

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

CNRS Rennes: Aurélien Cornet, Anne Siegel

Inra Toulouse: Caroline Barouck, Ludovic Cottret

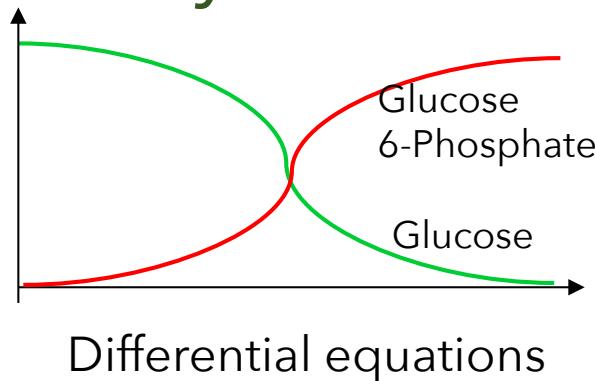
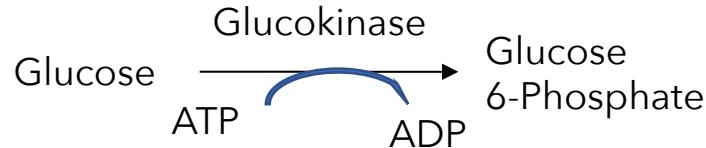
CNRS Saclay: Loïc Paulevé

Frei Berlin Univ: Alexander Bockmayr, Heike Siebert.



Modeling frameworks in systems biology

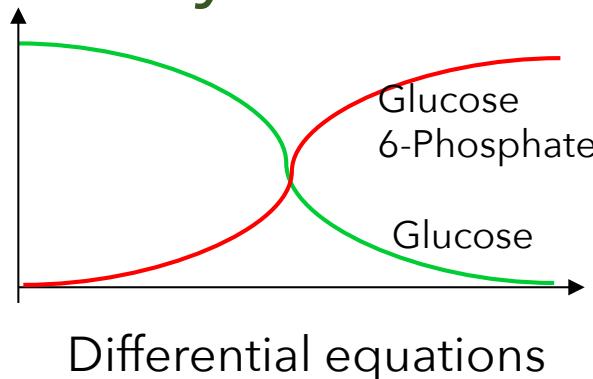
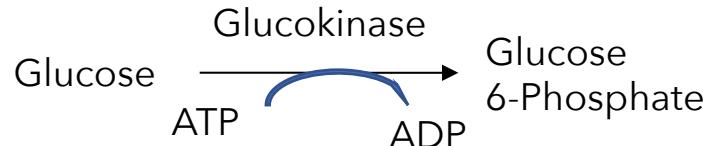
Metabolism



Entries are consumed by the reaction

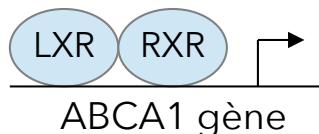
Modeling frameworks in systems biology

Metabolism



Entries are consumed by the reaction

Signaling and regulation



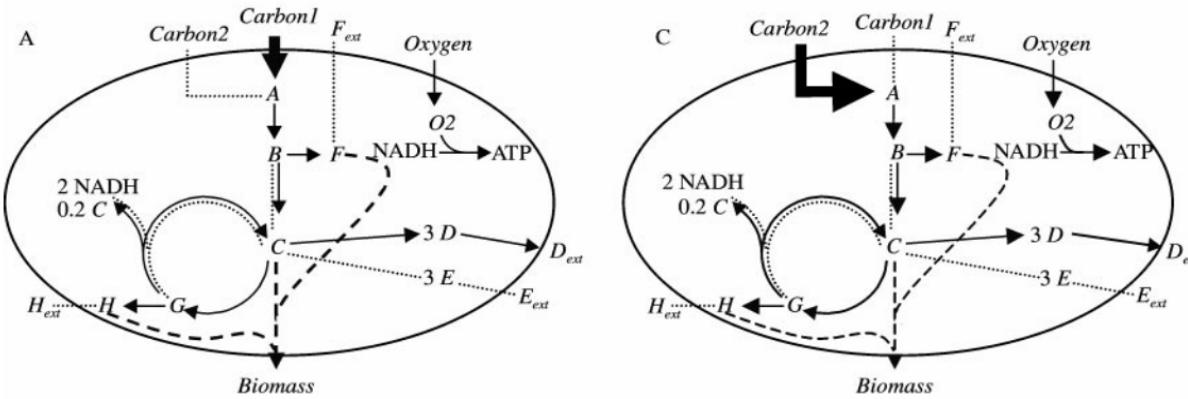
$$\begin{array}{c} \text{LXR} \\ \text{RXR} \\ \text{ABCA1} \end{array} \xrightarrow{\left(\begin{array}{c} 1 \\ 1 \\ 0 \end{array} \right)} \left(\begin{array}{c} 1 \\ 1 \\ 1 \end{array} \right)$$

Logical rules

Entries are not modified by the interaction

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

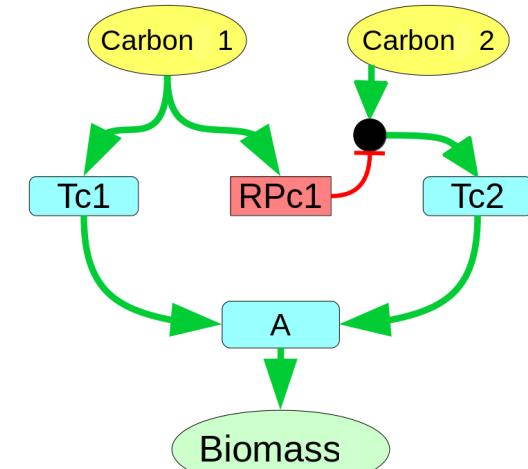
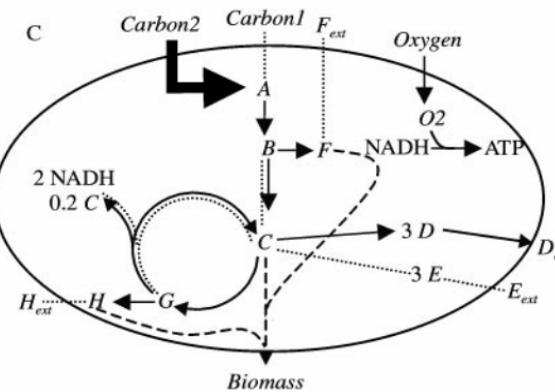
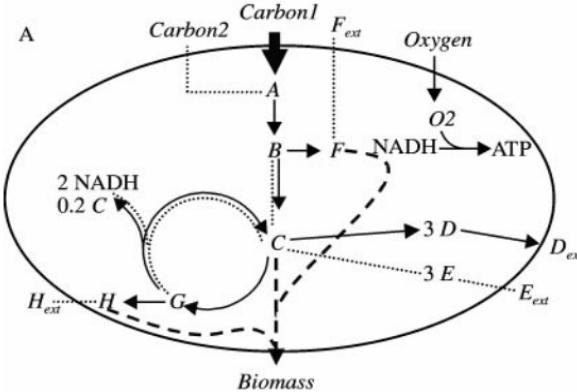
Carbon metabolism : diauxic shift



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: R_{PC1} regulates TC2**

Boolean rules

- $R_{Pc1} = \text{Carbon1}$
- $Tc2 = \text{Carbon2} + !R_{Pc1}$

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

Metabolic reactions

- 1 A - 1 ATP + 1 B
- 1 B + 2 ATP + 2 NADH + 1 C
- 1 C - 2 ATP - 2 NADH + 1 B
- 1 B + 1 F
- 1 C + 1 G
- 1 G + 0.8 C + 2 NADH
- 1 G + 0.8 C + 2 NADH
- 1 C + 2 ATP + 3 D
- 1 C - 4 NADH + 3 E
- 1 G - 1 ATP - 2 NADH + 1 H
- +1 G + 1 ATP + 2 NADH - 1 H
- 1 NADH - 1 O₂ + 1 ATP

Transport processes

- 1 Carbon1 + 1 A
- 1 Carbon2 + 1 A
- 1 F_{ext} + 1 F
- 1 D + 1 D_{ext}
- 1 E + 1 E_{ext}
- 1 H_{ext} + 1 H
- 1 Oxygen + 1 O₂

R1
R2a
R2b
R3
R4
R5a
R5b
R6
R7
R8a
R8b
Rres

IF NOT(RPh)

IF NOT (RPO2)
IF RPO2

IF NOT (RPb)
IF NOT (RPh)

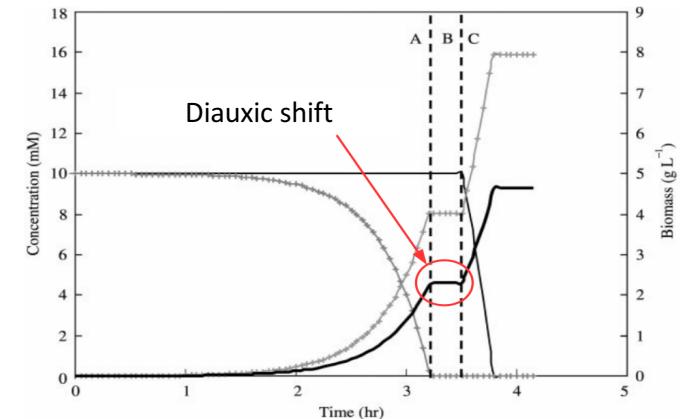
IF NOT (RPO2)

