Rational Design of Organelle Compartments in Cells

Claudio Angione

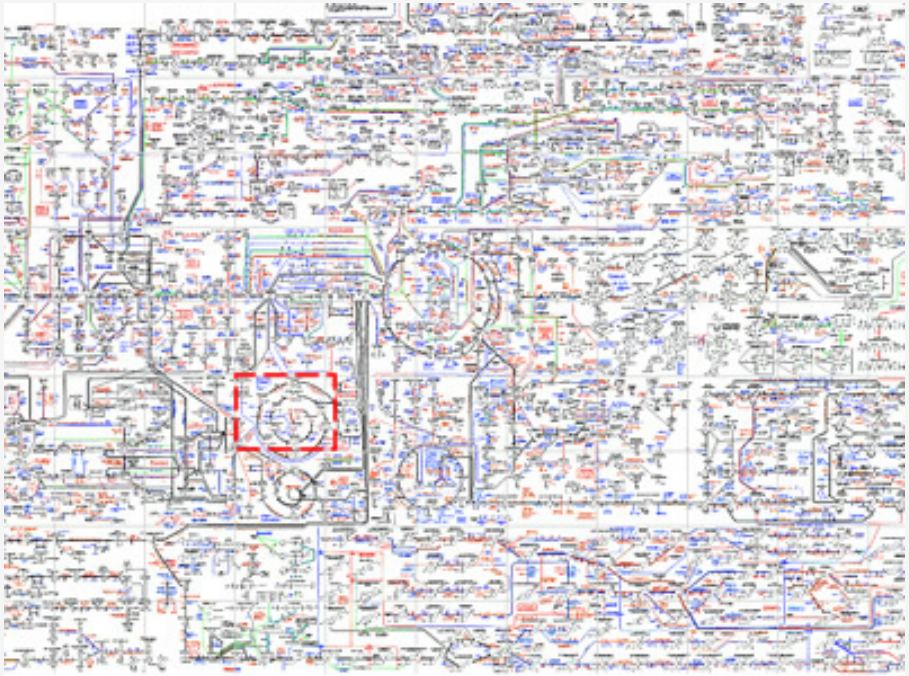
UNIVERSITY OF CAMBRIDGE

Nettab 2012

Motivation

- Metabolic engineering requires mathematical models for accurate design purposes
- **Aim**: overproducing desired substances
- **Problem**: identify the interventions needed to produce the metabolite of interest
- **Tools**: optimisation, sensitivity, robustness, identifiability

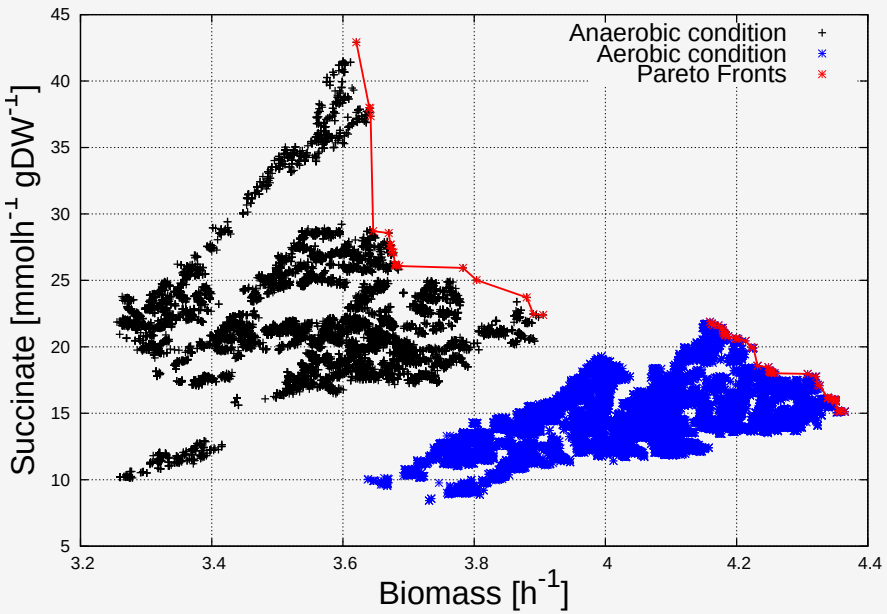
Obstacles



- Large number of reactions occurring in the cellular metabolism
- Large size of the solution space

Idea

- We use a multi-objective optimisation algorithm to seek the manipulation that optimise multiple cellular functions
- The idea is to use and improve the Pareto optimal solutions
- Pareto optimality is important to obtain not only a wide range of Pareto optimal solutions, but also the best trade-off design



Outline

Organelle models:

- Chloroplast model, 31 ODEs + equations for conserved quantities [Zhu et al., 2007]
- Mitochondrion model, 73 DAEs [Bazil et al., 2010]
- Hydrogenosome model, Flux Balance Analysis [Angione et al., submitted]

Common framework

1. *Sensitivity analysis*
2. *Multi-objective optimisation*
3. *Robustness analysis*
4. *Identifiability analysis*

