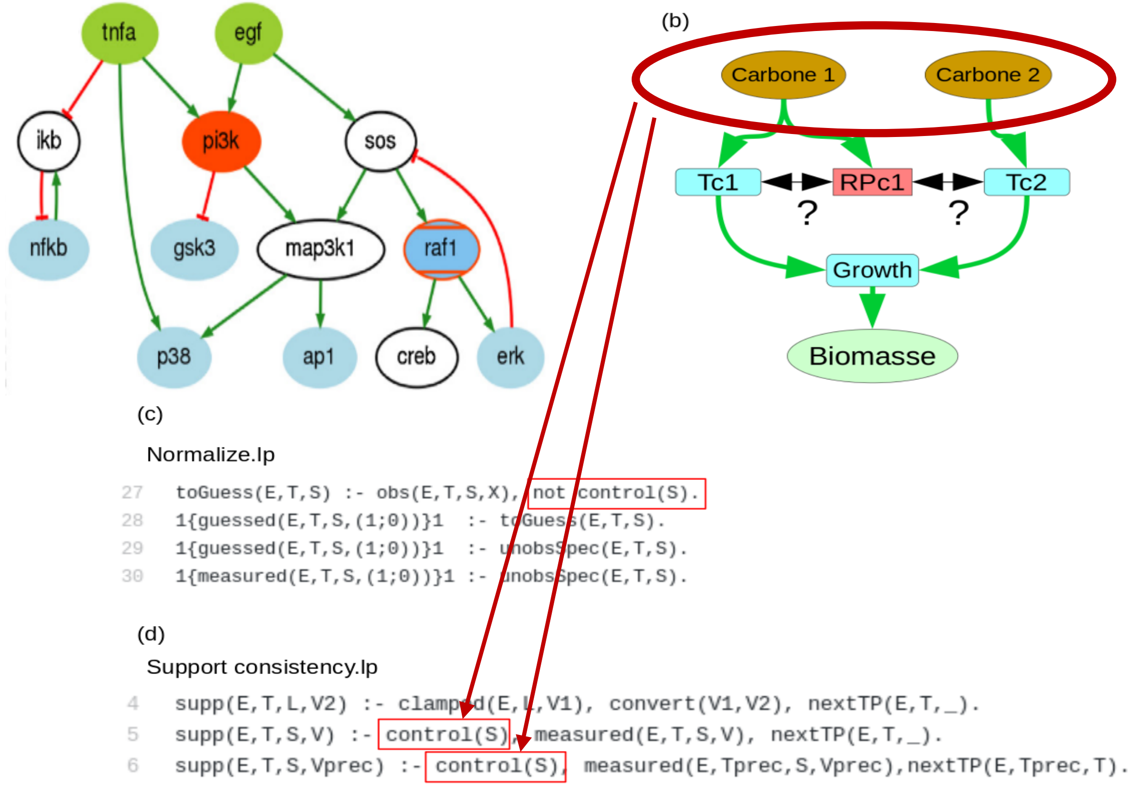




- Nodes can be
  - **Stimuli** : activated during the experimentation
  - **Inhibitors:** zero value during the experimentation
  - **Readout:** measured and controled by the boolean network dynamics
  - **Control:** measured and not controled by the BN dynamics.
- **Implementation** : « --control-nodes »

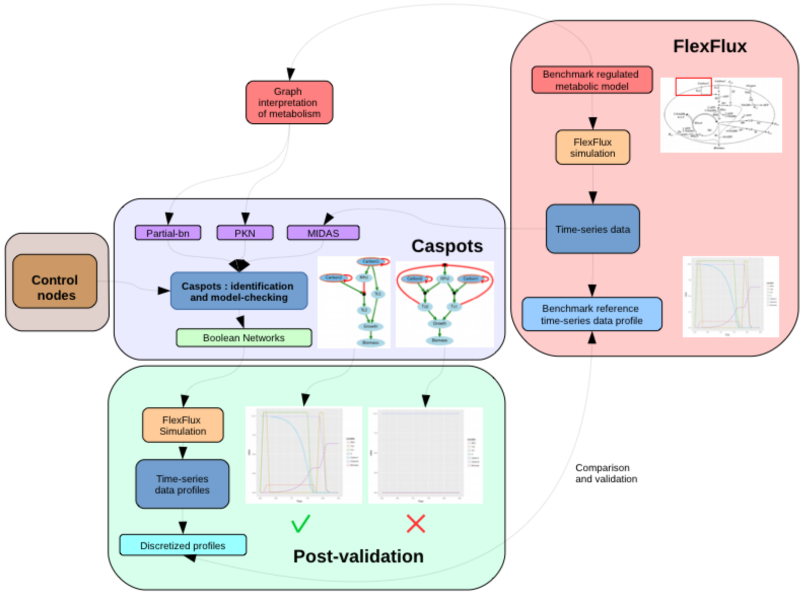
Adding a fourth class of nodes to caspots

- Nodes can be
  - **Stimuli** : activated during the experimentation
  - **Inhibitors:** zero value during the experimentation
  - **Readout:** measured and controled by the boolean network dynamics
  - **Control:** measured and not controled by the BN dynamics.
- **Implementation** : « --control-nodes »