IF NOT(RPc1) AND IF(Carbon2)

List of regulations

Tc1
Tc2
Tf
Td
Te
Th
To2

→
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

Metabolic reactions	
- 1 A	- 1 ATP + 1 B
- 1 B	+ 2 ATP + 2 NADH + 1 C
- 1 C	- 2 ATP - 2 NADH + 1 B
- 1 B	+ 1 F
- 1 C	+ 1 G
- 1 G	+ 0.8 C + 2 NADH
- 1 G	+ 0.8 C + 2 NADH
- 1 C	+ 2 ATP + 3 D
- 1 C	- 4 NADH + 3 E
- 1 G	- 1 ATP - 2 NADH + 1 H
+ 1 G	+ 1 ATP + 2 NADH - 1 H
- 1 NADH	- 1 O ₂ + 1 ATP
Transport processes	
- 1 Carbon1	+ 1 A
- 1 Carbon2	+ 1 A
- 1 F _{ext}	+ 1 F
- 1 D	+ 1 D _{ext}
- 1 E	+ 1 E _{ext}
- 1 H _{ext}	+ 1 H
- 1 Oxygen	+ 1 O ₂

R1
R2a
R2b
R3
R4
R5a
R5b
R6
R7
R8a
R8b
Rres

IF NOT(RPh)

IF NOT (RPO2)
IF RPO2

IF NOT (RPb)
IF NOT (RPh)

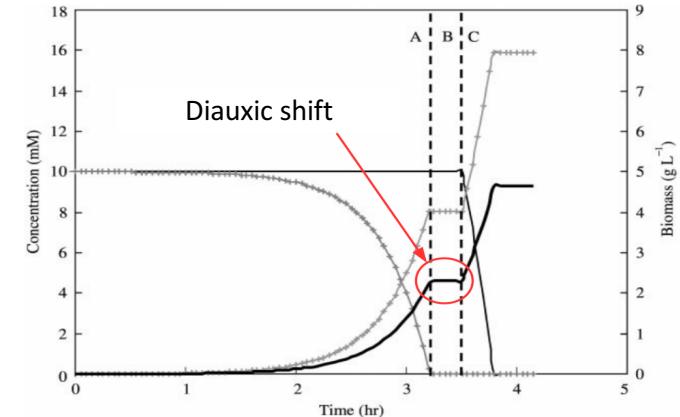
IF NOT (RPO2)

IF NOT(RPc1) AND IF(Carbon2)

List of regulations

Tc1
Tc2
Tf
Td
Te
Th
To2

→
FlexFlux,
(Cottret et al, 2015)

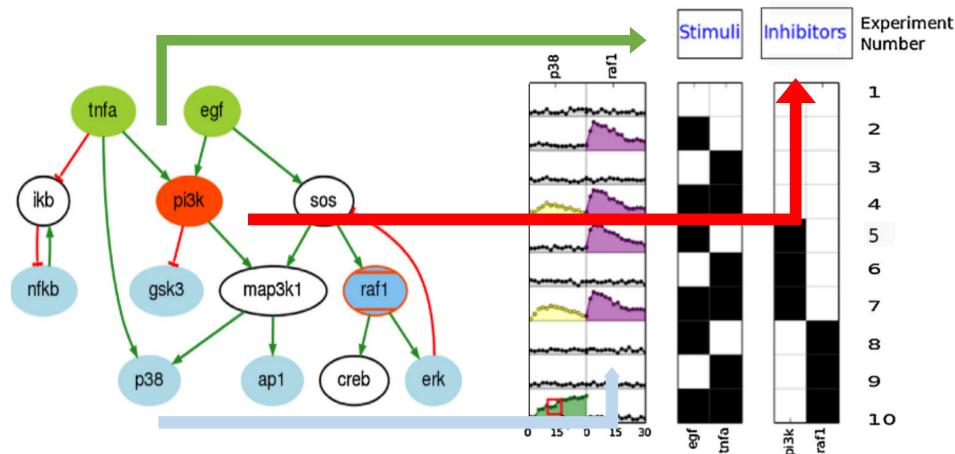


Covert et al., 2001

- **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

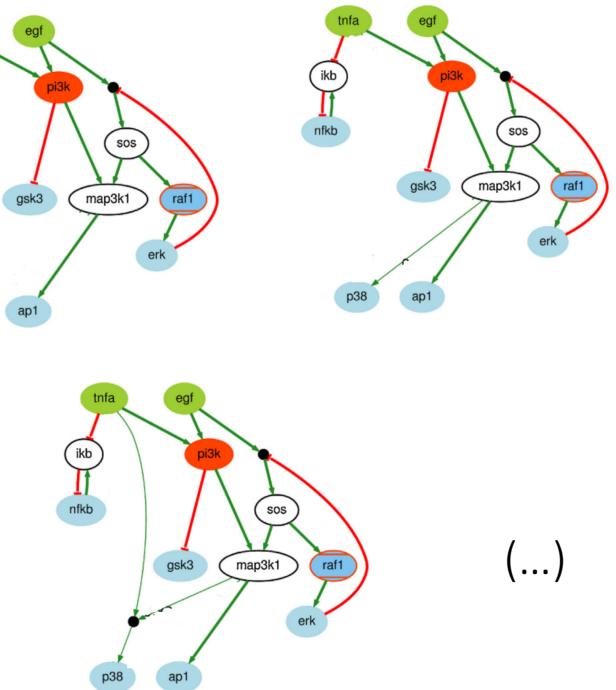


Caspo time series



Ostrowski et al., 2016

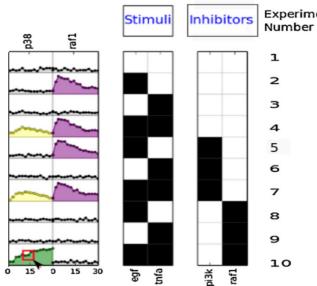
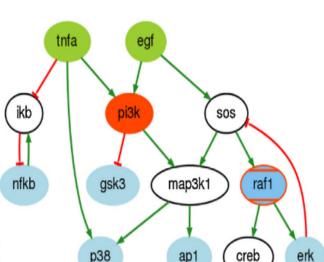
- **PKN** : interaction graph derived from literature and data
 - stimuli, inhibitors, readouts
- **Data**: measures of graph entities in different experimental conditions