Outline

Organelle models:

- Chloroplast model, 31 ODEs + equations for conserved quantities [Zhu et al., 2007]
- Mitochondrion model, 73 DAEs [Bazil et al., 2010]
- Hydrogenosome model, Flux Balance Analysis [Angione et al., submitted]

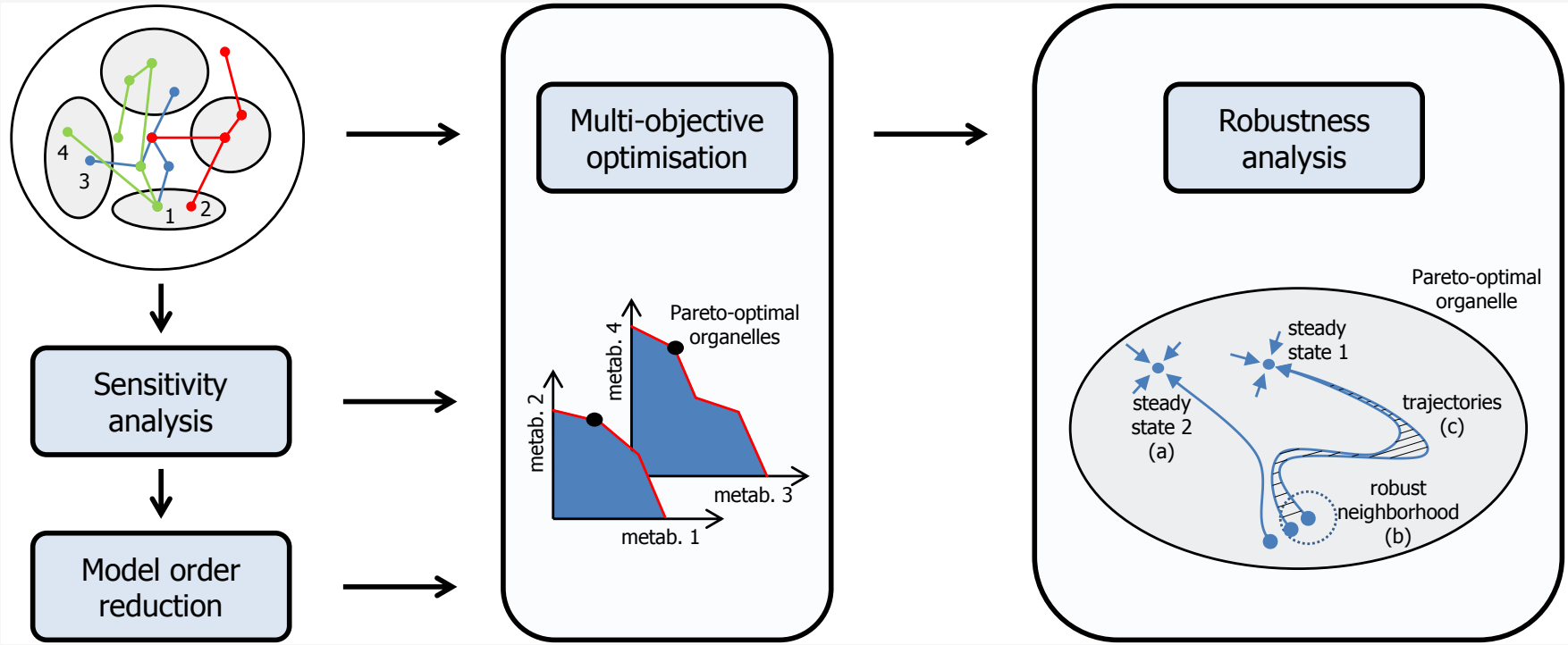
Common framework

1. *Sensitivity analysis*
2. *Multi-objective optimisation*
3. *Robustness analysis*
4. *Identifiability analysis*

Why a common framework for organelles?

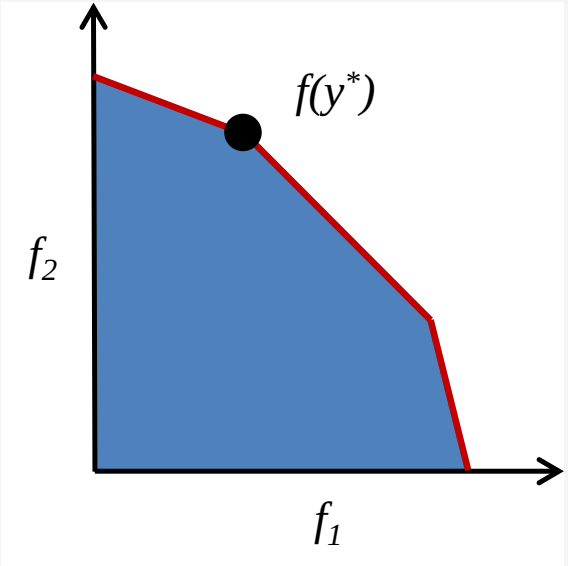
- All extant eukaryotes are descended from an ancestor that had a mitochondrion
- The evolutionary history of chloroplasts and mitochondria are intertwined
- The possibility of multi-objective optimisation related to the different tasks (e.g. maximising the ATP and the heat)
- Identify and cross-compare the most important components
- Assess the fragileness of the multi-optimised metabolic networks using the robustness analysis

