The TimoTimo: a revolutionary TomTom for analyzing metabolic highways

TimoTimo

BioSB Conference

20 May 2015, NL

Timo R. Maarleveld
Centrum Wiskunde & Informatica and VU University Amsterdam

TOMTOM®

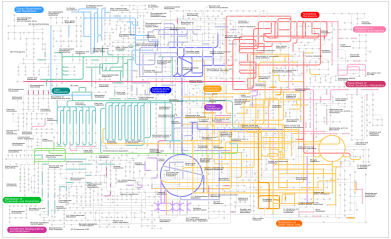
GSSMs

Tip of the Iceberg

FBA/CoPE FBA Primer

TimoTimo

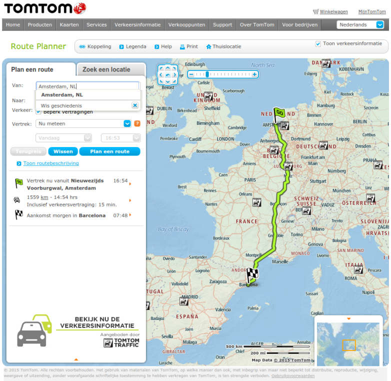
A Real Life Example: Synechocystis

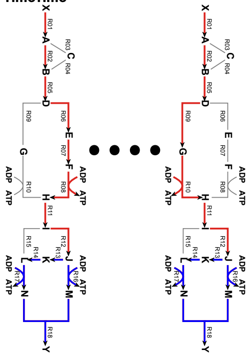


TOMTOM®

TimoTimo

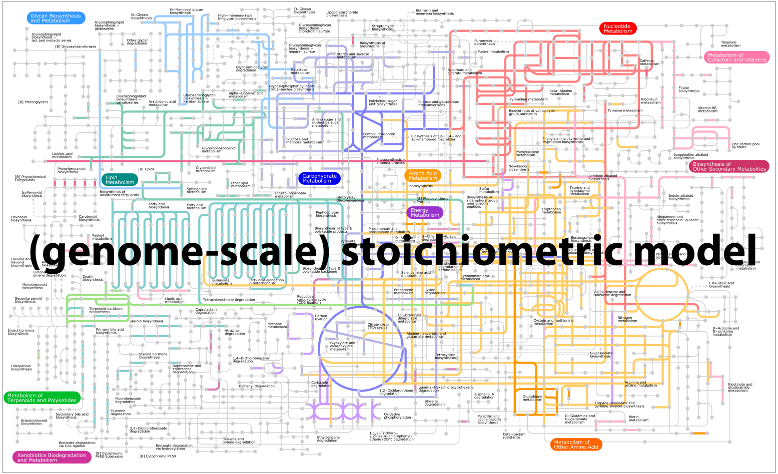
TomTom and TimoTimo





- TomTom gives you an optimal route from starting point to destination
- TimoTimo gives you **all** optimal routes from substrate(s) to product(s)

TimoTimo's input



(genome-scale) stoichiometric model

Genome-scale stoichiometric models in a nutshell

1. **metabolism**: the set of life-sustaining chemical transformations within the cells of living organisms

Genome-scale stoichiometric models in a nutshell

1. **metabolism**: the set of life-sustaining chemical transformations within the cells of living organisms
2. **genome-scale**: cover the total metabolic potential that is encoded in the genome of an organism

Genome-scale stoichiometric models in a nutshell

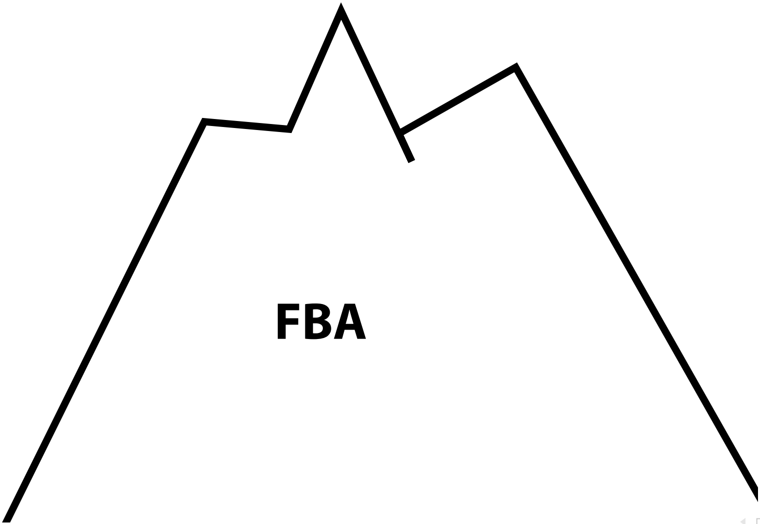
1. **metabolism**: the set of life-sustaining chemical transformations within the cells of living organisms
2. **genome-scale**: cover the total metabolic potential that is encoded in the genome of an organism
3. **stoichiometric**:
   - only relative quantities: $A + B \rightarrow C$
   - no kinetics: ~~$k_1 \times A[t] \times B[t]$~~
   - assume steady-state: $\frac{dA}{dt} = \frac{dB}{dt} = \frac{dC}{dt} = 0$