Several logical networks

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

caspo(ts): underlying optimization problem



Ostrowski et al., 2016

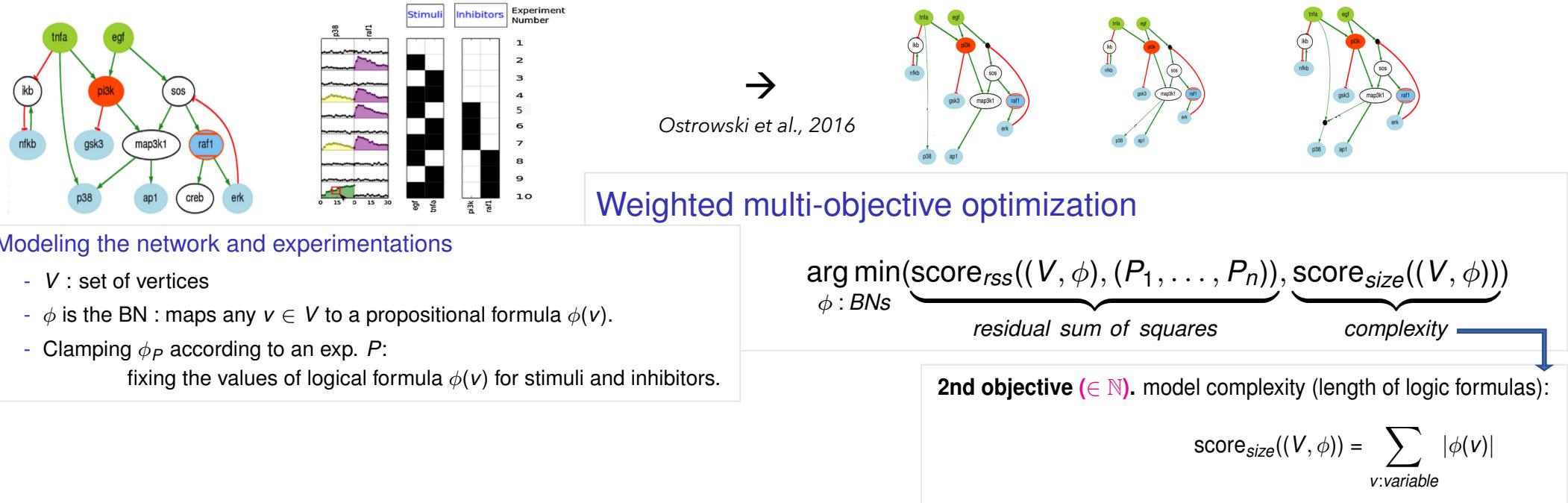
Weighted multi-objective optimization

Modeling the network and experimentations

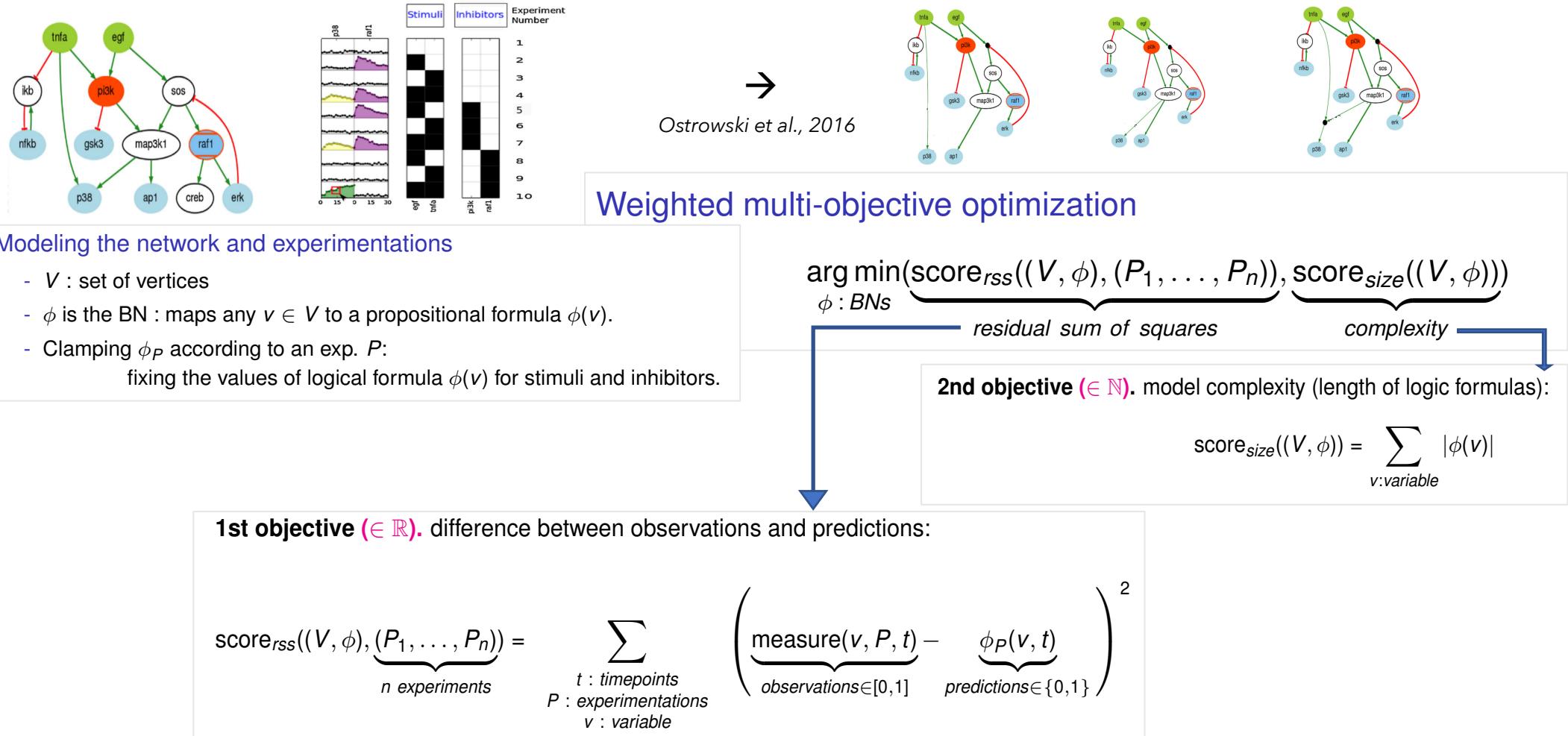
- V : set of vertices
- ϕ is the BN : maps any $v \in V$ to a propositional formula $\phi(v)$.
- Clamping ϕ_P according to an exp. P :
fixing the values of logical formula $\phi(v)$ for stimuli and inhibitors.

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

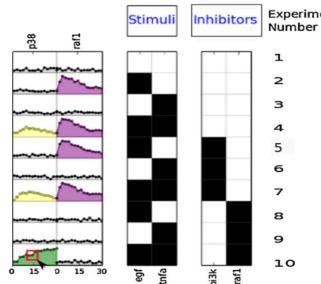
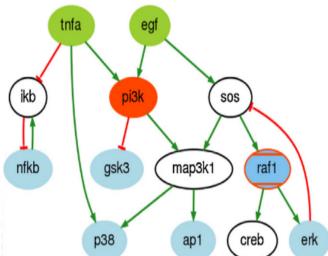
caspo(ts): underlying optimization problem



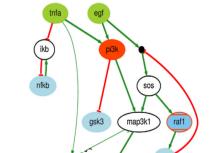
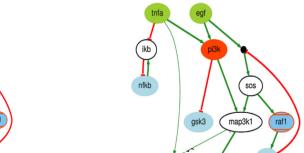
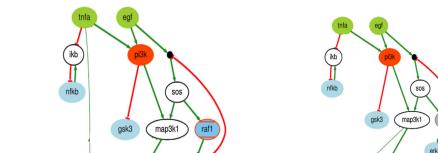
caspo(ts): underlying optimization problem



caspo(ts): underlying optimization problem



Ostrowski et al., 2016



Weighted multi-objective optimization

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

All optimal BNs equivalently explain the data

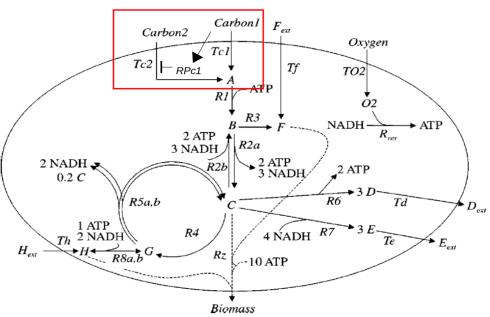
1st objective ($\in \mathbb{R}$). difference between observations and predictions:

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

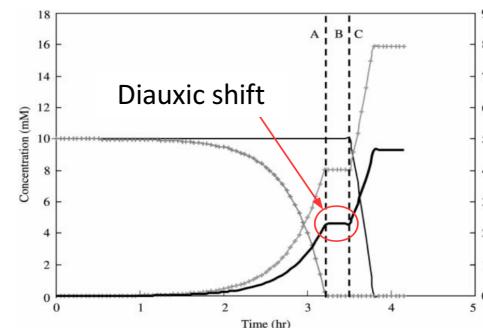
Introduce a threshold in data at 0.5

Stimuli and inhibitors have fixed values

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

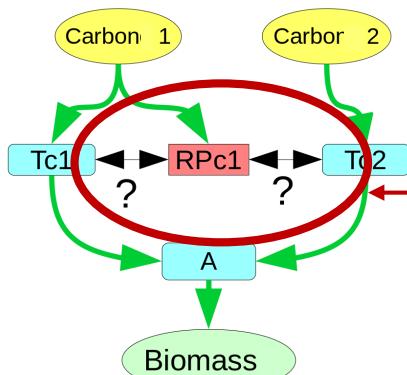


simulated data

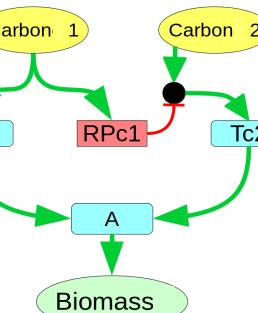
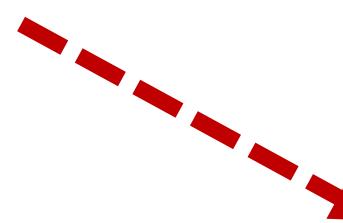


TR:Cell:CellLine	DA:RPC1	DA:RPc1	DA:Tc2	DA:Tc1	DA:Tc1	DA:Carbon1	DA:Carbon1	DA:Carbon2	DA:Carbon2	DA:Biomass	DA:Biomass
1	0	0	0	0	0	0	0	10	0	10	0
1	1	0	1	10.5	1	10.5	1	10	1	10	1
1	2	1	2	10.5	2	10.5	2	9.99	2	9.99	2
1	3	1	3	0	3	10.5	3	9.98	3	9.98	3
1	4	1	4	0	4	10.5	4	9.91	4	9.96	4
1	5	1	5	0	5	10.5	5	9.84	5	9.96	5
1	6	1	6	0	6	10.5	6	9.75	6	9.96	6
1	7	1	7	0	7	10.5	7	9.62	7	9.96	7
1	8	1	8	0	8	10.5	8	9.44	8	9.96	8
1	9	1	9	0	9	10.5	9	9.18	9	9.96	9
1	10	1	10	0	10	10.5	10	8.83	10	9.96	10
1	11	1	11	0	11	10.5	11	9.33	11	9.96	11
1	12	1	12	0	12	10.5	12	7.63	12	9.96	12
1	13	1	13	0	13	10.5	13	6.66	13	9.96	13
1	14	1	14	0	14	10.5	14	5.3	14	9.96	14
1	15	1	15	0	15	10.5	15	3.4	15	9.96	15
1	16	1	16	0	16	2.49	16	0.74	16	9.96	16
1	17	1	17	0	17	0	17	0	17	9.96	17
1	18	0	18	0	18	0	18	0	18	9.96	18
1	19	0	19	10.5	19	0	19	0	19	9.96	19
1	20	0	20	10.5	20	0	20	0	20	5.93	20
1	21	0	21	0.49	21	0	21	0	21	0.31	21
1	22	0	22	0	22	0	22	0	22	0	6.39