Smart changes in the ASP encoding of caspots

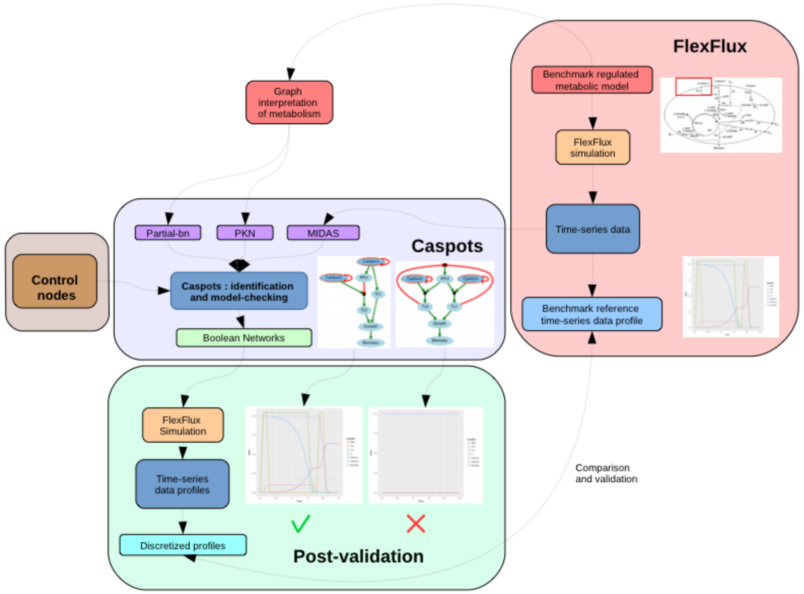
Revisiting the optimisation problem



	PKN modelling			Caspots version		Number of output networks		
Pipeline parameters	Integration Graph	Interaction graph with retro-controls	Interaction graph with artificial links	Without retro-control	With retro-controls	Learning	Comparison to benchmark dataset	Running time
1	x			x		40	1	seconds
2		x		x		15	1	minutes
3			x	x		4	1	hours
7	x				x	4	1	seconds
8		x			x	4	1	minutes
9			x		x	4	1	hours

- No more need to model any artificial relation
- Gain of performance

Revisiting the optimisation problem



Optimisation problem

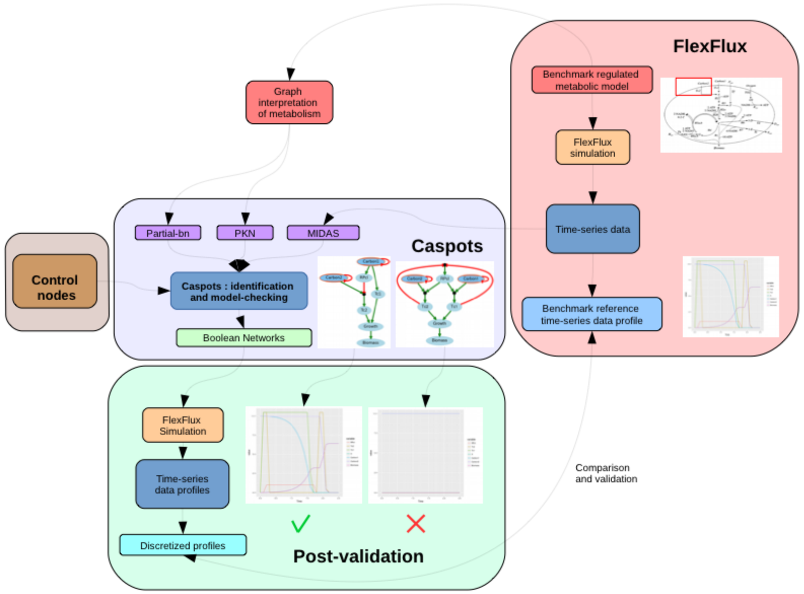
$$\arg\min_{\phi : BNs}\Big(\underbrace{score_{rss}((V,\phi),(P_1,\dots,P_n))}_{residual\ sum\ of\ squares},\ \underbrace{score_{size}((V,\phi))}_{network\ size},\ \underbrace{score_{simu}(met\_network,\phi)}_{flux-based\ simulation}\Big)$$

$$score_{rss}((V,\phi),\underbrace{(P_1,\dots,P_n)}_{n\ experiments}) = \sum_{\substack{t\ :\ timepoints\\ P\ :\ experimentations\\ v\ :\ variable}} \Big(\underbrace{measure(v,P,t)}_{observations\in[0,1]} - \underbrace{\phi_P(v,t)}_{predictions\in\{0,1\}}\Big)^2$$

	PKN modelling			Caspots version		Number of output networks		
Pipeline parameters	Integration Graph	Interaction graph with retro-controls	Interaction graph with artificial links	Without retro-control	With retro-controls	Learning	Comparison to benchmark dataset	Running time
1	x			x		40	1	seconds
2		x		x		15	1	minutes
3			x	x		4	1	hours
7	x				x	4	1	seconds
8		x			x	4	1	minutes
9			x		x	4	1	hours

