Studying Metabolic Networks Through Monte Carlo Simulations

Pablo Teixeira Héctor Cancela Luis Acerenza

Universidad de la República
Montevideo, Uruguay

MCQMC, Rennes, 2018

Talk Outline

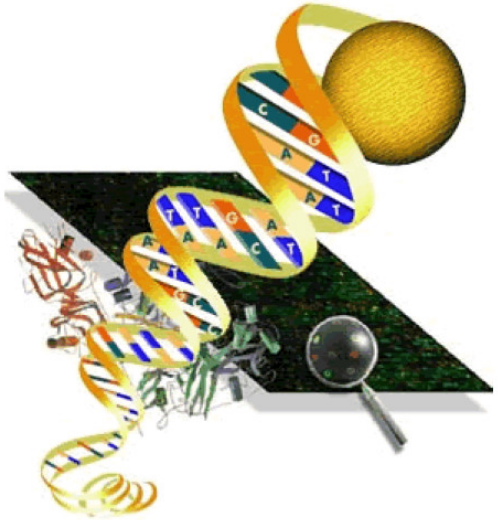
1. Background
2. Variability computation
3. Experiments
4. Conclusions and future work

Background

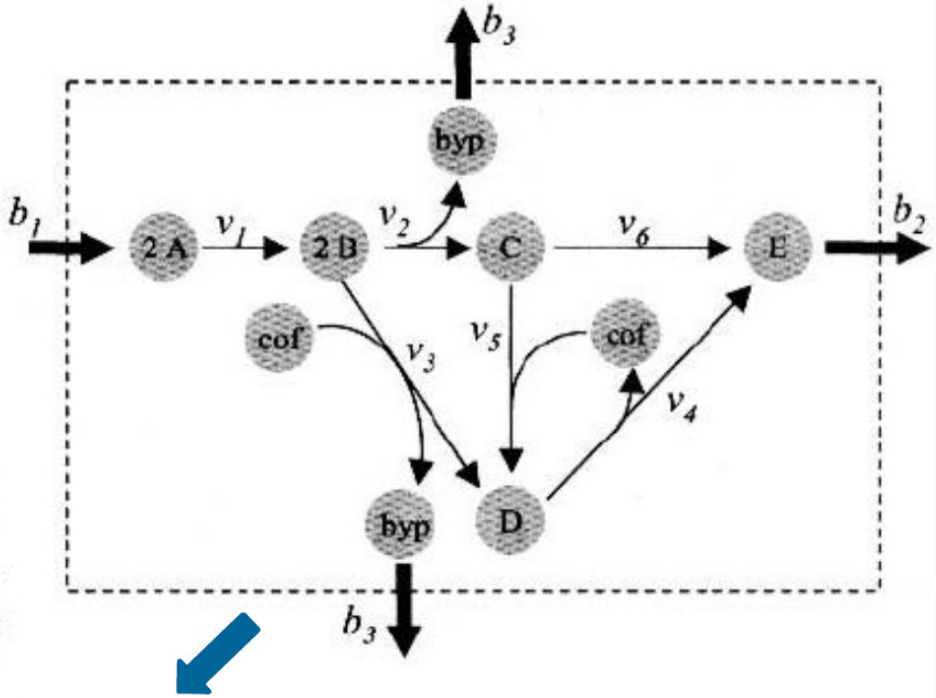
- Systems biology: "the computational and mathematical modeling of complex biological systems".
- Metabolic network: complete set of metabolic and physical processes that determine behavior of a cell.
- Genome sequencing allows to reconstruct full metabolic networks for organisms (from bacteria to humans).
- Reconstructed metabolic networks used to study different properties and behaviors.

Metabolic reconstruction

Genomic sequencing



Metabolic network



$$S = \begin{array}{c} \begin{matrix} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & b_1 & b_2 & b_3 \end{matrix} \\ \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 \\ +1 & -2 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & +1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & +1 & 0 & +1 & 0 & -1 & 0 \\ 0 & +1 & +1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & +1 & -1 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{matrix} A \\ B \\ C \\ D \\ E \\ byp \\ cof \end{matrix} \end{array}$$