↓ Prior knowledge network
On a very small toy-example

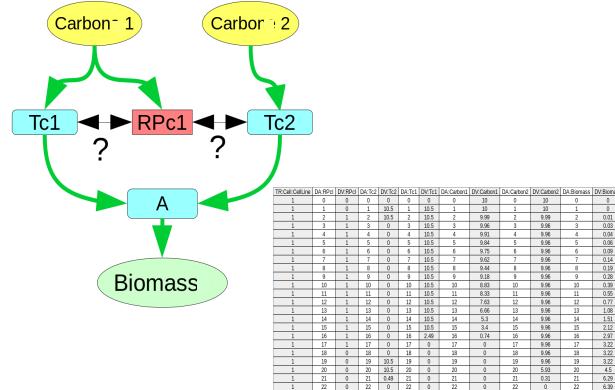


Search among all possible boolean interactions between RPc1, TC1 and TC2



TC2 = (Carbon2) AND (NOT RPc1)

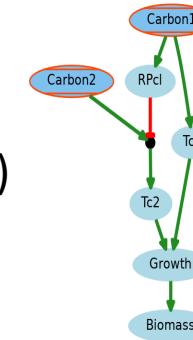
Basic approach based on casspots



40 models ...



(....)

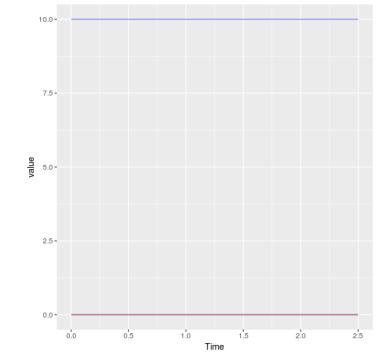
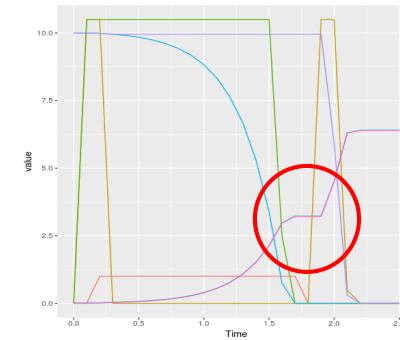
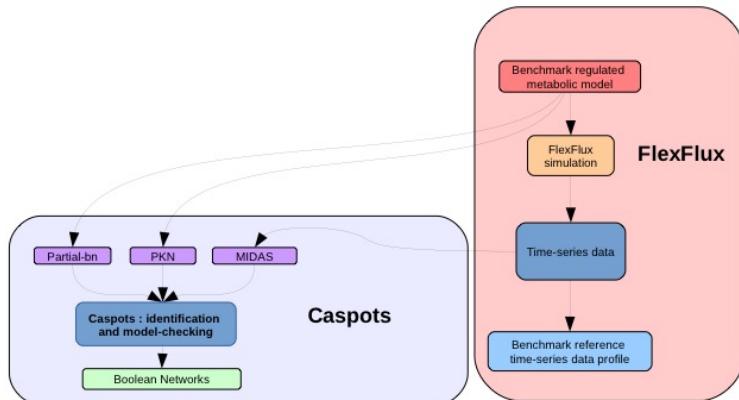


(....)

(....)

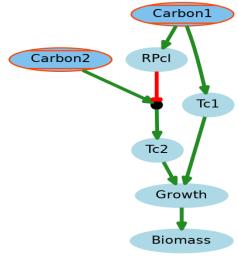
↓ Simulations ...

Feed casspots with simulation data

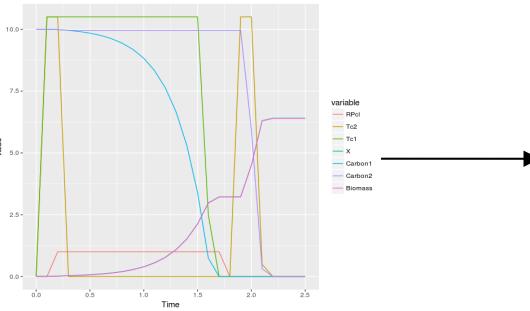


Only one model among the 40 solutions show a diauxic shift in the simulation

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)

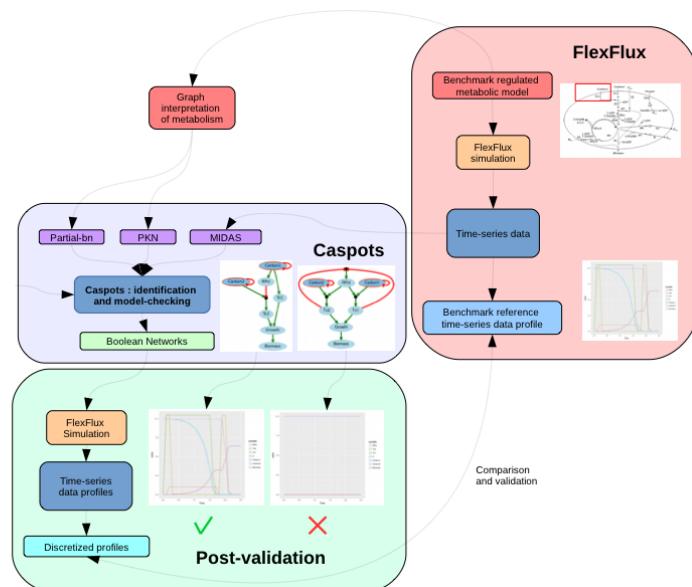
TR:Cell:CellLine	DA:RPcl	DV:RPcl	DA:Tc2	DV:Tc2	DA:Tc1	DV:Tc1	DA:Carbon1	DV:Carbon1	DA:Carbon2	DV:Carbon2	DA:Biomass	DV:Biomass
1	0	=	0	*	0	+	0	=	0	=	0	=
1	1	*	1	*	1	*	1	*	1	*	1	*
1	2	*	2	*	2	*	2	*	2	*	2	*
1	3	*	3	*	3	*	3	*	3	*	3	*
1	4	*	4	*	4	*	4	*	4	*	4	*
1	16	*	16	*	16	*	16	*	16	*	16	*
1	18	*	18	*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

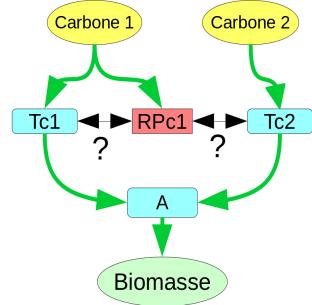
Weighted multi-objective optimization

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

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

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

Reducing the number of false-positive ?

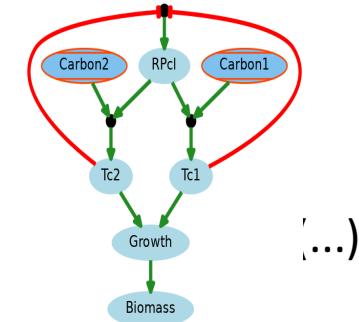


	Tc1	Carbone 1	Carbone 2	RPc1	Tc2	A	Biomasse
1	9	9	9	1	10.5	1	10
1	1	0	1	10.5	1	10	1
1	0	0	0	0	0	0	0
1	3	3	3	0	20.5	3	8.99
1	0	0	0	0	0	0	0
1	5	5	5	20.5	5	8.94	5
1	6	3	6	0	30.5	6	8.75
1	0	0	0	0	0	0	0
1	8	1	8	0	20.5	8	8.46
1	0	0	0	0	0	0	0
1	20	3	30	0	30.5	10	8.83
1	11	3	11	0	30.5	11	8.83
1	20	3	30	0	30.5	12	8.83
1	33	3	33	0	20.5	13	8.66
1	24	3	34	0	20.5	14	8.66
1	25	3	35	0	20.5	15	8.4
1	36	3	36	0	16	2.49	0.74
1	24	3	37	0	16	2.49	0.74
1	38	0	38	0	18	0	8.99
1	29	0	39	0	18	0	8.99
1	39	0	40	0	20	0	8.99
1	21	0	21	0.48	21	0	7.31
1	22	0	22	0	21	0	7.31

40 models ...

→ (...)

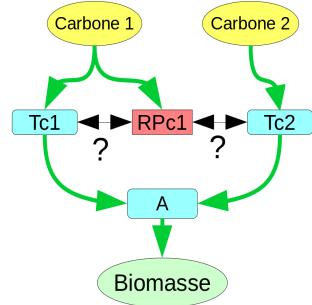
(...)



?

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

Reducing the number of false-positive ?