Genome-scale stoichiometric models in a nutshell

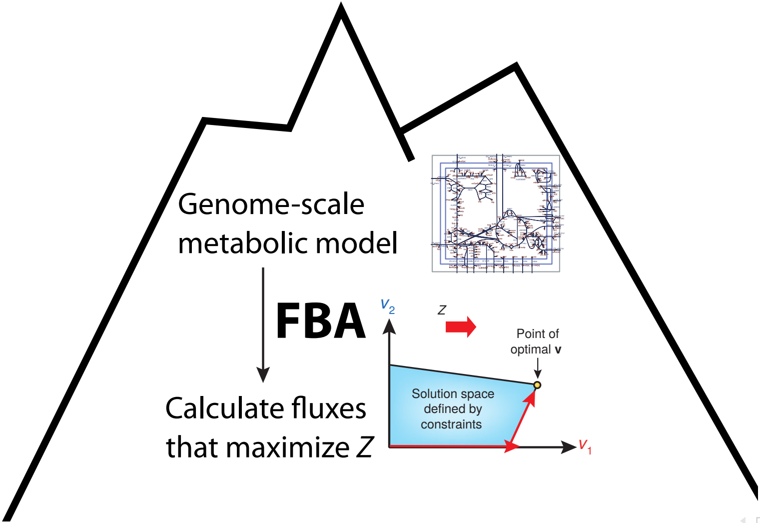
1. **metabolism**: the set of life-sustaining chemical transformations within the cells of living organisms
2. **genome-scale**: cover the total metabolic potential that is encoded in the genome of an organism
3. **stoichiometric**:
   - only relative quantities: $A + B \rightarrow C$
   - no kinetics: ~~$k_1 \times A[t] \times B[t]$~~
   - assume steady-state: $\frac{dA}{dt} = \frac{dB}{dt} = \frac{dC}{dt} = 0$
4. predict **steady-state flux distributions**

How can we simulate these models?