Stoichiometric matrix

Metabolic networks analysis

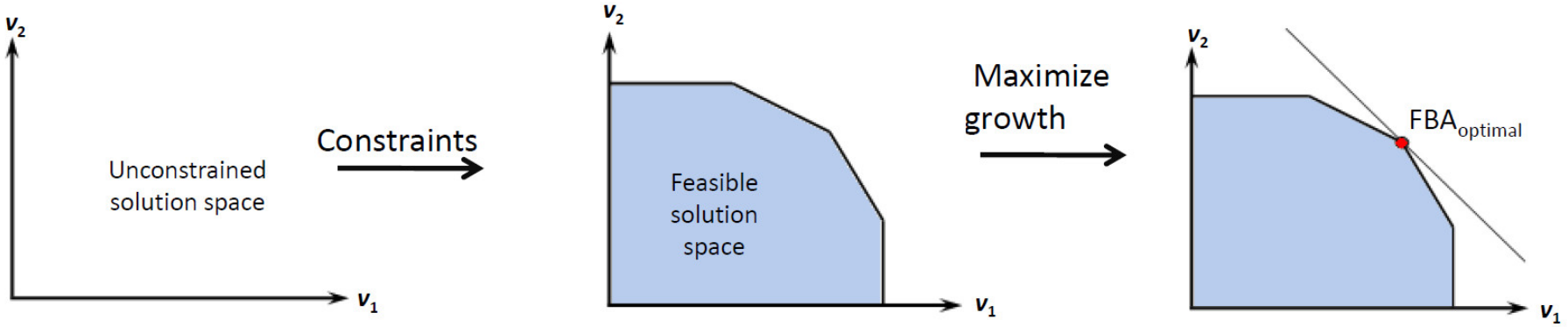
- Metabolic network $G = (M, R)$
- $M = 1..m$: set of nodes= metabolites
- $R = 1..r$: set of arcs = reactions (regulated by enzymes)
- S matrix $m \times r$, where each column corresponds to the stoichometric coefficients of the related metabolites.
- If v_i is the flux (reaction speed) for each $i \in R$, and $x_j(t)$ is the concentration for each metabolite $j \in M$, then this set of first order differential equations models the dynamic behavior of the cell:

$$\frac{d\vec{x}}{dt} = S.\vec{v}$$

Metabolic networks analysis (cont.)

- When the cell is in steady state, then system becomes $S.\vec{v} = 0$ and solutions $\vec{v}$ represent potential equilibrium configurations.
- There are more reactions than metabolites, this is usually a compatible, underdetermined linear system (usually not full rank), with infinite solutions.
- In any given solution, some reactions will be "turned off" (i.e., $v_i = 0$); other ones will be "turned on", at speed $v_i > 0$.
- Usually some additional constraints on reaction speeds are considered, $a_i \leq v_i \leq b_i$ for $i \in R$.
- Additionally, from all feasible solutions, we may select some by adding additional criteria (for instance, maximizing cell growth).

Flux Balance Analysis

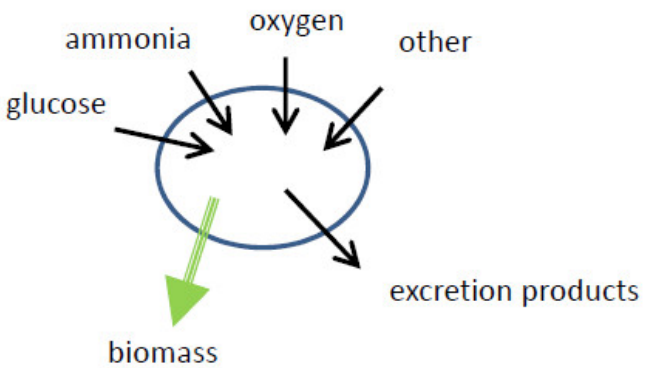


$$\begin{aligned} &maximize && v_{biomass} \\ &s.t. && Nv=0 \\ & && a_i \leq v_i \leq b_i \quad \text{for all } i \end{aligned}$$

Metabolic variability

- The different metabolic configurations play a role in resilience to changes in environment (different nutrients available, etc.).
- Larger set of feasible configurations can have an evolutionary advantage.
- How to measure this variability? What are its sources?
- Literature: Flux Variability Analysis - maximize/minimize each reaction flux, subject to the usual constraints.
- MSc. thesis by Magdalena San Roman, co-oriented by Acerenza and Cancela, analysis for FVA and other alternatives, studying internal/external/growth variability.
- Shortcomings: variability measure is a proxy, does not exactly reflect the size of the metabolic configurations space.
- This work: explore computation of size (volume) of this space.