The common framework



Multi-objective optimisation

Let f be the vector of r objective functions to optimise in the objective space



- Solution of a multi-objective problem: set of points called Pareto front
- Represents the best trade-off between two or more requirements
- A point y^* in the solution space is Pareto optimal if there does not exist a point y such that $f(y)$ dominates $f(y^*)$, i.e. $f_i(y) > f_i(y^*), \forall i = 1, ..., r$

Genetic design through multi-objective optimisation (GDMO) [Costanza et al. Bioinformatics, 2012]

Seek an optimal initial array of concentrations through an evolutionary algorithm inspired by NSGA-II [Deb et al., 2002]

1. generate initial population P(t)
2. evaluate the fitness of each individual in P(t)
3. while (not termination condition) do
    1. select parents, Pa(t) from P(t) based on their fitness in P(t)
    2. apply crossover to create offspring from parents: Pa(t) -> O(t)
    3. apply mutation to the offspring: O(t) -> O'(t)
    4. evaluate the fitness of each individual in O'(t)
    5. select population P(t+1) from current offspring O'(t) and parents Pa(t)

Genetic design through multi-objective optimisation (GDMO) [Costanza et al. Bioinformatics, 2012]

Seek an optimal initial array of concentrations through an evolutionary algorithm inspired by NSGA-II [Deb et al., 2002]

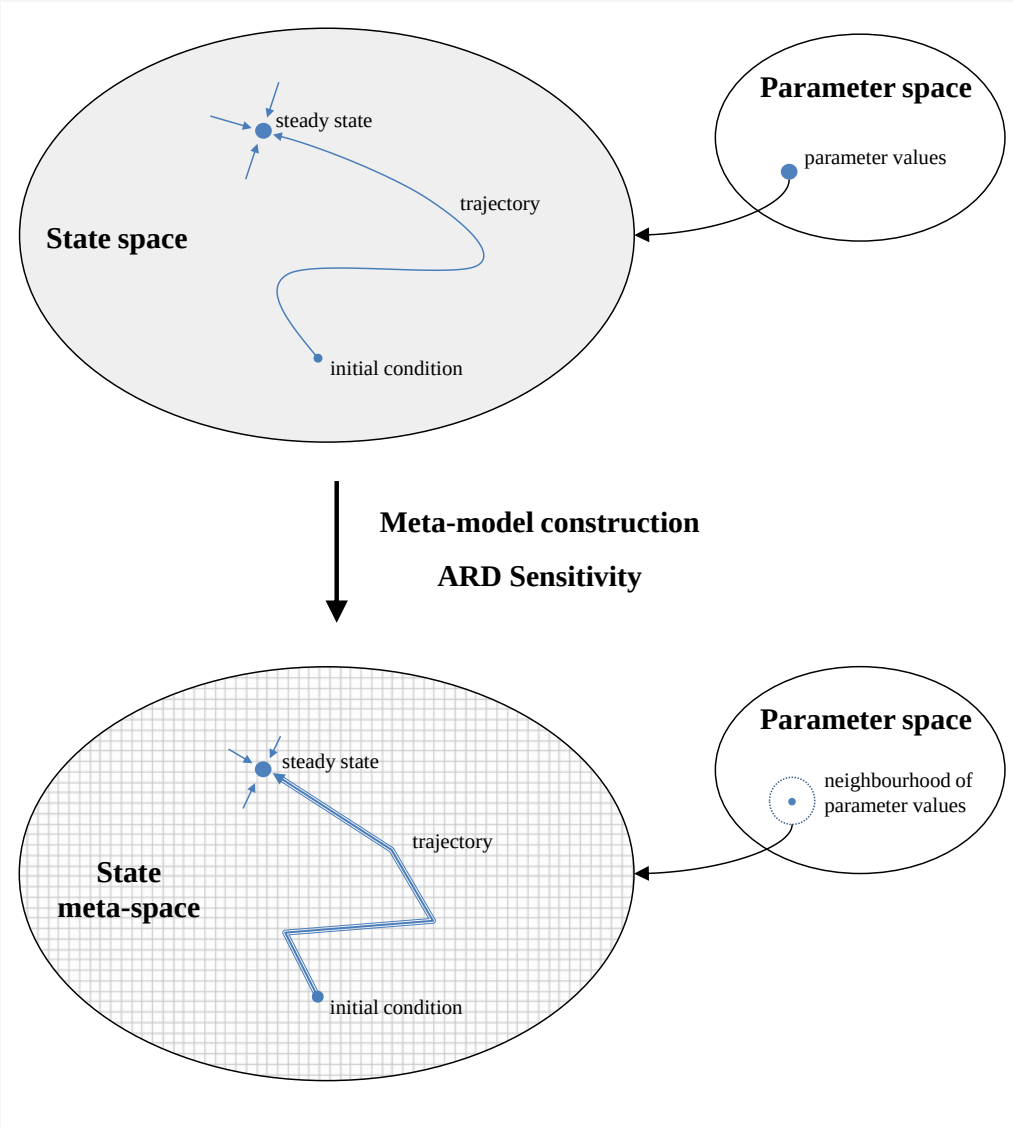
1. generate initial population P(t)
2. evaluate the fitness of each individual in P(t)
3. while (not termination condition) do
    1. select parents, Pa(t) from P(t) based on their fitness in P(t)
    2. apply crossover to create offspring from parents: Pa(t) -> O(t)
    3. apply mutation to the offspring: O(t) -> O'(t)
    4. evaluate the fitness of each individual in O'(t)
    5. select population P(t+1) from current offspring O'(t) and parents Pa(t)