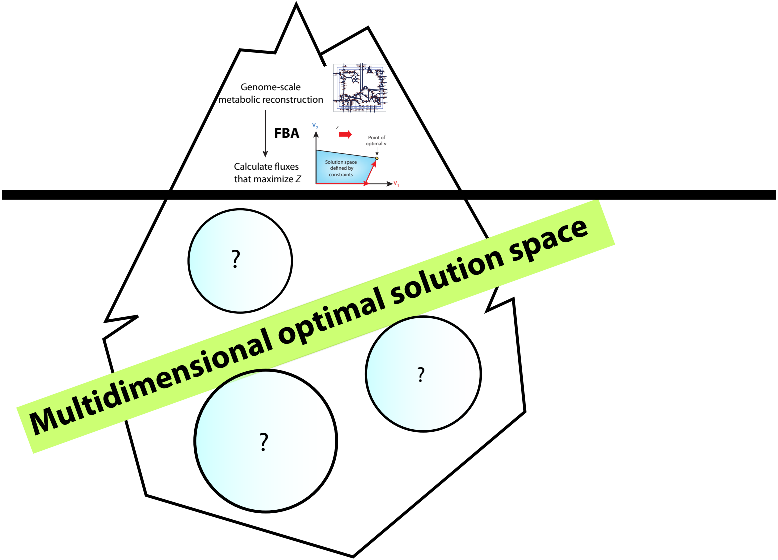
Flux **B**alance **A**nalysis

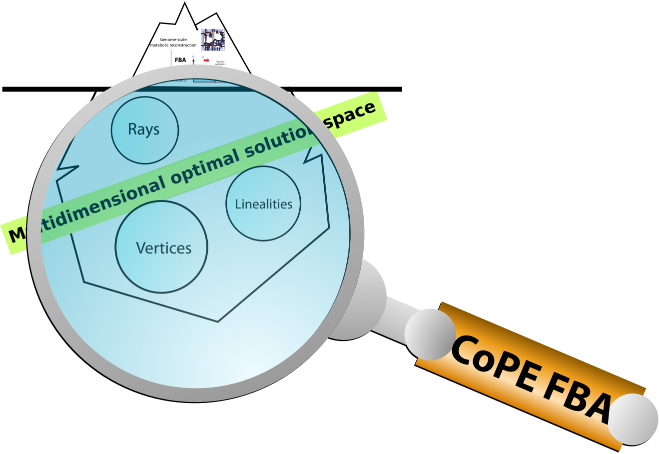


Flux Balance Analysis



Tip of the iceberg

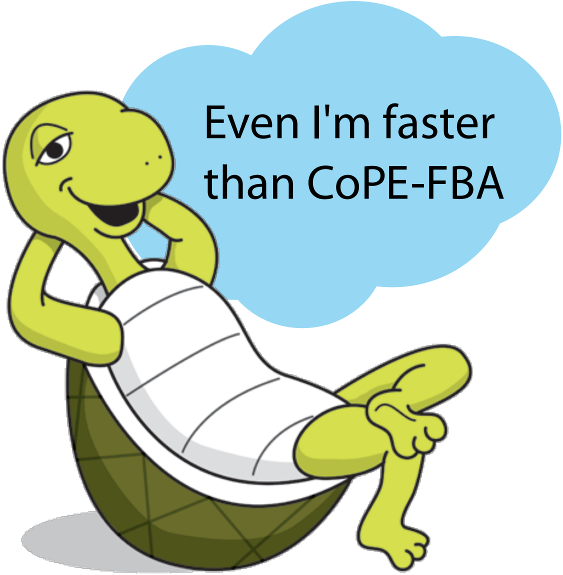




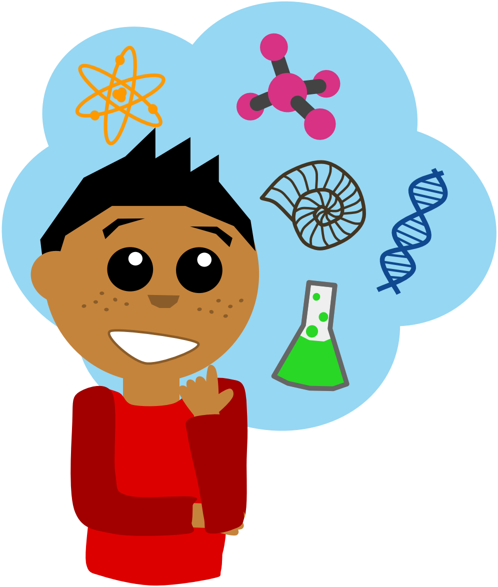
Topological understanding of the metabolic capacity in terms of metabolic flux routes

Kelk et al, 2012. Scientific Reports.

CoPE-FBA is slow!

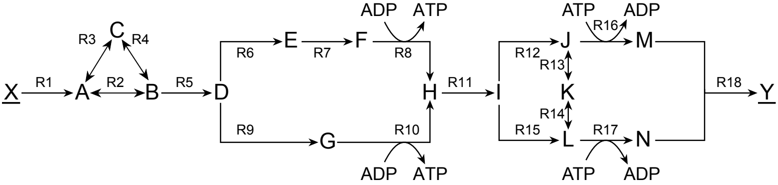


Can we speed-up CoPE-FBA?