- No more need to model any artificial relation
- Gain of performance

Revisiting the optimisation problem



Optimisation problem

$$\underset{\phi\,:\,BNs}{\arg\min}\Big(\underbrace{score_{rss}((V,\phi),(P_1,\ldots,P_n))}_{residual\ sum\ of\ squares},\ \underbrace{score_{size}((V,\phi))}_{network\ size},\ \underbrace{score_{simu}(met\_network,\phi)}_{flux-based\ simulation}\Big)$$

$$score_{rss}((V,\phi),\underbrace{(P_1,\ldots,P_n)}_{n\ experiments}) = \sum_{\substack{t\,:\,timepoints\\ P\,:\,experimentations\\ v\,:\,variable}} \Big(\underbrace{measure(v,P,t)}_{observations\in[0,1]} - \underbrace{\phi_P(v,t)}_{predictions\in\{0,1\}}\Big)^2$$

Remove control nodes from the score computation

	PKN modelling			Caspots version		Number of output networks		Running time
Pipeline parameters	Integration Graph	Interaction graph with retro-controls	Interaction graph with artificial links	Without retro-control	With retro-controls	Learning	Comparison to benchmark dataset	
1	x			x		40	1	seconds
2		x		x		15	1	minutes
3			x	x		4	1	hours
7	x				x	4	1	seconds
8		x			x	4	1	minutes
9			x		x	4	1	hours

$$scoreControl_{rss}((V,\phi),\underbrace{(P_1,\ldots,P_n)}_{n\ experiments}) = \sum_{\substack{t\,:\,timepoints\\ P\,:\,experimentations\\ v\,:\,variable\\ v\,\neq\,control}} \Big(\underbrace{measure(v,P,t)}_{observations\in[0,1]} - \underbrace{\phi_P(v,t)}_{predictions\in\{0,1\}}\Big)^2$$

- No more need to model any artificial relation
- Gain of performance

Focus on the last false positives

	PKN modelling			Caspots version		Number of output networks		
Pipeline parameters	Integration Graph	Interaction graph with retro-controls	Interaction graph with artificial links	Without retro-control	With retro-controls	Learning	Comparison to benchmark dataset	Running time
1	x			x		40	1	seconds
2		x		x		15	1	minutes
3			x	x		4	1	hours
7	x				x	4	1	seconds
8		x			x	4	1	minutes
9			x		x	4	1	hours

Focus on the last false positives

?

	PKN modelling			Caspots version		Number of output networks		
Pipeline parameters	Integration Graph	Interaction graph with retro-controls	Interaction graph with artificial links	Without retro-control	With retro-controls	Learning	Comparison to benchmark dataset	Running time
1	x			x		40	1	seconds
2		x		x		15	1	minutes
3			x	x		4	1	hours
7	x				x	4	1	seconds
8		x			x	4	1	minutes
9			x		x	4	1	hours

?

Same trend
→ accounted 6 times in the score computation

Time	Biomass	Carbon1	Carbon2
0.0	0.0	10.0	10.0
0.1	0.0	10.0	10.0
0.2	0.009547	9.985044	9.985044
0.3	0.028207	9.95581	9.95581
0.4	0.064681	9.898667	9.898667
0.5	0.135975	9.786972	9.786972
0.6	0.275331	9.568648	9.568648
0.7	0.547723	9.141901	9.141901
0.8	1.080155	8.307758	8.307758
0.9	2.120875	6.677297	6.677297
1.0	4.155121	3.490311	3.490311
1.1	7.100907	0.0	0.0
1.2	7.100907	0.0	0.0
1.3	7.100907	0.0	0.0
1.4	7.100907	0.0	0.0
1.5	7.100907	0.0	0.0
1.6	7.100907	0.0	0.0

The MSE score favors some non-accurante boolean values

- Biomass = 4,15/max → 1
- **Carbon 1 = 3,49/max → 0**

Several biais in the metabolic data

Data pre-treatment

TR:Cell:CellLine	DA:RPcl	DV:RPcl	DA:Tc2	DV:Tc2	DA:Tc1	DV:Tc1	DA:Carbon1	DV:Carbon1	DA:Carbon2	DV:Carbon2	DA:Biomass	DV:Biomass
1	0	0	0	0	0	0	0	10	0	10	0	0
1	1	0	1	10.5	1	10.5	1	10	1	10	1	0
1	2	1	2	10.5	2	10.5	2	9.99	2	9.99	2	0.01
1	3	1	3	0	3	10.5	3	9.96	3	9.96	3	0.03
1	4	1	4	0	4	10.5	4	9.91	4	9.96	4	0.04
1	5	1	5	0	5	10.5	5	9.84	5	9.96	5	0.06
1	6	1	6	0	6	10.5	6	9.75	6	9.96	6	0.09
1	7	1	7	0	7	10.5	7	9.62	7	9.96	7	0.14
1	8	1	8	0	8	10.5	8	9.44	8	9.96	8	0.19
1	9	1	9	0	9	10.5	9	9.18	9	9.96	9	0.28
1	10	1	10	0	10	10.5	10	8.83	10	9.96	10	0.39
1	11	1	11	0	11	10.5	11	8.33	11	9.96	11	0.55
1	12	1	12	0	12	10.5	12	7.63	12	9.96	12	0.77
1	13	1	13	0	13	10.5	13	6.66	13	9.96	13	1.08
1	14	1	14	0	14	10.5	14	5.3	14	9.96	14	1.51
1	15	1	15	0	15	10.5	15	3.4	15	9.96	15	2.12
1	16	1	16	0	16	2.49	16	0.74	16	9.96	16	2.97