	Tc1	Carbone 1	Carbone 2	RPc1	Tc2	A	Biomasse
1	0	0	0	0	0	0	0
1	1	0	1	10.5	1	10.5	1
1	2	0	2	21.0	2	21.0	2
1	3	0	3	31.5	3	31.5	3
1	4	0	4	42.0	4	42.0	4
1	5	0	5	52.5	5	52.5	5
1	6	0	6	63.0	6	63.0	6
1	7	0	7	73.5	7	73.5	7
1	8	0	8	84.0	8	84.0	8
1	9	0	9	94.5	9	94.5	9
1	10	0	10	105.0	10	105.0	10
1	11	0	11	10.5	11	8.33	11
1	12	0	12	21.0	12	18.33	12
1	13	0	13	31.5	13	28.33	13
1	14	0	14	42.0	14	38.33	14
1	15	0	15	52.5	15	48.33	15
1	16	0	16	2.49	16	0.74	16
1	17	0	17	10.5	17	3.12	17
1	18	0	18	0	18	0	18
1	19	0	19	21.0	19	6.33	19
1	20	0	20	31.5	20	9.33	20
1	21	0	21	42.0	21	12.33	21
1	22	0	22	52.5	22	15.33	22
1	23	0	23	0.48	23	0	23
1	24	0	24	0	24	0	24
1	25	0	25	0	25	0	25
1	26	0	26	0	26	0	26
1	27	0	27	0	27	0	27
1	28	0	28	0	28	0	28
1	29	0	29	0	29	0	29
1	30	0	30	0	30	0	30
1	31	0	31	0	31	0	31
1	32	0	32	0	32	0	32
1	33	0	33	0	33	0	33
1	34	0	34	0	34	0	34
1	35	0	35	0	35	0	35
1	36	0	36	0	36	0	36
1	37	0	37	0	37	0	37
1	38	0	38	0	38	0	38
1	39	0	39	0	39	0	39
1	40	0	40	0	40	0	40
1	41	0	41	0	41	0	41
1	42	0	42	0	42	0	42
1	43	0	43	0	43	0	43
1	44	0	44	0	44	0	44
1	45	0	45	0	45	0	45
1	46	0	46	0	46	0	46
1	47	0	47	0	47	0	47
1	48	0	48	0	48	0	48
1	49	0	49	0	49	0	49
1	50	0	50	0	50	0	50
1	51	0	51	0	51	0	51
1	52	0	52	0	52	0	52
1	53	0	53	0	53	0	53
1	54	0	54	0	54	0	54
1	55	0	55	0	55	0	55
1	56	0	56	0	56	0	56
1	57	0	57	0	57	0	57
1	58	0	58	0	58	0	58
1	59	0	59	0	59	0	59
1	60	0	60	0	60	0	60
1	61	0	61	0	61	0	61
1	62	0	62	0	62	0	62
1	63	0	63	0	63	0	63
1	64	0	64	0	64	0	64
1	65	0	65	0	65	0	65
1	66	0	66	0	66	0	66
1	67	0	67	0	67	0	67
1	68	0	68	0	68	0	68
1	69	0	69	0	69	0	69
1	70	0	70	0	70	0	70
1	71	0	71	0	71	0	71
1	72	0	72	0	72	0	72
1	73	0	73	0	73	0	73
1	74	0	74	0	74	0	74
1	75	0	75	0	75	0	75
1	76	0	76	0	76	0	76
1	77	0	77	0	77	0	77
1	78	0	78	0	78	0	78
1	79	0	79	0	79	0	79
1	80	0	80	0	80	0	80
1	81	0	81	0	81	0	81
1	82	0	82	0	82	0	82
1	83	0	83	0	83	0	83
1	84	0	84	0	84	0	84
1	85	0	85	0	85	0	85
1	86	0	86	0	86	0	86
1	87	0	87	0	87	0	87
1	88	0	88	0	88	0	88
1	89	0	89	0	89	0	89
1	90	0	90	0	90	0	90
1	91	0	91	0	91	0	91
1	92	0	92	0	92	0	92
1	93	0	93	0	93	0	93
1	94	0	94	0	94	0	94
1	95	0	95	0	95	0	95
1	96	0	96	0	96	0	96
1	97	0	97	0	97	0	97
1	98	0	98	0	98	0	98
1	99	0	99	0	99	0	99
1	100	0	100	0	100	0	100

40 models ...

(...)

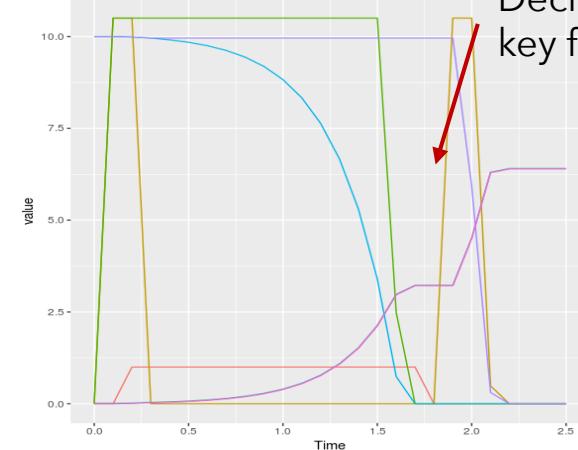
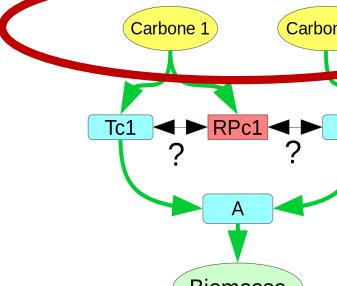
(...)

(...)

?

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

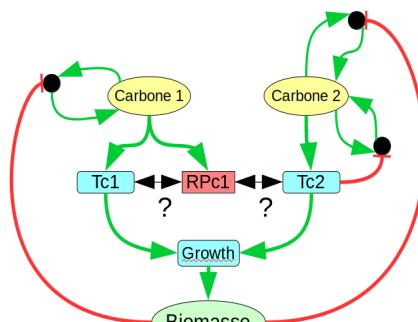
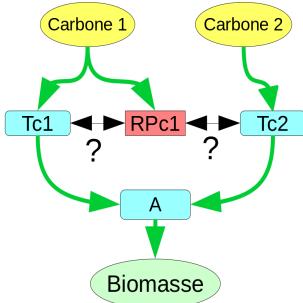
Fixed by the clamping operation in caspots