Our solution: CoPE-FBA 2.0 a.k.a. TimoTimo

Toy model for stoichiometric modeling



- both X (source) and Y (sink) are fixed
- co-factors ATP and ADP

Flux **B**alance **A**nalysis formulation

Linear Program:

Maximize (linear) obj. func.

subject to,

- steady-state
- flux constraints

Flux **B**alance **A**nalysis formulation

Linear Program:

Maximize $Z_{obj} = \mathbf{c}^T \mathbf{J}$

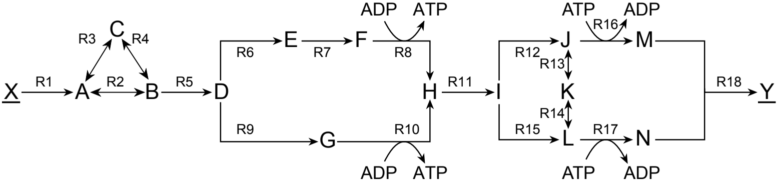
subject to,

$\boldsymbol{NJ} = \mathbf{0}$

$\boldsymbol{J^{min}} \leq \boldsymbol{J} \leq \boldsymbol{J^{max}}$

N: stoichiometric matrix
J: steady-state flux vector
c: vector of coefficients that represent the contribution of each flux in vector ***J*** to the objective function Z_{obj}.

Flux **B**alance **A**nalysis formulation, example



Maximize $Z_{obj} = \mathbf{c}^T \mathbf{J} = J_{18}$

subject to,

$$\boldsymbol{NJ} = \mathbf{0}$$

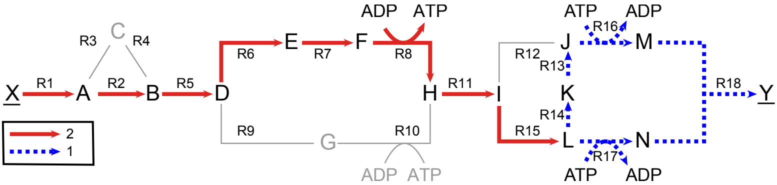
$$-\infty \leq J_r \leq \infty \qquad J_r \in \text{reversible reactions}$$

$$0 \leq J_i \leq \infty \qquad J_i \in \text{irreversible reactions}$$

$$0 \leq J_1 \leq 2$$

FBA example

Maximize $Z_{obj} \rightarrow J_{18} = 1$



with $\boldsymbol{J} = [2, 2, 0, 0, 2, 2, 2, 2, 0, 0, 2, 0, -1, -1, 2, 1, 1, 1]$

But this is only one of the possibilities ...

Optimal solution space characterization

Definition: $F_{opt} = \{\mathbf{J} : \mathbf{NJ} = \mathbf{0}, \boldsymbol{J}^{min} \leq \boldsymbol{J} \leq \boldsymbol{J}^{max}, \boldsymbol{c}^T \boldsymbol{J} = opt\}$

1. **vertices** --- optimal flux vectors
   - corner points of the optimal solution space
   - non-decomposable
2. **rays** --- irreversible cycles (or input-output pathways)
3. **linealities** --- reversible cycles (or input-output pathways)

Optimal solution space characterization

Definition: $F_{opt} = \{\mathbf{J} : \mathbf{NJ} = \mathbf{0}, \boldsymbol{J^{min}} \leq \boldsymbol{J} \leq \boldsymbol{J^{max}}, \boldsymbol{c^T J} = opt\}$

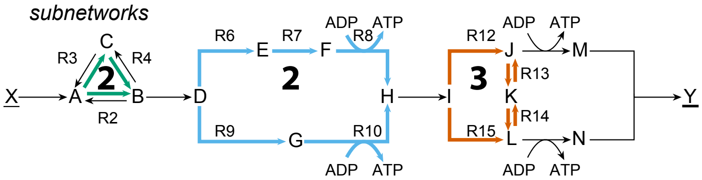
1. **vertices** --- optimal flux vectors
   - corner points of the optimal solution space
   - non-decomposable
2. **rays** --- irreversible cycles (or input-output pathways)
3. **linealities** --- reversible cycles (or input-output pathways)
4. **subnetworks** --- sets of variable correlated reactions
   - $\mathbf{N_A J_A} = \mathbf{d} \rightarrow$ fixed input-output relationship

Optimal solution space characterization

Definition: $F_{opt} = \{\mathbf{J} : \mathbf{NJ} = \mathbf{0}, \boldsymbol{J}^{min} \leq \boldsymbol{J} \leq \boldsymbol{J}^{max}, \boldsymbol{c}^{T}\boldsymbol{J} = opt\}$