Remove redundancies

TR:Cell:CellLine	DA:RPcl	DV:RPcl	DA:Tc2	DV:Tc2	DA:Tc1	DV:Tc1	DA:Carbon1	DV:Carbon1	DA:Carbon2	DV:Carbon2	DA:Biomass	DV:Biomass
1	0	0	0	0	0	0	0	10	0	10	0	0
1	1	0	1	10.5	1	10.5	1	10	1	10	1	0
1	2	1	2	10.5	2	10.5	2	9.99	2	9.99	2	0.01
1	3	1	3	0	3	10.5	3	9.96	3	9.96	3	0.03
1	4	1	4	0	4	10.5	4	9.91	4	9.96	4	0.04
1	16	1	16	0	16	2.49	16	0.74	16	9.96	16	2.97

Discretize data according to a pathway activation threashold

TR:Cell:CellLine	DA:RPcl	DV:RPcl	DA:Tc2	DV:Tc2	DA:Tc1	DV:Tc1	DA:Carbon1	DV:Carbon1	DA:Carbon2	DV:Carbon2	DA:Biomass	DV:Biomass
1	0	0	0	0	0	0	0	1	0	1	0	0
1	1	0	1	1	1	1	1	1	1	1	1	0
1	2	1	2	1	2	1	2	1	2	1	2	0
1	3	1	3	0	3	1	3	1	3	1	3	1
1	4	1	4	0	4	1	4	1	4	1	4	1
1	16	1	16	0	16	1	16	1	16	1	16	1

Data pre-treatment

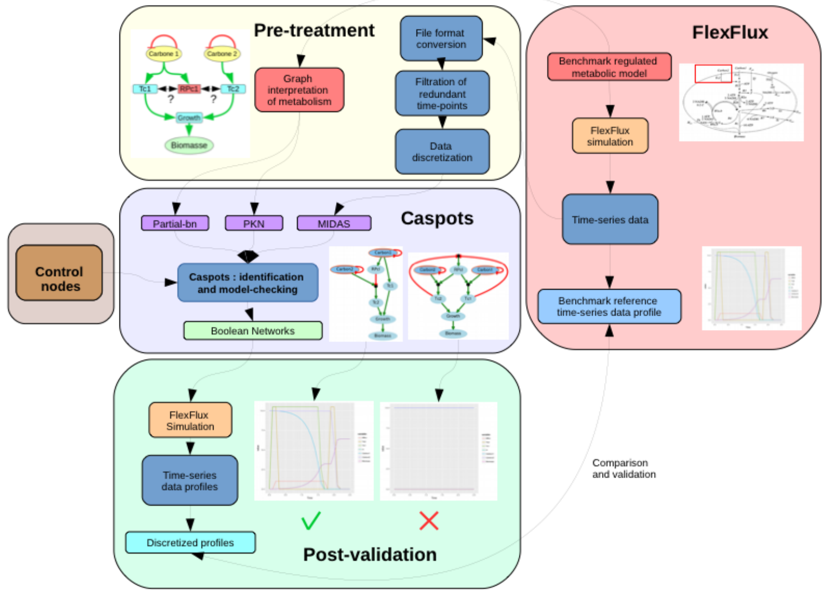
TR:Cell:CellLine	DA:RPcl	DV:RPcl	DA:Tc2	DV:Tc2	DA:Tc1	DV:Tc1	DA:Carbon1	DV:Carbon1	DA:Carbon2	DV:Carbon2	DA:Biomass	DV:Biomass
1	0	0	0	0	0	0	0	10	0	10	0	0
1	1	0	1	10.5	1	10.5	1	10	1	10	1	0
1	2	1	2	10.5	2	10.5	2	9.99	2	9.99	2	0.01
1	3	1	3	0	3	10.5	3	9.96	3	9.96	3	0.03
1	4	1	4	0	4	10.5	4	9.91	4	9.96	4	0.04
1	5	1	5	0	5	10.5	5	9.84	5	9.96	5	0.06
1	6	1	6	0	6	10.5	6	9.75	6	9.96	6	0.09
1	7	1	7	0	7	10.5	7	9.62	7	9.96	7	0.14
1	8	1	8	0	8	10.5	8	9.44	8	9.96	8	0.19
1	9	1	9	0	9	10.5	9	9.18	9	9.96	9	0.28
1	10	1	10	0	10	10.5	10	8.83	10	9.96	10	0.39
1	11	1	11	0	11	10.5	11	8.33	11	9.96	11	0.55
1	12	1	12	0	12	10.5	12	7.63	12	9.96	12	0.77
1	13	1	13	0	13	10.5	13	6.66	13	9.96	13	1.08
1	14	1	14	0	14	10.5	14	5.3	14	9.96	14	1.51
1	15	1	15	0	15	10.5	15	3.4	15	9.96	15	2.12
1	16	1	16	0	16	2.49	16	0.74	16	9.96	16	2.97

Remove redundancies

TR:Cell:CellLine	DA:RPcl	DV:RPcl	DA:Tc2	DV:Tc2	DA:Tc1	DV:Tc1	DA:Carbon1	DV:Carbon1	DA:Carbon2	DV:Carbon2	DA:Biomass	DV:Biomass
1	0	0	0	0	0	0	0	10	0	10	0	0
1	1	0	1	10.5	1	10.5	1	10	1	10	1	0
1	2	1	2	10.5	2	10.5	2	9.99	2	9.99	2	0.01
1	3	1	3	0	3	10.5	3	9.96	3	9.96	3	0.03
1	4	1	4	0	4	10.5	4	9.91	4	9.96	4	0.04
1	16	1	16	0	16	2.49	16	0.74	16	9.96	16	2.97