Genetic design through multi-objective optimisation (GDMO) [Costanza et al. Bioinformatics, 2012]

Seek an optimal initial array of concentrations through an evolutionary algorithm inspired by NSGA-II [Deb et al., 2002]

1. generate initial population P(t)
2. evaluate the fitness of each individual in P(t)
3. while (not termination condition) do
    1. select parents, Pa(t) from P(t) based on their fitness in P(t)
    2. apply crossover to create offspring from parents: Pa(t) -> O(t)
    3. apply mutation to the offspring: O(t) -> O'(t)
    4. evaluate the fitness of each individual in O'(t)
    5. select population P(t+1) from current offspring O'(t) and parents Pa(t)

Model Reduction and Sensitivity Analysis



Organelle complete model

State space reflects metabolism

Very accurate

High-dimensional parameter space

Computationally expensive to analyse

Organelle metamodel

Approximation of the real model

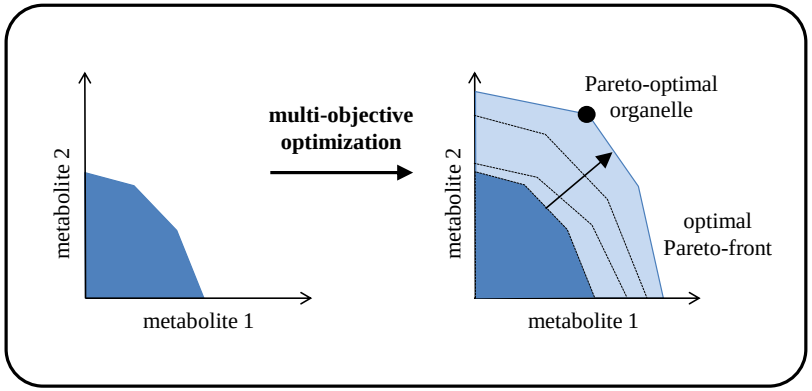
Easy to analyse

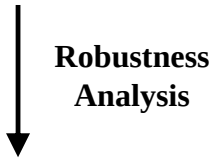
Investigate the sensitivity and robustness

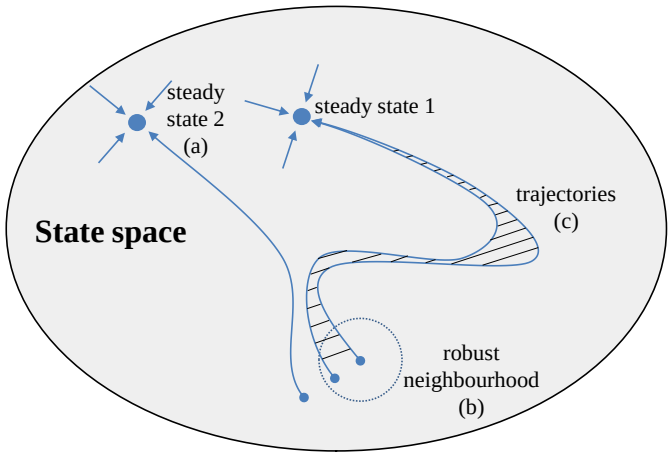
Same initial condition

Slightly different trajectories

Multi-objective Optimisation and Robustness Analysis







Multi-objective optimisation

Move the front towards the best Pareto-front

Maximise metabolites (e.g. ATP vs NADH)