Decreasing carbons in a key feature of the process

variable
RPc1
Tc2
Tc1
X
Carbon1
Carbon2
Biomass

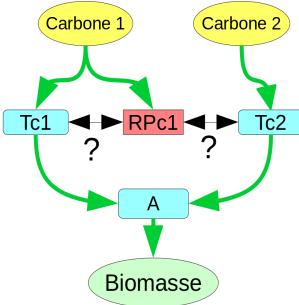
A better modeling for carbon inputs



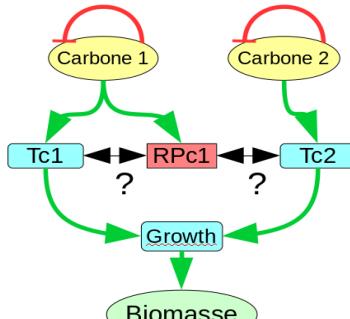
Feed-forward loop

Artificial logical dependency

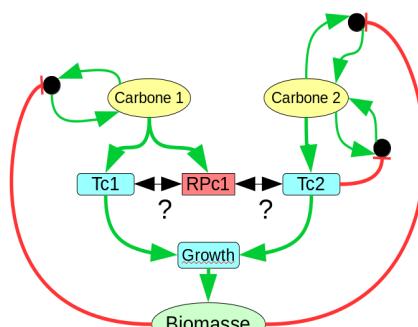
A better modeling for carbon inputs



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



Feed-forward loop

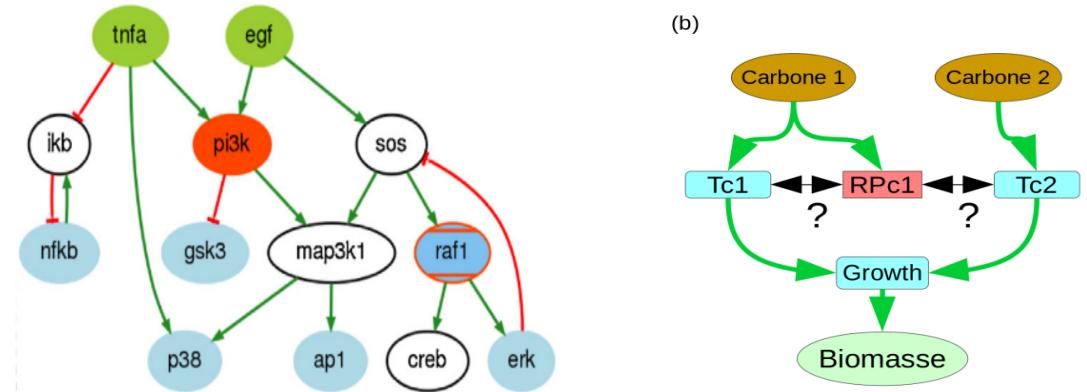


Artificial logical dependency

Introducing artificial dependencies highly reduces the number of false-positive

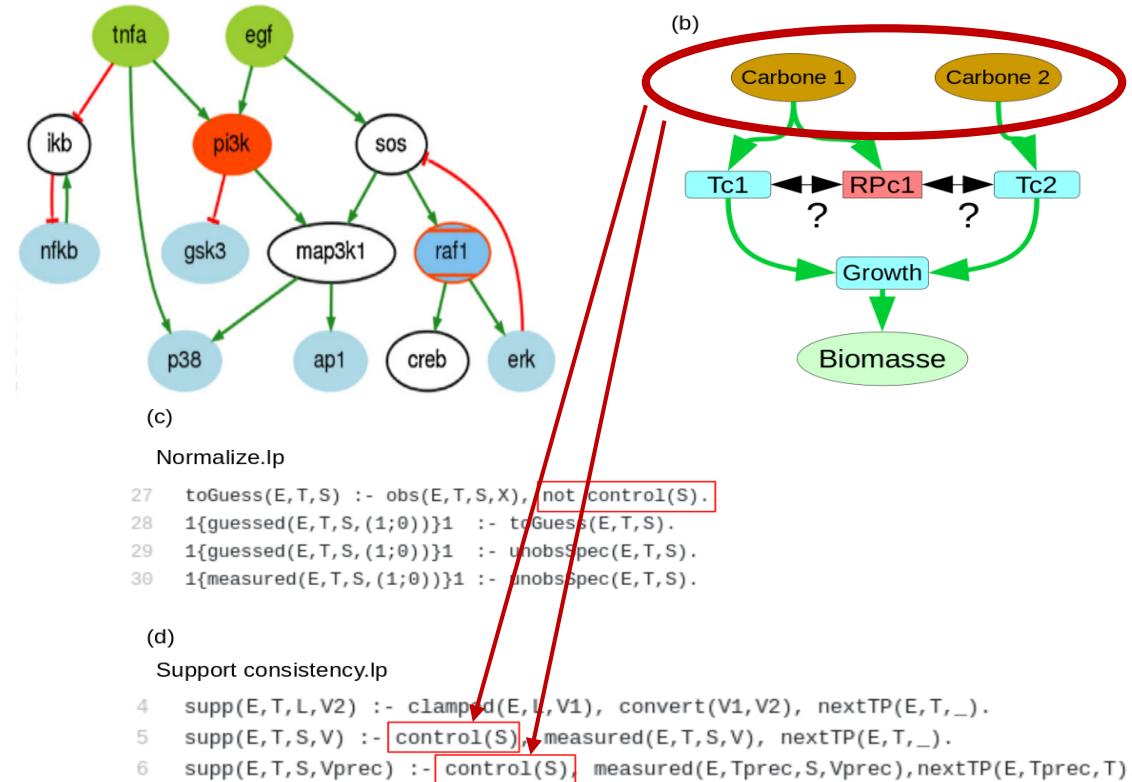
Adding a fourth class of nodes to caspots

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



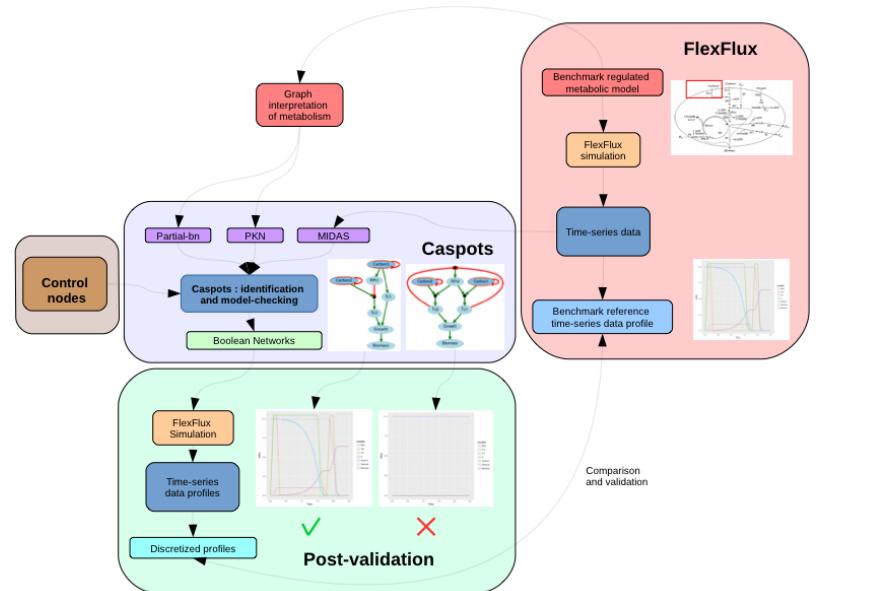
Adding a fourth class of nodes to caspots

- Nodes can be
 - Stimuli**: activated during the experimentation
 - Inhibitors**: zero value during the experimentation
 - Readout**: measured and controlled 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

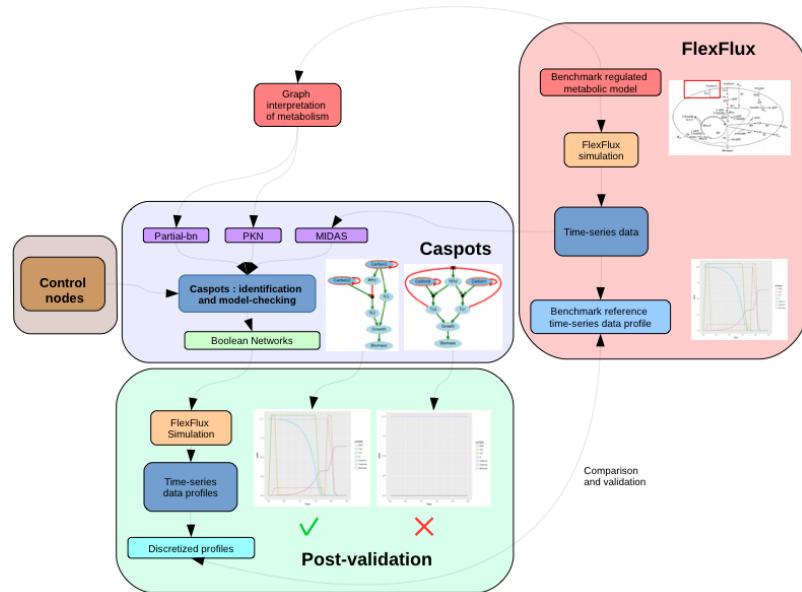
Revisiting the optimisation problem



Pipeline parameters	PKN modelling		Caspots version		Number of output networks		Running time
	Integration Graph	Interaction graph with retro-controls	Interaction graph with artificial links	Without retro-control	With retro-controls	Learning	
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

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

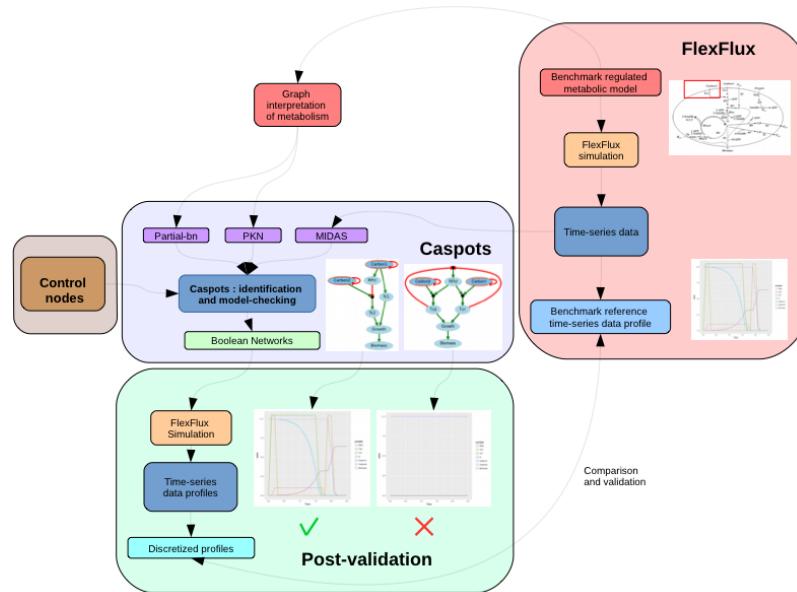
↓

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

Pipeline parameters	PKN modelling			Caspots version		Number of output networks	Comparison to benchmark dataset	Running time
	Integration Graph	Interaction graph with retro-controls	Interaction graph with artificial links	Without retro-control	With retro-controls			
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	x	x	4	1	hours

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

Revisiting the optimisation problem



Optimisation problem

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

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

Remove control nodes from the score computation

Pipeline parameters	PKN modelling			Caspots version		Number of output networks	Comparison to benchmark dataset	Running time
	Integration Graph	Interaction graph with retro-controls	Interaction graph with artificial links	Without retro-control	With retro-controls			
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

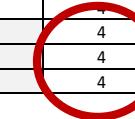
- No more need to model any artificial relation
- Gain of performance

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

Focus on the last false positives

?

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

?