Discretize data according to a pathway activation threashold

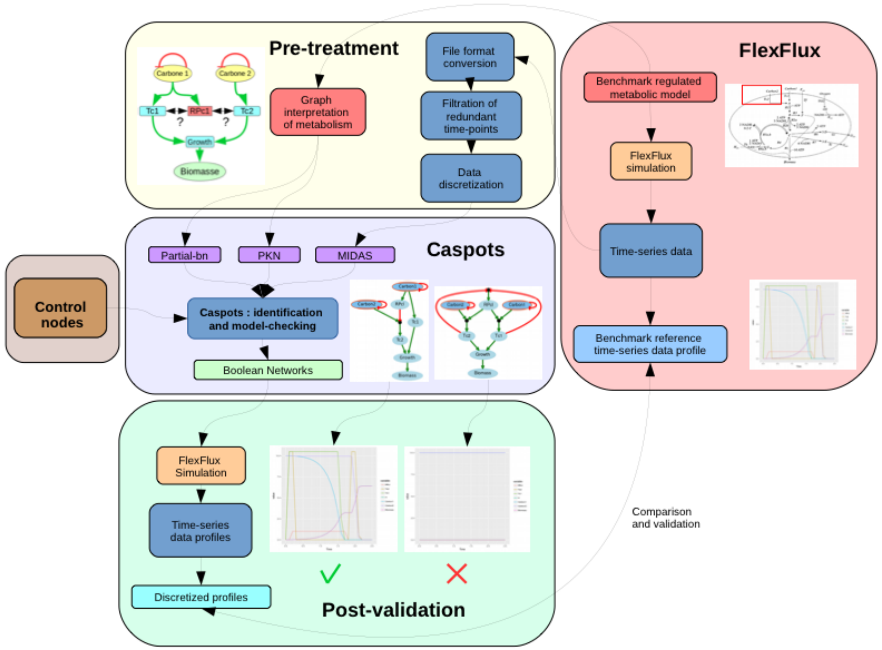
TR:Cell:CellLine	DA:RPcl	DV:RPcl	DA:Tc2	DV:Tc2	DA:Tc1	DV:Tc1	DA:Carbon1	DV:Carbon1	DA:Carbon2	DV:Carbon2	DA:Biomass	DV:Biomass
1	0	0	0	0	0	0	0	1	0	1	0	0
1	1	0	1	1	1	1	1	1	1	1	1	0
1	2	1	2	1	2	1	2	1	2	1	2	0
1	3	1	3	0	3	1	3	1	3	1	3	1
1	4	1	4	0	4	1	4	1	4	1	4	1
1	16	1	16	0	16	1	16	1	16	1	16	1



	PKN modelling			Input DataSet		Caspots version		Number of output networks		Running time
Pipeline parameters	Integration Graph	Interaction graph with retro-controls	Interaction graph with artificial links	Direct flux measurements	Discretized fluxes (threshold=0.1)	Without retro-control	With retro-controls	Learning	Comparison to benchmark dataset	
1	x			x		x		40	1	seconds
2		x		x		x		15	1	minutes
3			x	x		x		4	1	hours
4	x				x	x		37	1	seconds
5		x			x	x		12	1	minutes
6			x		x	x		1	1	hours
7	x			x			x	4	1	seconds
8		x		x			x	4	1	minutes
9			x	x			x	4	1	hours
10	x				x		x	1	1	seconds
11		x			x		x	1	1	minutes
12			x		x		x	1	1	hours

False-positive after caspots have disapeared

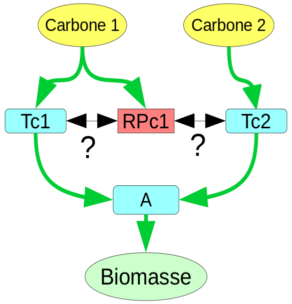
Which optimisation problem ?



$$\arg\min_{\phi : BNs}\Big(\underbrace{\mathrm{scoreControl}_{rss}((V,\phi),(P_1,\dots,P_n))}_{residual\ \ sum\ \ of\ \ squares},\ \underbrace{\mathrm{score}_{size}((V,\phi))}_{network\ size},\ \underbrace{\mathrm{score}_{simu}(met\_network,\phi)}_{flux-based\ simulation}\Big)$$

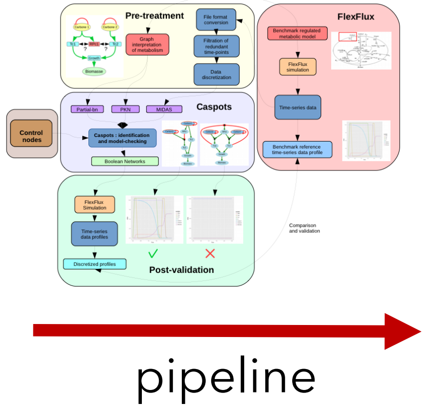
$$\mathrm{scoreControlFiltered}((V,\phi),\underbrace{(P_1,\dots,P_n)}_{n\ experiments}) = \sum_{\substack{t\ :\ \textbf{non redundant}\\ time\ points\\ P\ :\ experimentations\\ v\ :\ variable\\ v\ \neq\ control}} \left(\underbrace{\mathrm{discretised}(\mathrm{measure}(v,P,t))}_{observations\in\{0,1\}} - \underbrace{\phi_P(v,t)}_{predictions\in\{0,1\}}\right)$$