Variability - example of prior work



$$\Delta^{tot} = \Delta^{int} + \Delta^{ext} + \Delta^{gro}$$

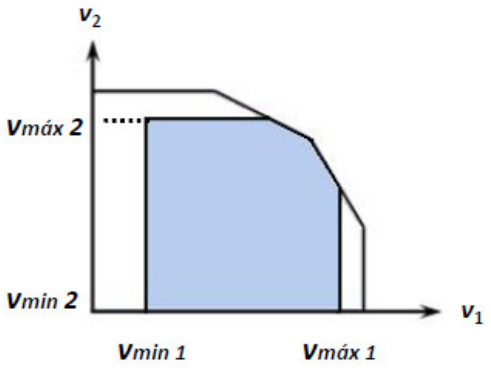
Jgro free, Jext free

$$\Delta^{int} + \Delta^{ext}$$

Jgro fixed, Jext free

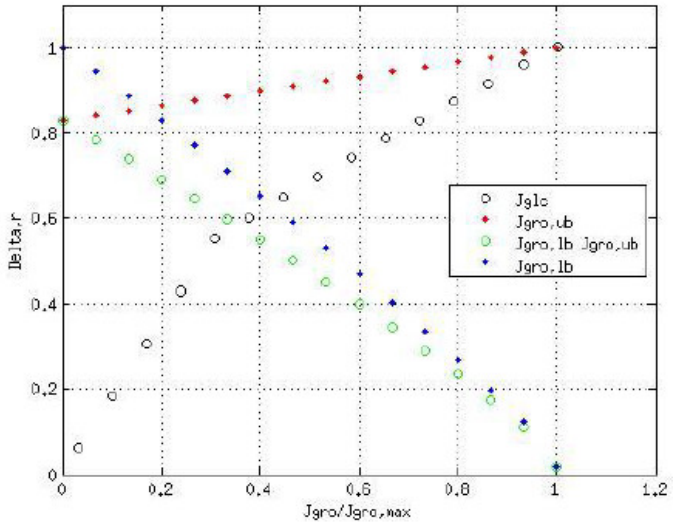
$$\Delta^{int}$$

Jgro fixed, Jext fixed



$$\Delta_i = \frac{v_{max,i} - v_{min,i}}{V_{max,i} - V_{min,i}}$$

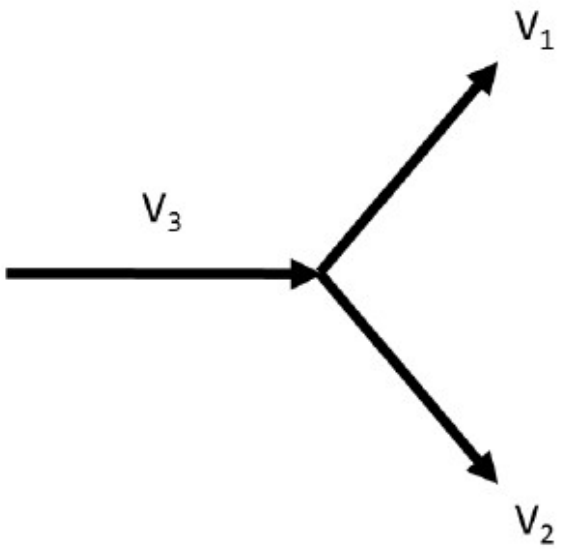
$$\Delta_r = \left(\sum_{i \in R} \Delta_i \right) / r$$