Pipeline parameters	PKN modelling			Casplots version		Number of output networks	Comparison to benchmark dataset	Running time
	Integration Graph	Interaction graph with retro-controls	Interaction graph with artificial links	Without retro-control	With retro-controls			
1	x			x		40	1	seconds
2		x		x		15	1	minutes
3			x	x		+	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:RPc1	DV:RPc1	DA:Tc2	DV:Tc2	DA:Tc1	DV:Tc1	DA:Carbon1	DV:Carbon1	DA:Carbon2	DV:Carbon2	DA:Biomass	DV:Biomass
1	0	0	0	10.5	1	10.5	1	10	0	10	1	0
1	1	0	1	10.5	2	10.5	2	9.99	2	9.99	2	0.01
1	2	1	0	10.5	3	10.5	3	9.96	3	9.96	3	0.03
1	3	1	3	0	3	10.5	4	9.91	4	9.96	4	0.04
1	4	1	4	0	4	10.5	5	9.84	5	9.96	5	0.06
1	5	1	5	0	5	10.5	6	9.75	6	9.96	6	0.09
1	6	1	6	0	6	10.5	7	9.62	7	9.96	7	0.14
1	7	1	7	0	7	10.5	8	9.44	8	9.96	8	0.19
1	8	1	8	0	8	10.5	9	9.3	9	9.96	9	0.28
1	9	1	9	0	9	10.5	10	8.63	10	9.96	10	0.39
1	10	1	10	0	10	10.5	11	8.33	11	9.96	11	0.55
1	11	1	11	0	11	10.5	12	7.63	12	9.96	12	0.77
1	12	1	12	0	12	10.5	13	6.66	13	9.96	13	1.08
1	13	1	13	0	13	10.5	14	5.3	14	9.96	14	1.51
1	14	1	14	0	14	10.5	15	3.4	15	9.96	15	2.12
1	15	1	15	0	15	10.5	16	0.74	16	9.96	16	2.97
1	16	1	16	0	16	2.49	16	0.74	16	9.96	16	2.97

Remove redundancies

TR:Cell:CellLine	DA:RPc1	DV:RPc1	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 threshold

TR:Cell:CellLine	DA:RPc1	DV:RPc1	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

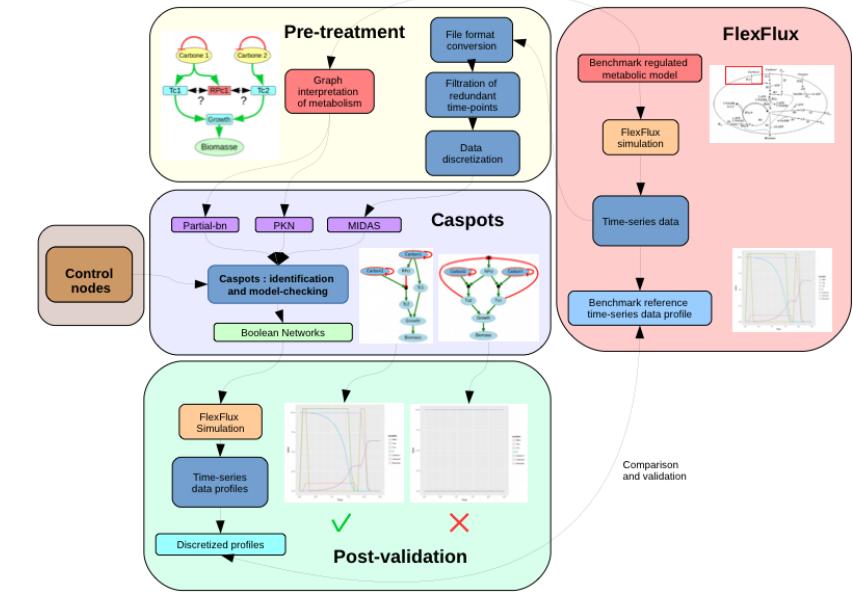
TR:Cell:CellLine	DA:RPcl	DV:RPcl	DA:Tc2	DV:Tc2	DA:Tc1	DV:Tc1	DA:Carbon1	DV:Carbon1	DA:Carbon2	DV:Carbon2	DA:Biomass	DV:Biomass
1	0	0	0	0	0	0	0	10	0	10	0	0
1	1	0	1	10.5	1	10.5	1	10	1	10	1	0
1	2	1	2	10.5	2	10.5	2	9.99	2	9.99	2	0.01
1	3	1	3	0	3	10.5	3	9.96	3	9.96	3	0.03
1	4	1	4	0	4	10.5	4	9.91	4	9.96	4	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	9.38	9	9.96	9	0.28
1	10	1	10	0	10	10.5	10	8.63	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 threshold

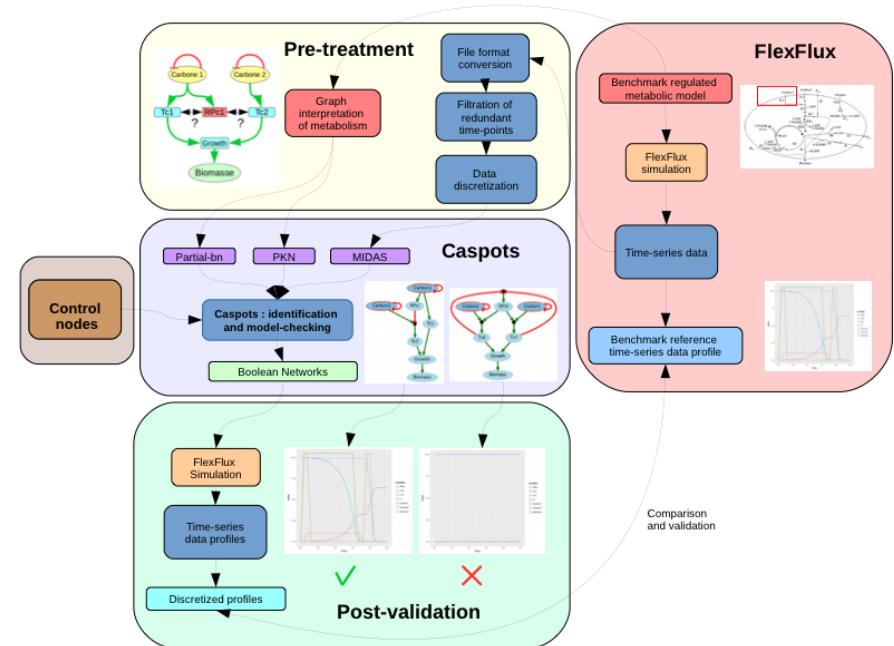
TR:Cell:CellLine	DA:RPcl	DV:RPcl	DA:Tc2	DV:Tc2	DA:Tc1	DV:Tc1	DA:Carbon1	DV:Carbon1	DA:Carbon2	DV:Carbon2	DA:Biomass	DV:Biomass
1	0	0	0	0	0	0	0	1	0	1	0	0
1	1	0	1	1	1	1	1	1	1	1	1	0
1	2	1	2	1	2	1	2	1	2	1	2	0
1	3	1	3	0	3	1	3	1	3	1	3	1
1	4	1	4	0	4	1	4	1	4	1	4	1
1	16	1	16	0	16	1	16	1	16	1	16	1



Pipeline parameters	PKN modelling		Input DataSet		Caspots version		Number of output networks	Comparison to benchmark dataset	Running time
	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		
1	x				x		x	40	1 seconds
2		x			x		x	15	1 minutes
3			x	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	x	4	1 hours
10	x				x	x	x	1	1 seconds
11		x			x	x	x	1	1 minutes
12			x		x	x	x	1	1 hours

False-positive after caspots have disappeared

Which optimisation problem ?



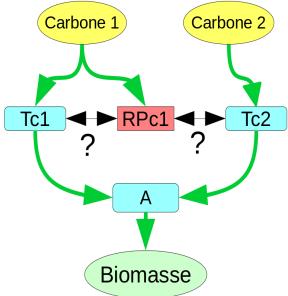
$$\arg \min_{\phi : BNs} (\underbrace{\text{scoreControl}_{rss}((V, \phi), (P_1, \dots, P_n))}_{\text{residual sum of squares}}, \underbrace{\text{score}_{size}((V, \phi))}_{\text{network size}}, \underbrace{\text{score}_{simu}(\text{met_network}, \phi)}_{\text{flux-based simulation}})$$

↓

$$\text{scoreControlFiltered}((V, \phi), (P_1, \dots, P_n)) = \sum_{t : \text{non redundant time points}} \left(\underbrace{\text{discretised(measure}(v, P, t)) -}_{\text{observations} \in \{0,1\}} \underbrace{\phi_P(v, t)}_{\text{predictions} \in \{0,1\}} \right)$$

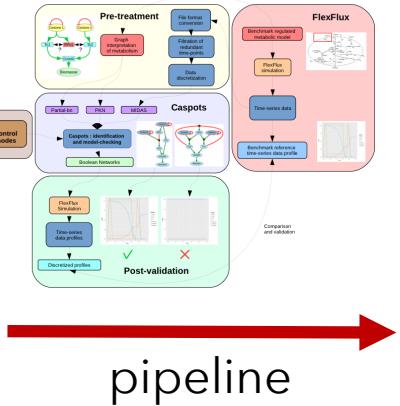
$\underbrace{t : \text{non redundant time points}}$
 $P : \text{experimental}$
 $v : \text{variable}$
 $v \neq \text{control}$

Conclusion : we did it...