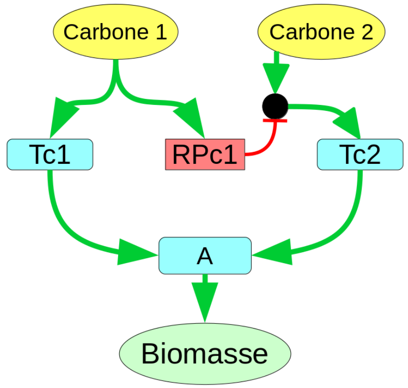
Conclusion : we did it…



+

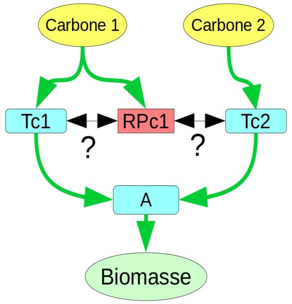
TR:Cell:CellLine	DA:RPcl	DV:RPcl	DA:Tc2	DV:Tc2	DA:Tc1	DV:Tc1	DA:Carbon1	DV:Carbon1	DA:Carbon2	DV:Carbon2	DA:Biomass	DV:Biomass
1	0	0	0	0	0	0	0	10	0	10	0	0
1	1	0	1	10.5	1	10.5	1	10	1	10	1	0
1	2	1	2	10.5	2	10.5	2	9.99	2	9.99	2	0.01
1	3	1	3	0	3	10.5	3	9.96	3	9.96	3	0.03
1	4	1	4	0	4	10.5	4	9.91	4	9.96	4	0.04
1	5	1	5	0	5	10.5	5	9.84	5	9.96	5	0.06
1	6	1	6	0	6	10.5	6	9.75	6	9.96	6	0.09
1	7	1	7	0	7	10.5	7	9.62	7	9.96	7	0.14
1	8	1	8	0	8	10.5	8	9.44	8	9.96	8	0.19
1	9	1	9	0	9	10.5	9	9.18	9	9.96	9	0.28
1	10	1	10	0	10	10.5	10	8.83	10	9.96	10	0.39
1	11	1	11	0	11	10.5	11	8.33	11	9.96	11	0.55
1	12	1	12	0	12	10.5	12	7.63	12	9.96	12	0.77
1	13	1	13	0	13	10.5	13	6.66	13	9.96	13	1.08
1	14	1	14	0	14	10.5	14	5.3	14	9.96	14	1.51
1	15	1	15	0	15	10.5	15	3.4	15	9.96	15	2.12
1	16	1	16	0	16	2.49	16	0.74	16	9.96	16	2.97
1	17	1	17	0	17	0	17	0	17	9.96	17	3.22
1	18	0	18	0	18	0	18	0	18	9.96	18	3.22
1	19	0	19	10.5	19	0	19	0	19	9.96	19	3.22
1	20	0	20	10.5	20	0	20	0	20	5.93	20	4.5
1	21	0	21	0.49	21	0	21	0	21	0.31	21	6.29
1	22	0	22	0	22	0	22	0	22	0	22	6.39





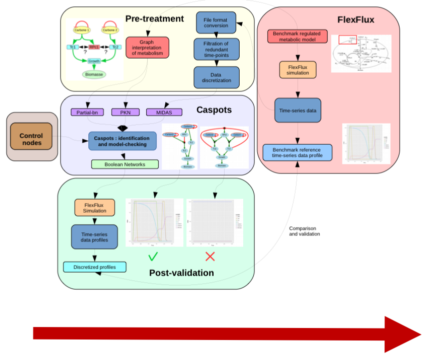
Uniquely recover the rule of a very simple regulated metabolic network