Choose Pareto-optimal organelle

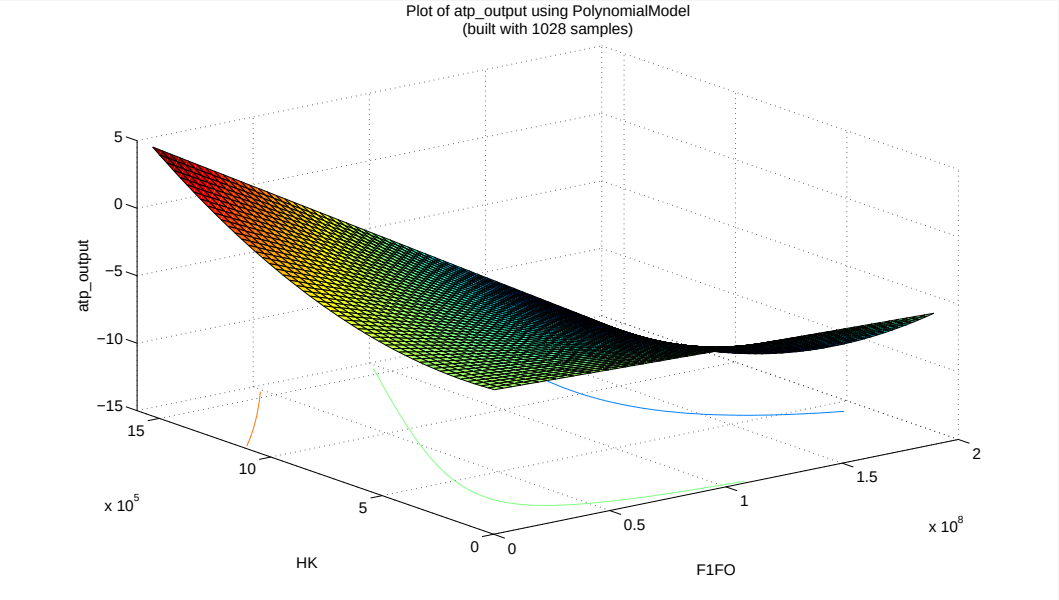
Robustness of the Pareto optimal organelle

(a) Maintains its functionality if it transits through a new steady state [Kitano, 2007]

(b) Robustness to change of initial conditions [Gunawardena, 2009]

(c) Percentage of perturbation trials such that the output remains in a given interval [Stracquadanio & Nicosia, 2011]

ARD Sensitivity and Reduction in the Mitochondrion Metamodel



Metamodel

Closer look at the model behaviour

Polynomial surrogate models

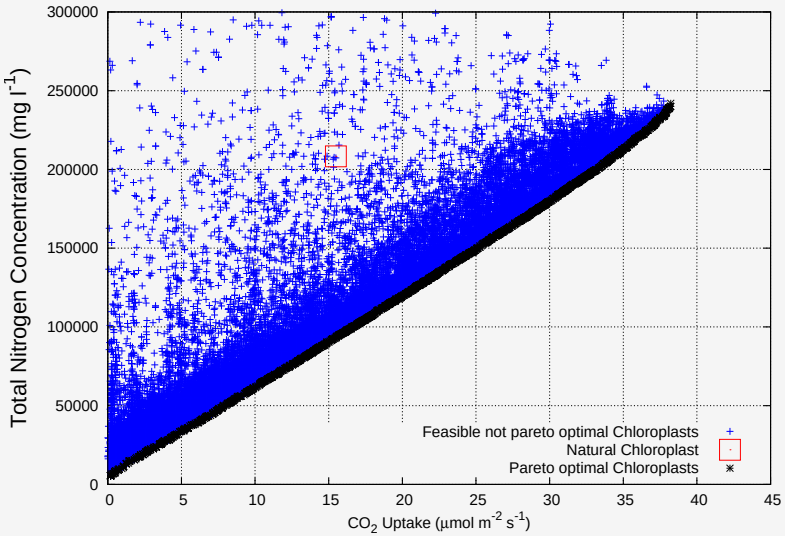
1028 samples in the parameter space

Second order model: $a_0 + c^\intercal p + p^\intercal A p$

p = array of parameters

- Most sensitive parameters: Hexokinase max rate (HK), F_1F_0 ATP synthase activity
- Low values of HK: changes in F_1F_0 have little effect on the ATP production
- High values of HK: the mitochondrion is highly sensitive to variations of F_1F_0

Multi-objective Optimisation in the Chloroplast Model



CO_2 uptake rate vs. protein nitrogen consumption

Sensitive domain: 11 most sensitive enzymes

Multi-objective optimisation in the "sensitive domain"

The other 12 enzymes kept at their nominal value

Goal: Higher CO_2 uptake employing less nitrogen

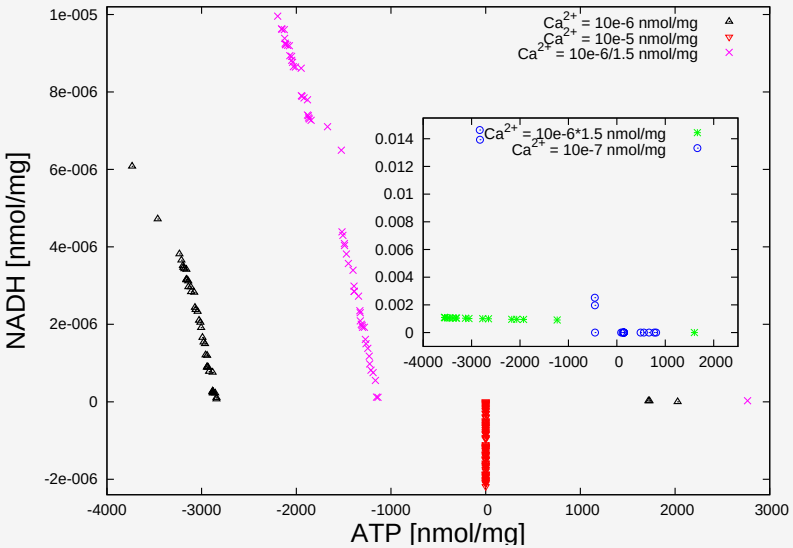
Absorbing more CO_2 while consuming less "leaf-fuel"