Simple example - volume vs. FVA

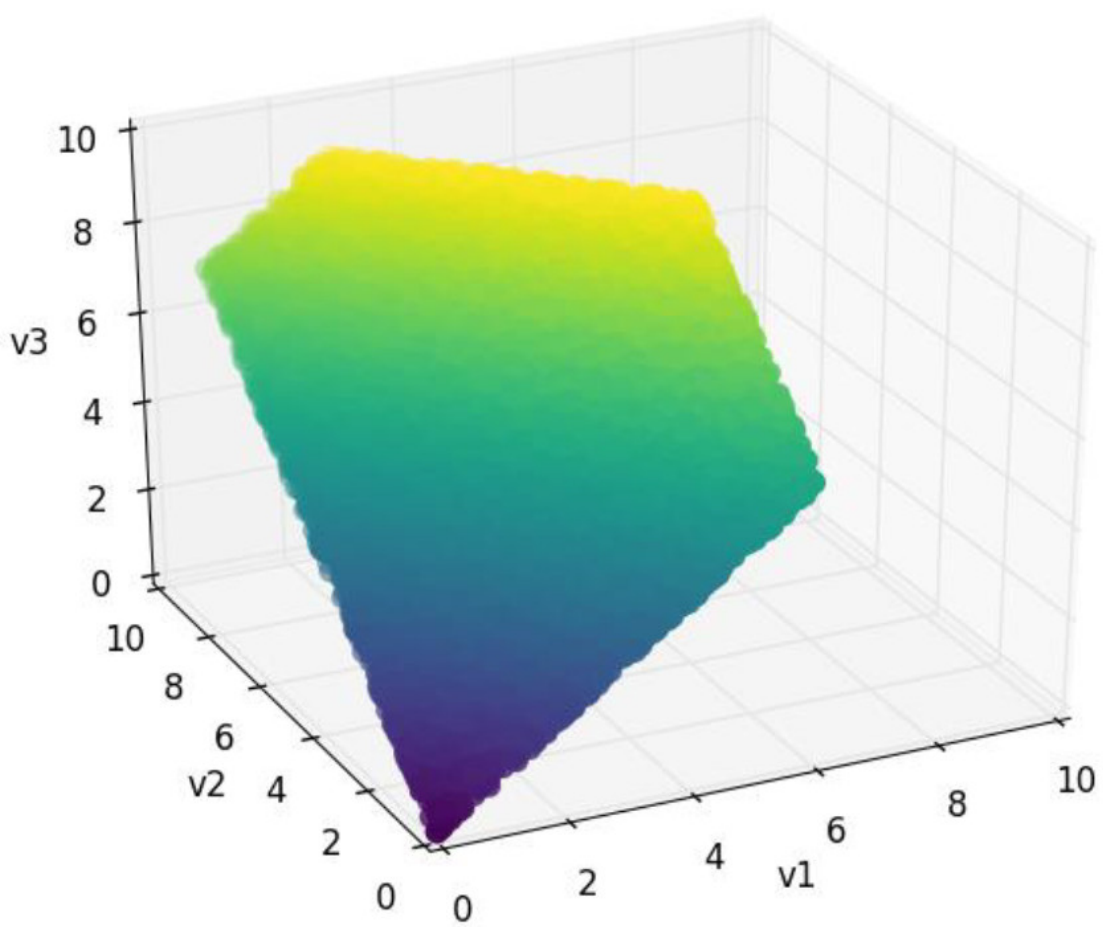
Suppose we have a very simple system, with just three flows:

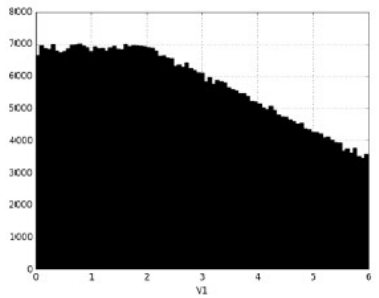
	v_1	v_2	v_3
max	6	8	10
min	0	0	0

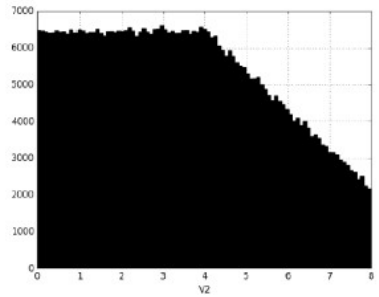


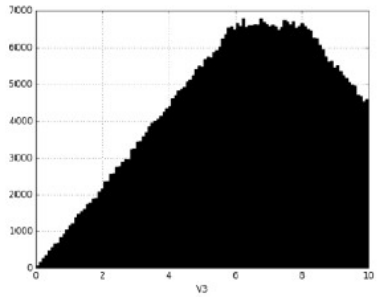
Simple example - volume (cont)

We can see that the actual set of configurations is a 2D surface, and that its projections on a set of arbitrary coordinates do not have the same size.