Conclusion : we did it…

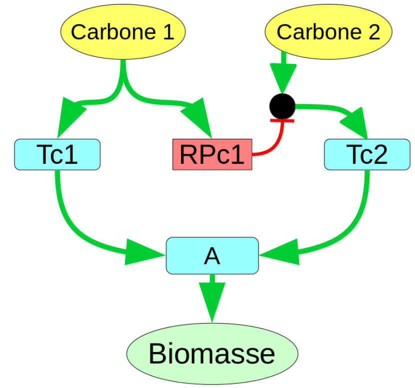


+

TR:Cell:CellLine	DA:RPcl	DV:RPcl	DA:Tc2	DV:Tc2	DA:Tc1	DV:Tc1	DA:Carbon1	DV:Carbon1	DA:Carbon2	DV:Carbon2	DA:Biomass	DV:Biomass
1	0	0	0	0	0	0	0	10	0	10	0	0
1	1	0	1	10.5	1	10.5	1	10	1	10	1	0
1	2	1	2	10.5	2	10.5	2	9.99	2	9.99	2	0.01
1	3	1	3	0	3	10.5	3	9.96	3	9.96	3	0.03
1	4	1	4	0	4	10.5	4	9.91	4	9.96	4	0.04
1	5	1	5	0	5	10.5	5	9.84	5	9.96	5	0.06
1	6	1	6	0	6	10.5	6	9.75	6	9.96	6	0.09
1	7	1	7	0	7	10.5	7	9.62	7	9.96	7	0.14
1	8	1	8	0	8	10.5	8	9.44	8	9.96	8	0.19
1	9	1	9	0	9	10.5	9	9.18	9	9.96	9	0.28
1	10	1	10	0	10	10.5	10	8.83	10	9.96	10	0.39
1	11	1	11	0	11	10.5	11	8.33	11	9.96	11	0.55
1	12	1	12	0	12	10.5	12	7.63	12	9.96	12	0.77
1	13	1	13	0	13	10.5	13	6.66	13	9.96	13	1.08
1	14	1	14	0	14	10.5	14	5.3	14	9.96	14	1.51
1	15	1	15	0	15	10.5	15	3.4	15	9.96	15	2.12
1	16	1	16	0	16	2.49	16	0.74	16	9.96	16	2.97
1	17	1	17	0	17	0	17	0	17	9.96	17	3.22
1	18	0	18	0	18	0	18	0	18	9.96	18	3.22
1	19	0	19	10.5	19	0	19	0	19	9.96	19	3.22
1	20	0	20	10.5	20	0	20	0	20	5.93	20	4.5
1	21	0	21	0.49	21	0	21	0	21	0.31	21	6.29
1	22	0	22	0	22	0	22	0	22	0	22	6.39



pipeline



Uniquely recover the rule of a very simple regulated metabolic network

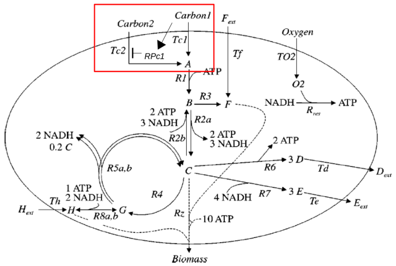
$$\underset{\phi\,:\,BNs}{\arg\min}(\underbrace{\text{score}_{rss}((V,\phi),(P_1,\ldots,P_n))}_{residual\ sum\ of\ squares},\underbrace{\text{score}_{size}((V,\phi))}_{complexity})$$

$$\underset{\phi\,:\,BNs}{\arg\min}(\underbrace{\text{scoreControlFiltered}((V,\phi),(P_1,\ldots,P_n))}_{distance\ to\ pretreated\ data},\underbrace{\text{score}_{size}((V,\phi))}_{network\ size},\underbrace{\text{score}_{simu}(met\_network,\phi)}_{flux-based\ simulation})$$

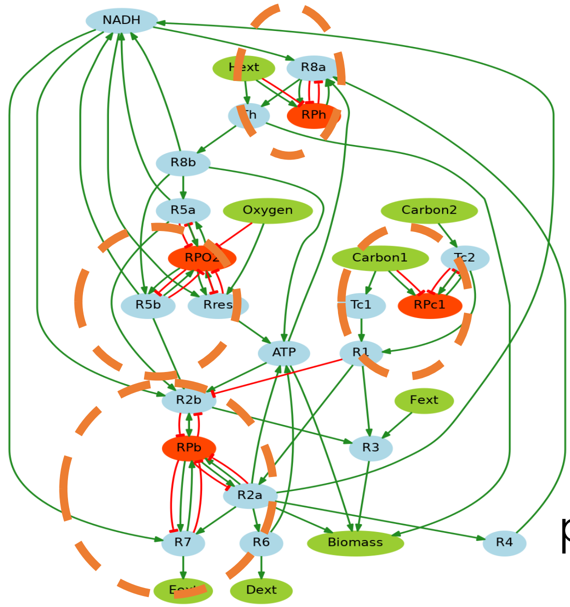
$$\text{scoreControlFiltered}((V,\phi),\underbrace{(P_1,\ldots,P_n)}_{n\ experiments}) = \sum_{\substack{t\,:\,\textbf{non redundant}\\ time\ points\\ P\,:\,experimentations\\ v\,:\,variable\\ v\,\neq\,control}} \Big(\underbrace{\text{discretised}(\text{measure}(v,P,t))}_{observations\in\{0,1\}} - \underbrace{\phi_P(v,t)}_{predictions\in\{0,1\}}\Big)$$

Complete reformulation
of the optimisation problem.

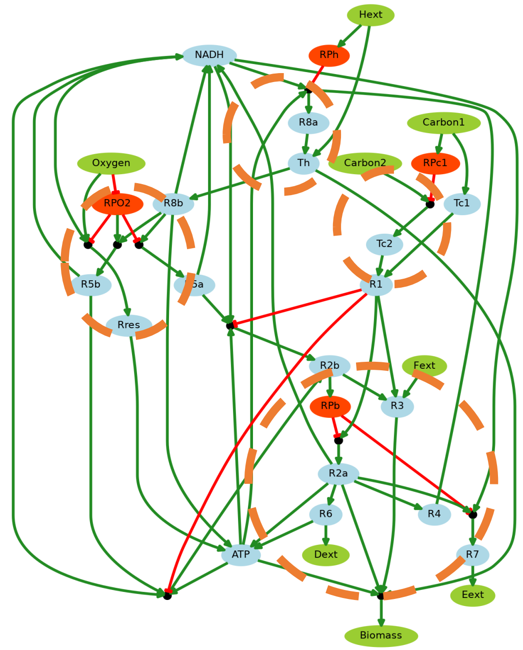
What happens for a larger example ?