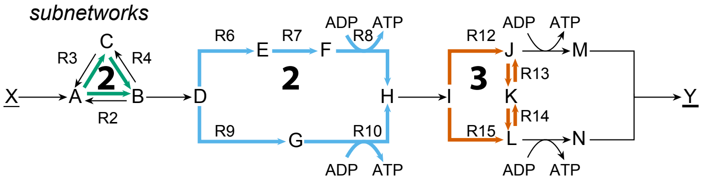
1. **vertices** --- optimal flux vectors
   - corner points of the optimal solution space
   - non-decomposable
   - no convex combination of other optimal flux vectors
2. rays --- irreversible cycles (or input-output pathways)
3. linealities --- reversible cycles (or input-output pathways)
4. **subnetworks** --- sets of variable correlated reactions
   - $\mathbf{N_A J_A} = d \rightarrow$ fixed input-output relationship

Subnetworks explain # vertices



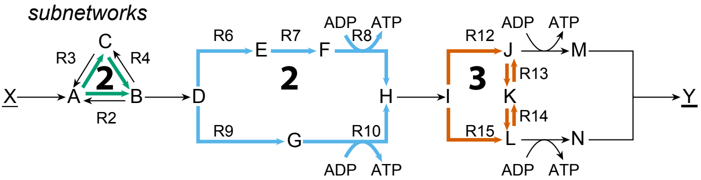
- $\mathbf{N_A J_A} = \mathbf{d} \neq \mathbf{0}$ i.e. input-output relationship
  1. $A \rightarrow B$
  2. $D + ADP \rightarrow H + ATP$
  3. $I \rightarrow 0.5(J + L)$

Subnetworks explain # vertices



- $\mathbf{N_A J_A} = \mathbf{d} \neq \mathbf{0}$ i.e. input-output relationship
  1. $A \rightarrow B$
  2. $D + ADP \rightarrow H + ATP$
  3. $I \rightarrow 0.5(J + L)$
- Subnetworks have 2, 2, and 3 vertices
  1. $\{R2^+\}$ & $\{R3^+, R4^+\}$
  2. $\{R6, R7, R8\}$ & $\{R9, R10\}$
  3. $\{R12, R13^+, R14^+\}$, $\{R15, R13^-, R14^-\}$ & $\{R12, R15\}$

Subnetworks explain # vertices



- $\mathbf{N_A J_A} = \mathbf{d} \neq \mathbf{0}$ i.e. input-output relationship
  1. $A \to B$
  2. $D + ADP \to H + ATP$
  3. $I \to 0.5(J + L)$
- Subnetworks have 2, 2, and 3 vertices
  1. $\{R2^+\}$ & $\{R3^+, R4^+\}$
  2. $\{R6, R7, R8\}$ & $\{R9, R10\}$
  3. $\{R12, R13^+, R14^+\}$, $\{R15, R13^-, R14^-\}$ & $\{R12, R15\}$
- Each subnetwork is an independent module with a fixed $\mathbf{d}$: $2 \times 2 \times 3 = 12$ network vertices

Maarleveld et al, 2015. PLoS Comp. Biol.

Reversible-reaction splitting

Reversible-reaction splitting yields a unique characterization of the optimal solution space, i.e. all non-decomposable **J** in the optimum.

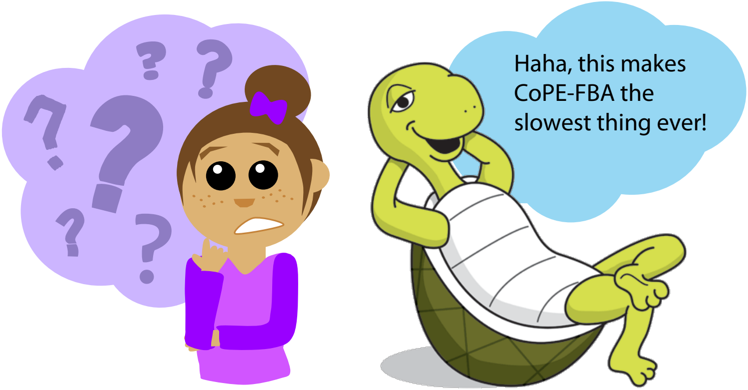
Organism (model)	Source	# vertices	
		without splitting	with splitting
S. PC6803 (iTM686)	glycogen	4.423.680	1.399.652.352
S. PC6803 (iTM686)	light	368.640	41.803.776
E. coli (iAF1260,ox)	glucose	839.808	120.932.352
E. coli (iAF1260,noox)	glucose	31104	1.492.992
L. Lactis	glucose	192	1.218.240
S. thermophilus	lactose	96	280
M. tuberculosis (iNJ661)	glycerol	1.327.104	7.05×10^{11}
M. barkeri (iAF692)	methanol	512	104.832
E. coli (iJR904)	malate	320	6912
E. coli (iJR904)	fumarate	640	7680