Find all those sensitive enzyme concentration vectors $\hat{x} = (c_1, c_2, \ldots, c_{11})$ such that the resulting CO_2 uptake function is maximised and the nitrogen consumption is minimised. This renders the metabolism cycle more efficient.

$$\max_{\hat{x} \in \mathbb{R}^{11}} \left(f_1(\hat{x}), -f_2(\hat{x})\right)^T$$

$f_1 = CO_2$ uptake, $f_2 =$ nitrogen consumption

Multi-objective Optimisation in the Mitochondrion Model



NADH vs. ATP Pareto fronts

Different Ca^{2+} concentrations

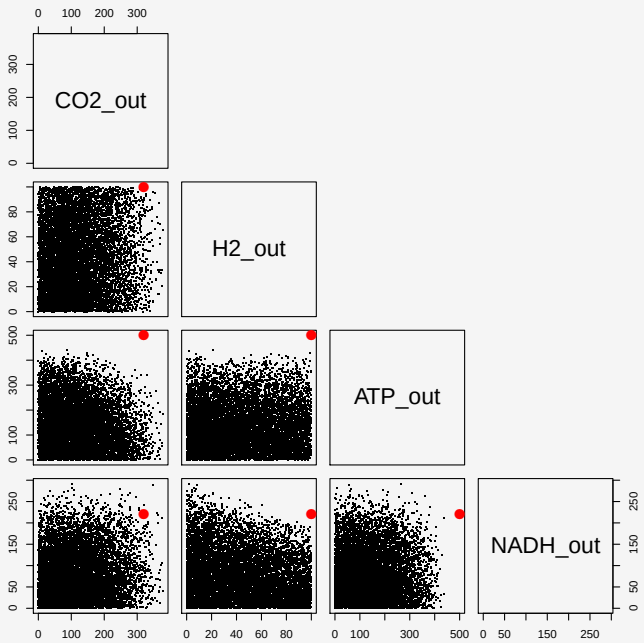
Goal: Higher ATP and NADH

Genetic algorithm to move the Pareto-front

Before the optimisation
$\text{NADH} = 1.5987 \cdot 10^{-10}$ nmol/mg (formation)
$\text{ATP} = -0.0014$ nmol/mg (consumption)

- $f_1 =$ ATP production, $f_2 =$ NADH production

Multi-objective Optimisation in the Hydrogenosome Model



Hydrogenosome Pareto front

MCMC sampling of the reaction network

Trade-offs among the maximisations

Most reactions are coupled

NADH and H_2 are in contrast

H_2 and CO_2 seem uncorrelated

Red points are the optimal points

CO_2-NADH plot: the hydrogenosome is versatile and can produce both, but it cannot specialise in producing only one metabolite (higher curvature of the front).

Robustness Analysis

- Assess the ability of a system to preserve its behaviour despite internal or external perturbations
- Perturbation $\gamma(\Psi, \sigma)$: applies a stochastic noise σ to the system Ψ
- Generate a set T of trial samples $\tau = \gamma(\Psi, \sigma)$

An element $\tau \in T$ is said to be robust to the perturbation [Stracquadanio & Nicosia, 2011], due to stochastic noise σ, for a given property (or metric) ϕ, if:

$$\rho(\Psi, \tau, \phi, \epsilon) = \begin{cases} 1 & \text{if } |\phi(\Psi) - \phi(\tau)| \leq \epsilon \\ 0 & \text{otherwise,} \end{cases}$$

where Ψ is the *reference system*, ϵ is a *robustness threshold*.

Robustness of a system Ψ: the percentage of robust trials

$$\Gamma(\Psi, T, \phi, \epsilon) = \frac{\sum_{\tau \in T} \rho(\Psi, \tau, \phi, \epsilon)}{|T|}.$$

Robustness Analysis

- Assess the ability of a system to preserve its behaviour despite internal or external perturbations
- Perturbation $\gamma(\Psi, \sigma)$: applies a stochastic noise σ to the system Ψ
- Generate a set T of trial samples $\tau = \gamma(\Psi, \sigma)$

An element $\tau \in T$ is said to be robust to the perturbation [Stracquadanio & Nicosia, 2011], due to stochastic noise σ, for a given property (or metric) ϕ, if:

$$\rho(\Psi, \tau, \phi, \epsilon) = \begin{cases} 1 & \text{if } |\phi(\Psi) - \phi(\tau)| \leq \epsilon \\ 0 & \text{otherwise,} \end{cases}$$

where Ψ is the *reference system*, ϵ is a *robustness threshold*.