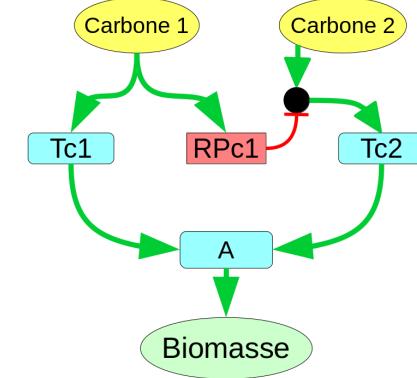


+

TR Cell	CellLine	DA:Rho	DA:RPc1	DA:Tc2	DA:Tc1	DA:Tc	DA:Carbone1	DA:Carbone2	DA:Carbone	DA:Biomasse
1	1	0	1	0	1	0.5	1	0	0	0
1	2	1	2	0.5	2	0.5	2	0.5	2	0.05
1	3	1	0	0	1	0.5	1	0.5	1	0.05
1	4	1	4	0	4	0.5	4	0.5	4	0.04
1	5	1	5	0	5	0.5	5	0.5	5	0.06
1	6	1	6	0	6	0.5	6	0.5	6	0.06
1	7	1	7	0	7	0.5	7	0.5	7	0.14
1	8	1	8	0	8	0.5	8	0.5	8	0.19
1	9	1	9	0	9	0.5	9	0.5	9	0.29
1	10	1	10	0	10	0.5	10	0.5	10	0.39
1	11	1	11	0	11	0.5	11	0.5	11	0.55
1	12	1	12	0	12	0.5	12	0.5	12	0.77
1	13	1	13	0	13	0.5	13	0.5	13	1.06
1	14	1	14	0	14	0.5	14	0.5	14	1.35
1	15	1	15	0	15	0.5	15	0.5	15	1.73
1	16	1	16	0	16	2.49	16	0.5	16	2.97
1	17	1	17	0	17	0.5	17	0.5	17	3.27
1	18	0	18	0	18	0	18	0	18	3.22
1	19	0	19	0.5	19	0	19	0	19	3.22
1	20	0	20	0	20	0	20	0	20	3.22
1	21	0	21	0.49	21	0	21	0	21	3.23
1	22	0	22	0	22	0	22	0	22	3.39

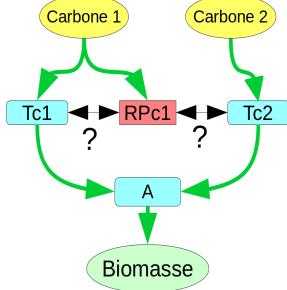


pipeline



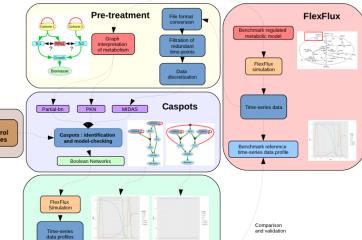
Uniquely recover the rule of a very simple regulated metabolic network

Conclusion : we did it...

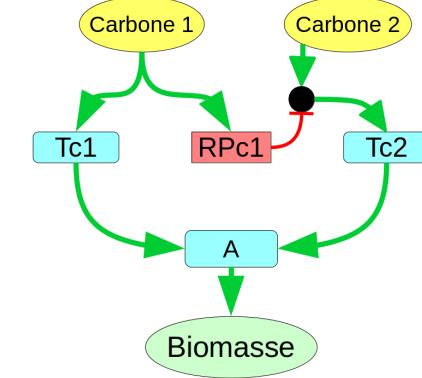


+

TR Cell CellLine	DA-Rpc1	DA-Rpc2	DA-Tc1	DA-Tc2	DA-Carbon1	DA-Carbon2	DA-Biomass	DA-Samples
1	1	0	1	0	3	10	1	0
1	2	1	2	0.5	2	10.5	2	9.99
1	3	1	3	0.5	3	10.5	3	9.99
1	4	1	4	0	4	10.5	4	9.99
1	5	1	5	0	5	10.5	5	9.99
1	6	1	6	0	6	10.5	6	9.99
1	7	1	7	0	7	10.5	7	9.99
1	8	1	8	0	8	10.5	8	9.99
1	9	1	9	0	9	10.5	9	9.99
1	10	1	10	0	10	10.5	10	9.99
1	11	1	11	0	11	10.5	11	9.99
1	12	1	12	0	12	10.5	12	9.99
1	13	1	13	0	13	10.5	13	9.99
1	14	1	14	0	14	10.5	14	9.99
1	15	1	15	0	15	10.5	15	9.99
1	16	1	16	0	16	10.5	16	9.99
1	17	1	17	0	17	10.5	17	9.99
1	18	0	18	0	18	0	18	9.99
1	19	0	19	0.5	19	0	19	9.99
1	20	0	20	0	20	0	20	9.99
1	21	0	21	0.49	21	0	21	0.31
1	22	0	22	0	22	0	22	0.29

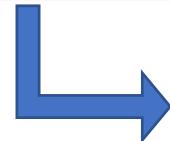


pipeline



Uniquely recover the rule of a very simple regulated metabolic network

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



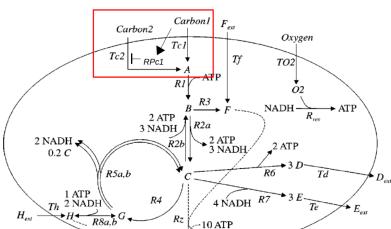
$$\arg \min_{\phi : BNs} \underbrace{\text{scoreControlFiltered}((V, \phi), (P_1, \dots, P_n))}_{\text{distance to pretreated data}}, \underbrace{\text{score}_{size}((V, \phi))}_{\text{network size}}, \underbrace{\text{score}_{simu}(met_network, \phi)}_{\text{flux-based simulation}}$$

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

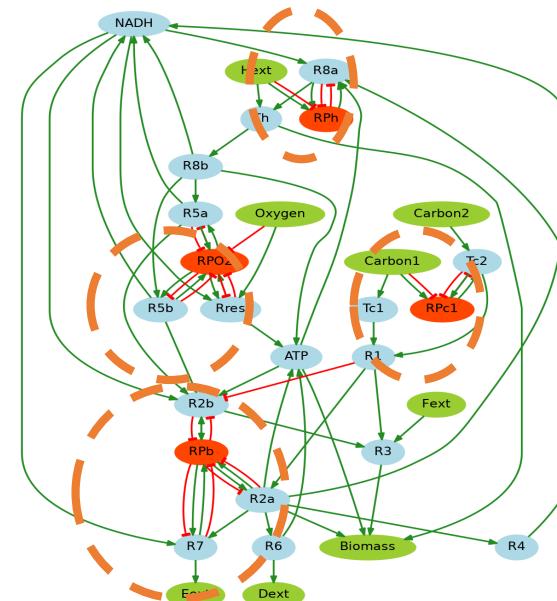
*P : experimentations
v : variable
v ≠ control*

Complete reformulation
of the optimisation problem.

What happens for a larger example ?



Extended
interaction
graph



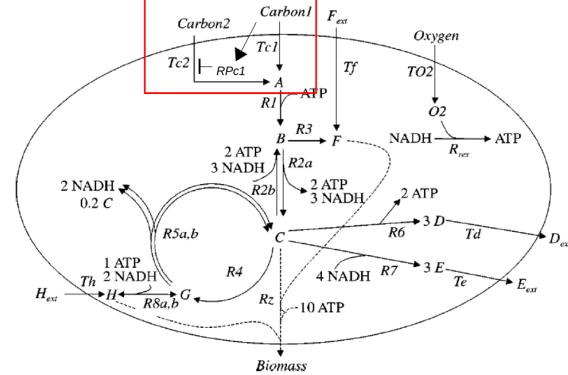
pipeline

+ data

TR Cell CellLine	DA.RPc1	DV.RPc1	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	10	0	10	0	1	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	9.49	8	9.96	8	0.19
1	9	1	9	0	9	10.5	9	9.19	9	9.96	9	0.23
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

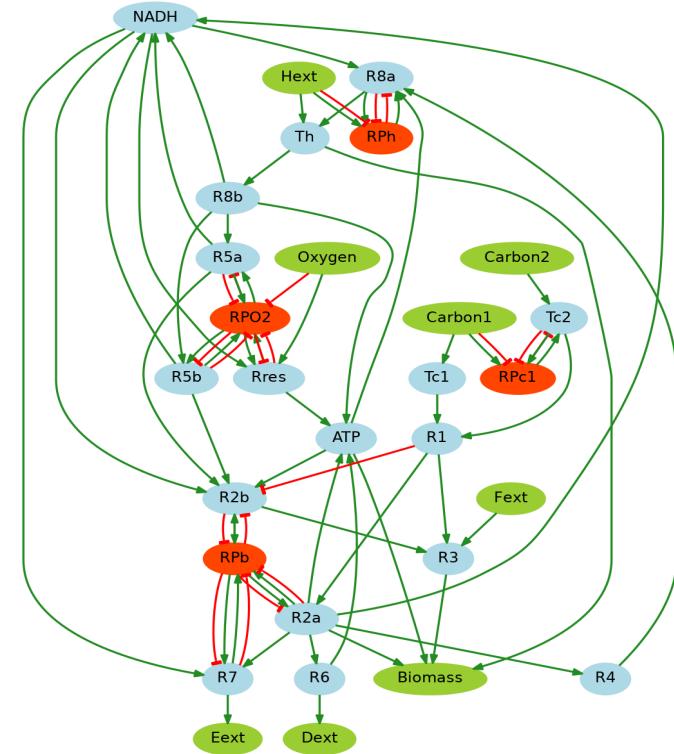
Uniquely recover the expected regulations

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 stoichiometry constraints
- (...)

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