CoPE-FBA was designed to enumerate the optimal solution space **without** reversible-reaction splitting

Maarleveld et al, 2015. PLoS Comp. Biol.

Enumerating the optimal solution space with reversible-reaction splitting ...

These simulations are going to take forever right??



That's why we developed the TimoTimo

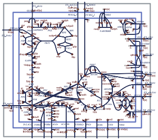
How fast is TimoTimo?

Organism (model)	Source	Running time (s) CoPE-FBA Kelk et al. 2012	Running time (s) TimoTimo
S. PC6803 (iTM686)	glycogen	FAILED	117
S. PC6803 (iTM686)	light	FAILED	26
E. coli (iAF1260,ox)	glucose	FAILED	36
E. coli (iAF1260,noox)	glucose	FAILED	27
L. Lactis	glucose	432495	16
S. thermophilus	lactose	345362	15
M. tuberculosis (iNJ661)	glycerol	FAILED	41
M. barkeri (iAF692)	methanol	265760	12
E. coli (iJR904)	malate	324793	10
E. coli (iJR904)	fumarate	584690	16

"Faster than lightning", but why?

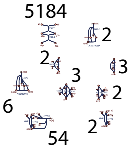
CoPE-FBA approach (in a nutshell)

iAF1260 (glc, aerobic)



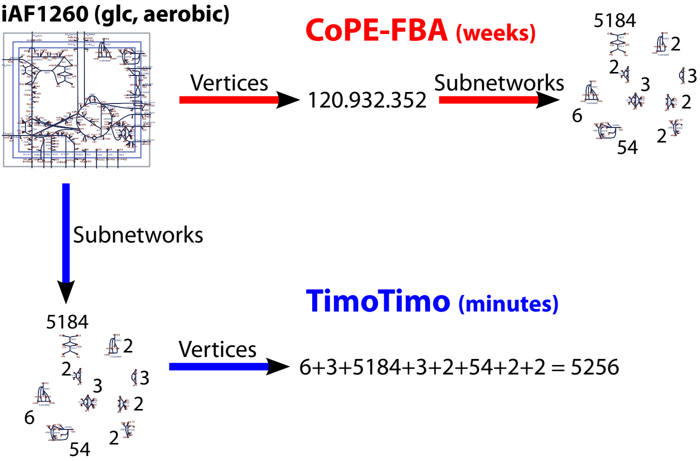
CoPE-FBA (weeks)

Vertices → 120.932.352 Subnetworks →



Kelk et al, 2012. Scientific Reports

TimoTimo uses a divide-and-conquer approach



Enumerate 0.004% of the vertices to get all

Maarleveld et al, 2015. PLoS Comp. Biol.

Occasionally ...

there exists 1 subnetwork that captures (nearly) all variability. Then, also TimoTimo is "slow".

Organism (model)	Source	Running time (s) CoPE-FBA Kelk et al. 2012	Running time (s) TimoTimo
S. PC6803 (iTM686)	glycogen	FAILED	117
S. PC6803 (iTM686)	light	FAILED	26
E. coli (iAF1260)	glucose	FAILED	147600
E. coli (iAF1260,ox)	glucose	FAILED	36
E. coli (iAF1260,noox)	glucose	FAILED	27
L. Lactis	glucose	432000+	16
S. thermophilus	lactose	345362	15
M. tuberculosis (iNJ661)	glycero	FAILED	41
M. barkeri (iAF692)	methanol	265760	12
E. coli (iJR904)	malate	324793	10
E. coli (iJR904)	fumarate	584690	16

E.coli iAF1260 growing on glucose in an aerobic restricted environment is such an example.