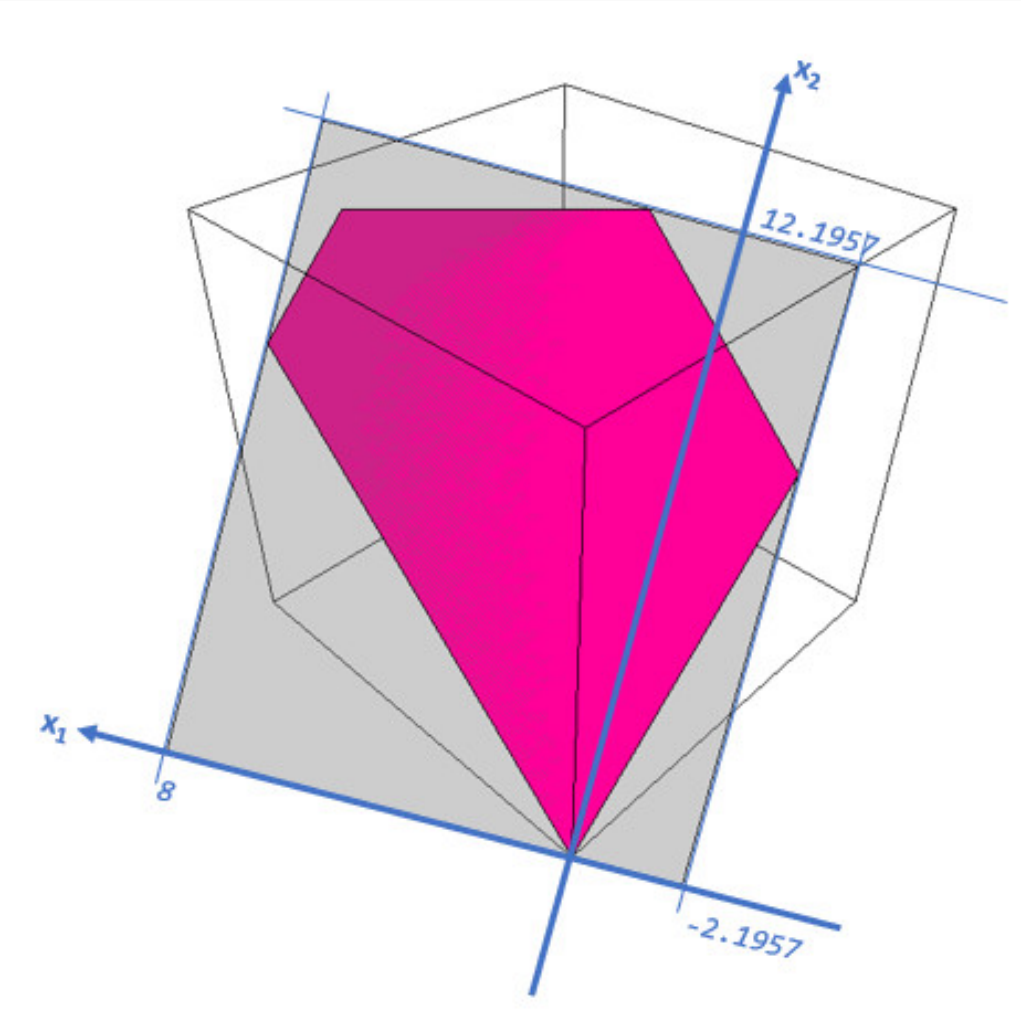








Coordinates change



Volume computation procedure

1. Eliminate linearly dependent rows of stoichiometric matrix S (S' full rank matrix, rank d - dimension of the volume of interest).
2. Transpose S', find QR decomposition. The last d columns of Q are an orthonormal basis for the null space of S'^T - python libraries.
3. Find bounding box (minimizing and maximizing each direction, subject to problem constraints) - $2d$ linear programming problems, glpk. Compute volume of bounding box B (product of differences between max and min value in each dimension).
4. Sample N points within bounding box. Each point X is a vector of independent uniform values, when multiplied by W it is transformed back into a vector in the original space,where it can be checked whether it satisfies all constraints.
5. Let N_c be the number of points which satisfy the constraints, $\widehat{V} = N_c/N \times B$ is the estimation of the volume of interest.