Extended interaction graph



pipeline

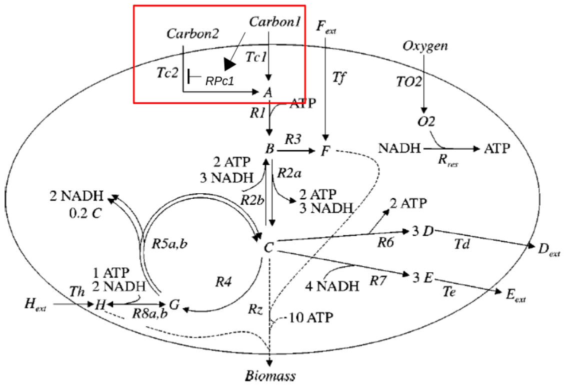


+ data

TR:Cell:CellLine	DA:RPcl	DV:RPcl	DA:Tc2	DV:Tc2	DA:Tc1	DV:Tc1	DA:Carbon1	DV:Carbon1	DA:Carbon2	DV:Carbon2	DA:Biomass	DV:Biomass
1	0	0	0	0	0	0	0	10	0	10	0	0
1	1	0	1	10.5	1	10.5	1	10	1	10	1	0
1	2	1	2	10.5	2	10.5	2	9.99	2	9.99	2	0.01
1	3	1	3	0	3	10.5	3	9.96	3	9.96	3	0.03
1	4	1	4	0	4	10.5	4	9.91	4	9.96	4	0.04
1	5	1	5	0	5	10.5	5	9.84	5	9.96	5	0.06
1	6	1	6	0	6	10.5	6	9.75	6	9.96	6	0.09
1	7	1	7	0	7	10.5	7	9.62	7	9.96	7	0.14
1	8	1	8	0	8	10.5	8	9.44	8	9.96	8	0.19
1	9	1	9	0	9	10.5	9	9.18	9	9.96	9	0.28
1	10	1	10	0	10	10.5	10	8.83	10	9.96	10	0.39
1	11	1	11	0	11	10.5	11	8.33	11	9.96	11	0.55
1	12	1	12	0	12	10.5	12	7.63	12	9.96	12	0.77
1	13	1	13	0	13	10.5	13	6.66	13	9.96	13	1.08
1	14	1	14	0	14	10.5	14	5.3	14	9.96	14	1.51
1	15	1	15	0	15	10.5	15	3.4	15	9.96	15	2.12
1	16	1	16	0	16	2.49	16	0.74	16	9.96	16	2.97
1	17	1	17	0	17	0	17	0	17	9.96	17	3.22
1	18	0	18	0	18	0	18	0	18	9.96	18	3.22
1	19	0	19	10.5	19	0	19	0	19	9.96	19	3.22
1	20	0	20	10.5	20	0	20	0	20	5.93	20	4.5
1	21	0	21	0.49	21	0	21	0	21	0.31	21	6.29
1	22	0	22	0	22	0	22	0	22	0	22	6.39

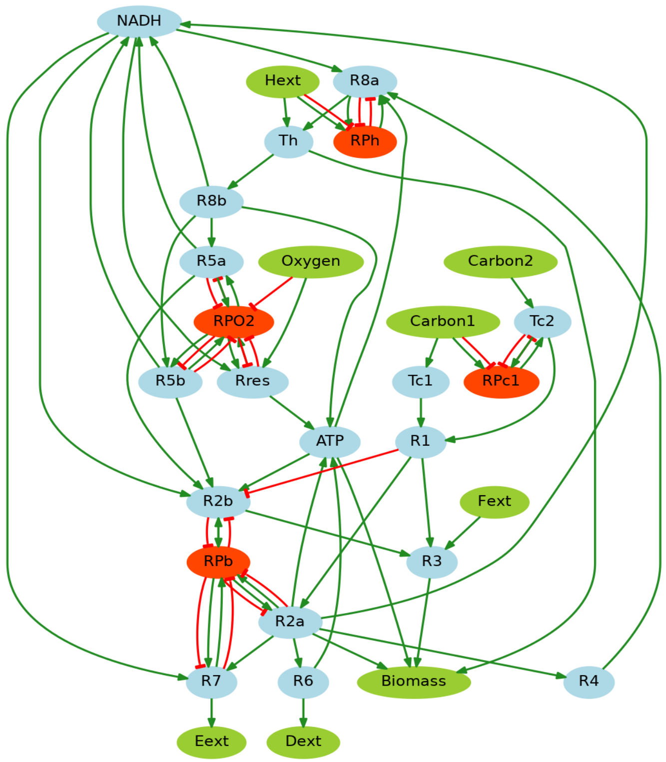
Uniquely recover the expected regulations

Where did we cheat… to be continued



?

Extended interaction graph



- Remove spurious metabolic cycles
- Interpret optimization flux-based constraints as forbidden co-activations
- Model the role of co-factors
- Model stoechiometry constraints
- (…)

Last (but not least) bottleneck : how to automatically interprete a metabolic network into an interaction graph ?