A real life example: cyanobacterium *Synechocystis*

CoPE-FBA

T.R. Maarleveld

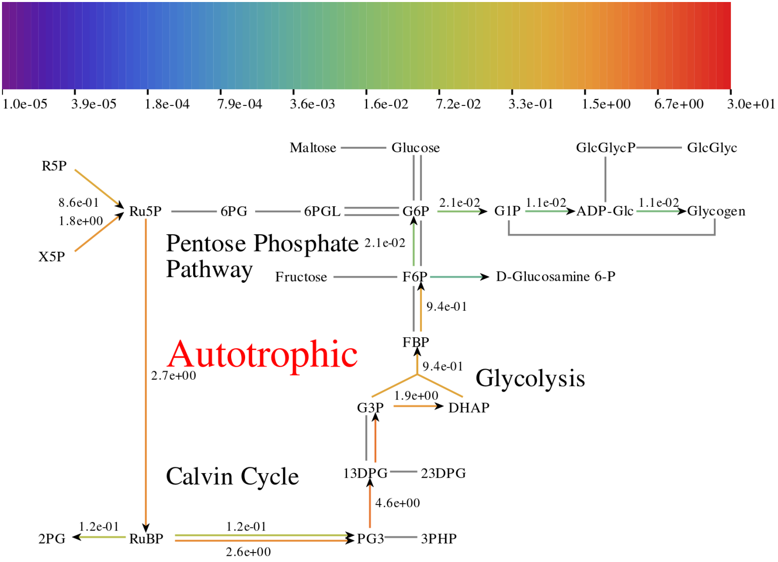
GSSMs

Tip of the Iceberg

FBA/CoPE FBA Primer

TimoTimo

A Real Life Example: Synechocystis



Maarleveld et al, 2014. Plant Physiology

Optimal solution space characterization
after reversible-reaction splitting

	Synechocystis iTM686	
	autotrophic (LLS)	heterotrophic (glycogen)
Simulation time (s)	135	228
Rays	242	242
Linealities	0	0
Vertices	**41.803.776**	**1.399.652.352**
Subnetworks	11	11

What's the effect of secondary objectives?

$$P_L = |\{J_i : J_i \neq 0\}| \quad (1)$$

$$P_J = \sum_{i=1}^{r} |J_i| \quad (2)$$

What's the effect of secondary objectives?

$$P_L = |\{J_i : J_i \neq 0\}| \tag{1}$$

$$P_J = \sum_{i=1}^{r} |J_i| \tag{2}$$

	Synechocystis iTM686 growth condition	
	autotrophic (LLS)	heterotrophic (glycogen)
Vertices	41.803.776	1.399.652.352
min(P_L)	184.320	460.800
min(P_J)	24	4

What's the effect of secondary objectives?

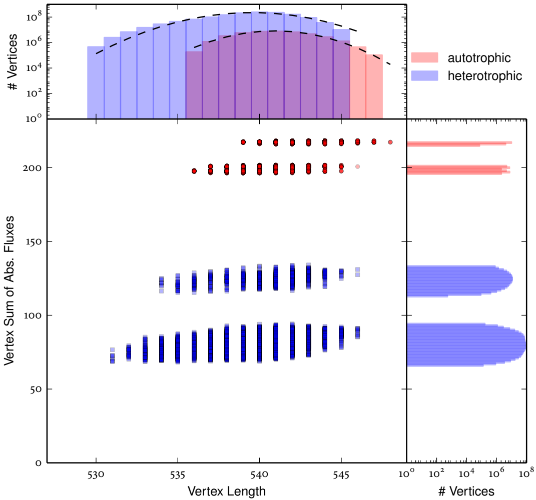
$$P_L = |\{J_i : J_i \neq 0\}| \quad (1)$$

$$P_J = \sum_{i=1}^{r} |J_i| \quad (2)$$

	Synechocystis iTM686 growth condition	
	autotrophic (LLS)	heterotrophic (glycogen)
Vertices	41.803.776	1.399.652.352
min(P_L)	184.320	460.800
min(P_J)	24	4

What kind of distributions do you expect?