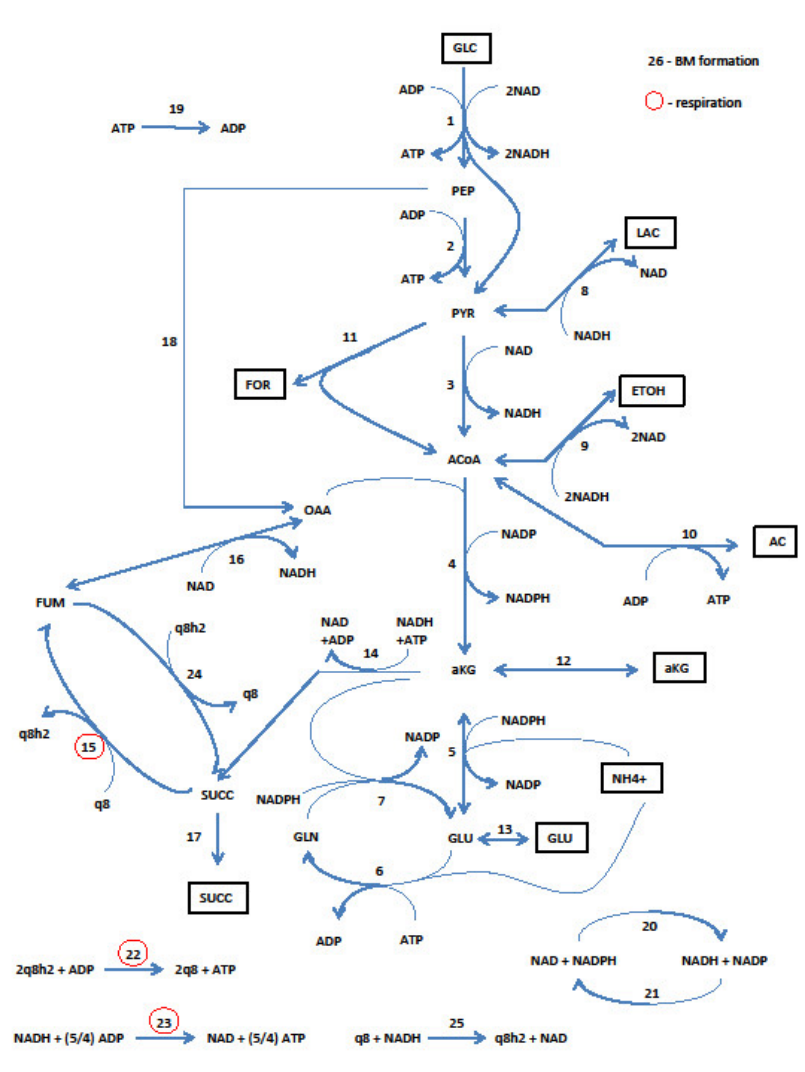
When the volume of interest is a small proportion of the bounding box, "rare event". Alternatives: importance sampling, MCMC.

Computational tools used

- Python, anaconda installation with numpy, scipy, linalg libraries.
- GLPK linear programming solver.
- Tried out COBRApy MCMC sampler, found errors reported to the community (sampling restricted to a subspace of the whole volume).
- Equipments used: desktop computers for development and initial computatons; small server, 72 Gb RAM, 24 cores AMD Opteron 6172 processor, for larger experiments.

Reduced core model for Escherichia Coli

- 16 metabolites, 26 reactions.
- Includes respiration and 6 fermentation modules.



Main energy generating reactions within cell:

Respiration (reactions 15, 22 and 23 in model).
Fermentation reactions:

- LAC Lactic acid
- ETOH Ethanol
- AKG, α-ketoglutarate
- AC Acetate
- FOR Formate
- GLU Glutamic acid

Volumes computed for fermentation reactions

Base case (only respiration, no fermentations): Bounding Box= 1.8018e+08, Volume≃1.6648e+06, CoV≃0.1% .