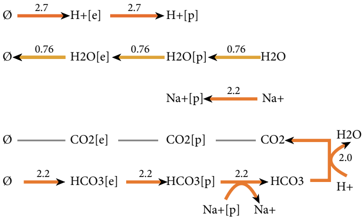
Gaussian and bimodal distributions of secondary optimization of vertices



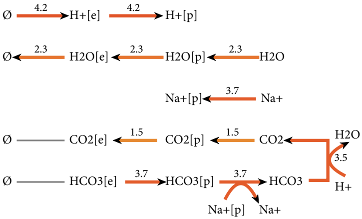
What explains the bimodal distributions?

autotrophic growth

min P_j



max P_j



- strategy: minimal HCO_3^- vs. maximal HCO_3^- import
- relative effect: 206 %
- total effect: 11 %

Summary

- we present a **faster** and more **user-friendly** tool to enumerate the optimal solution space

- Topological understanding of the metabolic capacity
- Investigate flexibility

- further reduction can be done by using secondary objectives
- unexpected bimodal distribution (also found for other organisms)

Acknowledgments

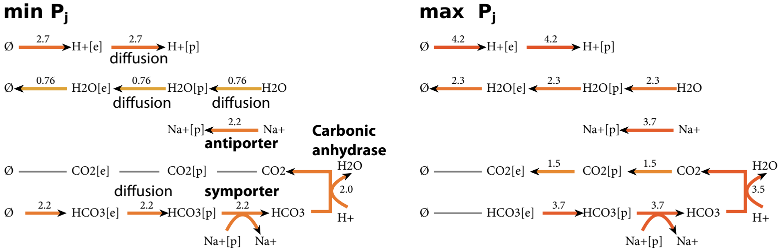
- Meike Wortel , Brett Olivier, Bas Teusink, Frank Bruggeman (VU)
- Arne Reimers (CWI)

- Maarleveld et al. PLoS Comp. Biol. 2015 - Interplay between Constraints, Objectives, and Optimality for Genome-Scale Stoichiometric Models
- Maarleveld et al. Plant Physiology 2014 - A Data Integration and Visualization Resource for the Metabolic Network of *Synechocystis* sp. PCC 6803

- Contact information: t.r.maarleveld[at]cwi.nl
- Website: `memesa-tools.sf.net`

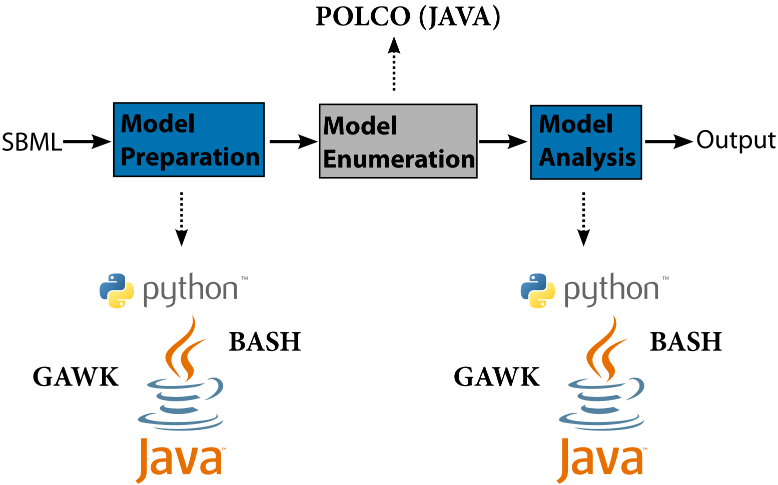
What explains the bimodal distributions?

autotrophic growth

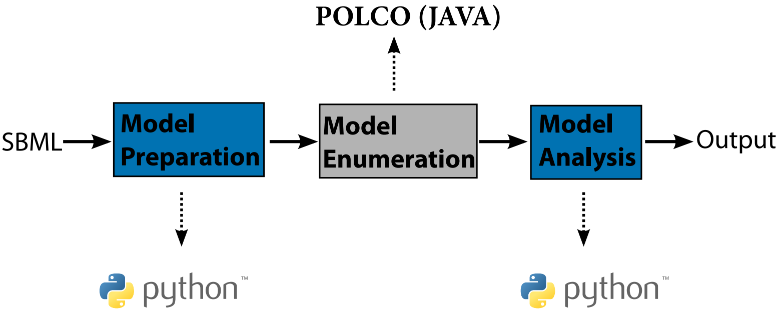


- strategy: minimal HCO_3^- vs. maximal HCO_3^- import
- some reactions are related to diffusion
- requires additional expression of transporters and enzymes

Old implementation



New implementation



- (more) user-friendly
- cross platform