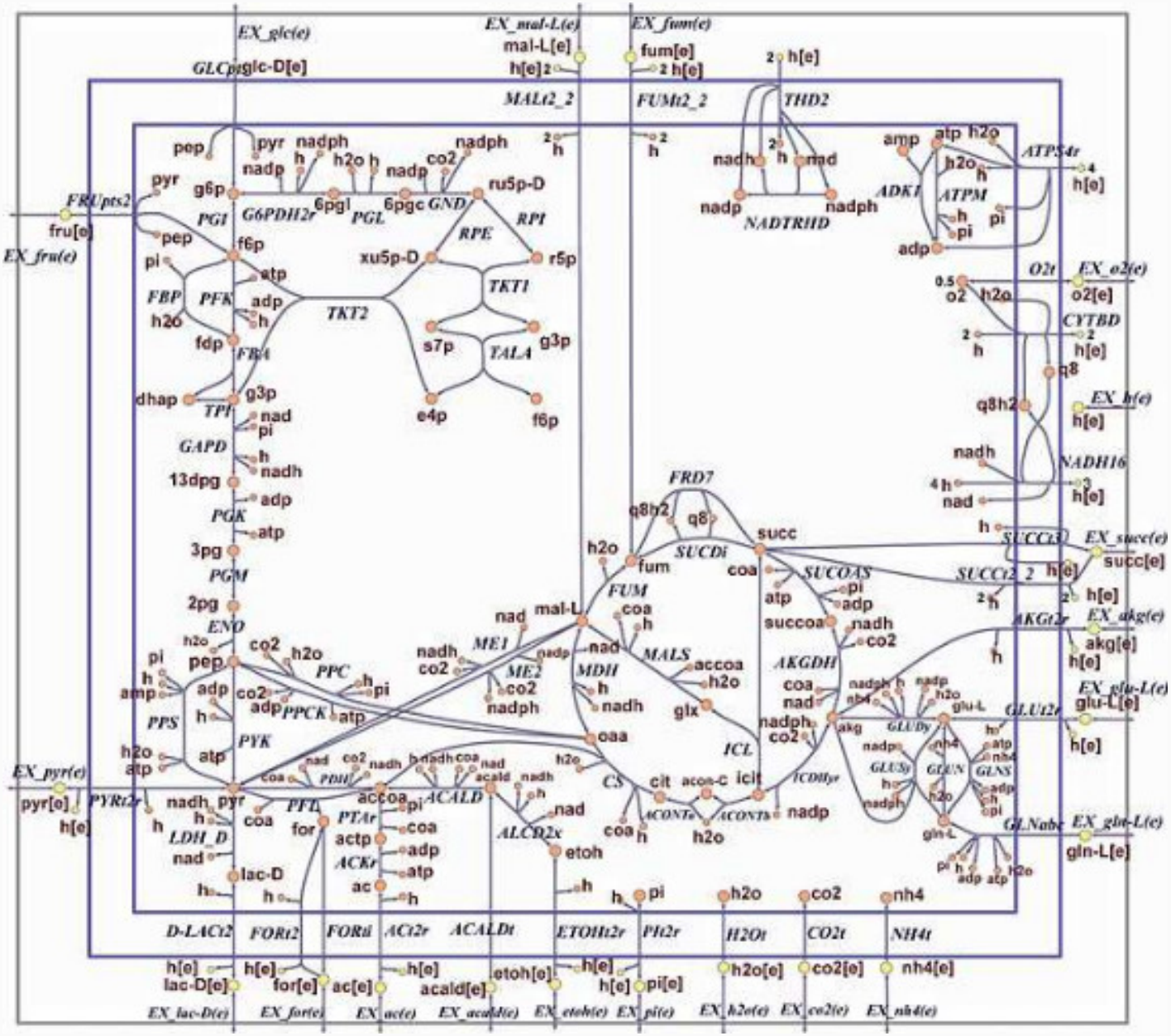
Fermentation:	Bounding box	Volume
AC	6.2797e+11	3.3938e+08
GLU	1.7571e+11	2.1866e+08
aKG	3.1622e+11	1.2583e+08
ETOH	1.2758e+11	5.3491e+07
LAC	6.5172e+10	4.3457e+07
FOR	2.7336e+10	9.6865e+06

Fermentations are ordered by decreasing associated variability (with strong differencies). Observe that if instead of actual volume, the bounding box volume is used (like FVA), then aKG would be thought of inducing almost twice the variability induced by GLU (while actually the opposite is true).

Conclusions, work in progress and future work

- Volume computation of metabolic configurations can help understanding behavior of cells.
- Monte Carlo methods (using linear programming in sampling set up) are a powerful tool to approximately compute these volumes.
- Proxies for volume (like Flux Variability Analysis) do not always reflect properly actual volumes/variabilities.
- (wip) Study interactions among different fermentation modes, and discusion about actual co-existence of fermentation and respiration.
- (tbd) Analyze internal, external and growth variabilities (like the work by Magdalena San Román) using volume indicators.
- (tbd) Employ more sophisticated sampling strategies to tackle higher dimensionality models.

Full core model for E.Coli



Thanks!