Robustness of a system Ψ: the percentage of robust trials

$$\Gamma(\Psi, T, \phi, \epsilon) = \frac{\sum_{\tau \in T} \rho(\Psi, \tau, \phi, \epsilon)}{|T|}.$$

Robustness Analysis

- Assess the ability of a system to preserve its behaviour despite internal or external perturbations
- Perturbation $\gamma(\Psi, \sigma)$: applies a stochastic noise σ to the system Ψ
- Generate a set T of trial samples $\tau = \gamma(\Psi, \sigma)$

An element $\tau \in T$ is said to be robust to the perturbation [Stracquadanio & Nicosia, 2011], due to stochastic noise σ, for a given property (or metric) ϕ, if:

$$\rho(\Psi, \tau, \phi, \epsilon) = \begin{cases} 1 & \text{if } |\phi(\Psi) - \phi(\tau)| \leq \epsilon \\ 0 & \text{otherwise,} \end{cases}$$

where Ψ is the *reference system*, ϵ is a *robustness threshold*.

Robustness of a system Ψ: the percentage of robust trials

$$\Gamma(\Psi, T, \phi, \epsilon) = \frac{\sum_{\tau \in T} \rho(\Psi, \tau, \phi, \epsilon)}{|T|}.$$

Robustness of Enzymes in Natural Chloroplast

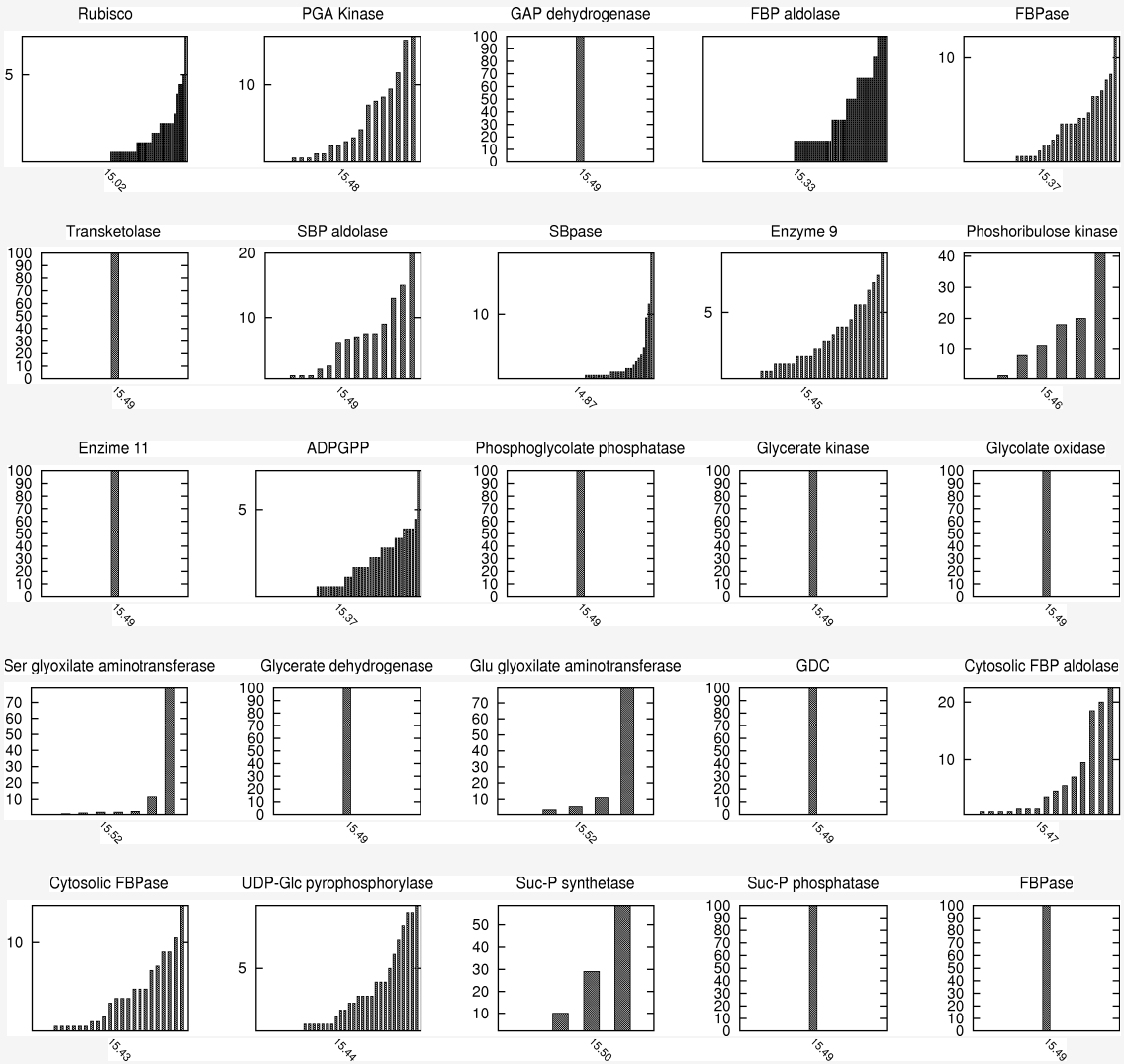
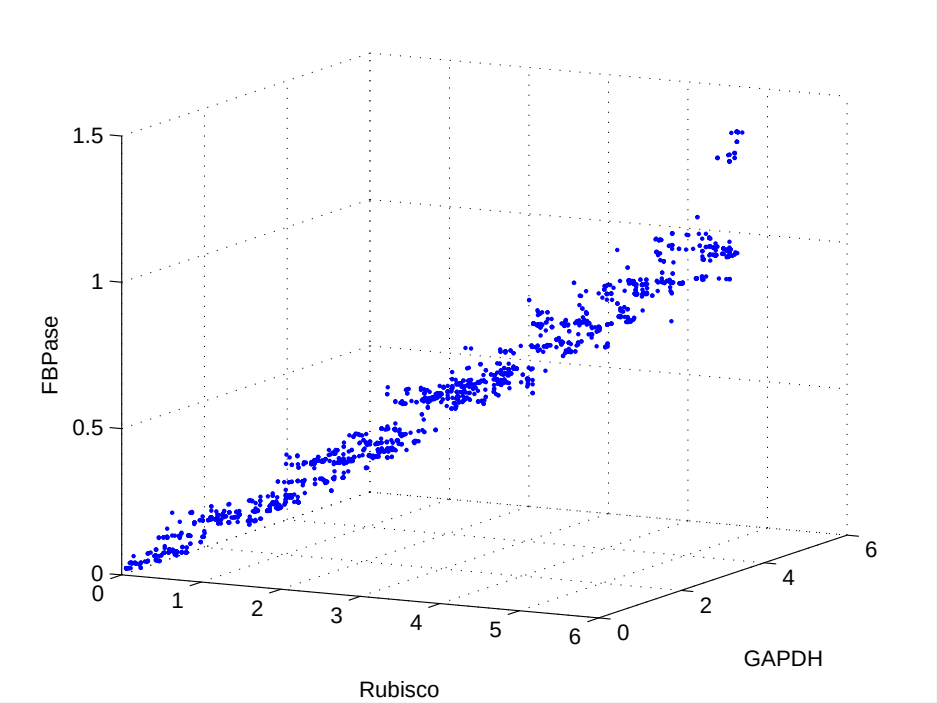
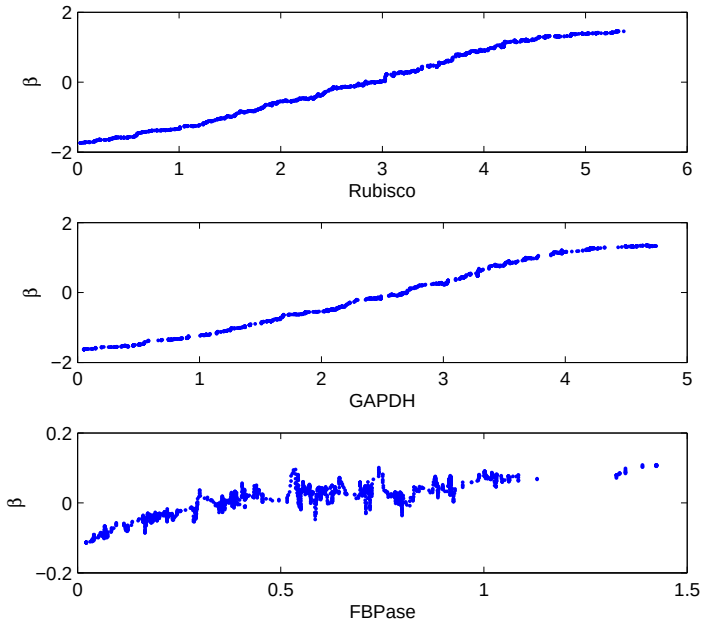


Figure: The chloroplast is robust to perturbations of the enzyme concentration if the CO_2 uptake rate is close to the nominal value ($15.48 \mu mol/m^2 s$) in the majority of the perturbation trials.

Identifiability Analysis in Chloroplast





- Detect relations among decision variables of the optimisation
- Structural non-identifiability: functional relation among decision variables

Network of Organelles

- Systems composed of different species living and interacting in the same organism
- Design, analyse and optimise the "global" metabolism
- Optimise two or more objectives in different organelles simultaneously
- Highlight the complementarity of different metabolisms

Example

- *Mitochondria and chloroplasts are (usually) both found in plants*
- *Part of the same functional pipeline*
- *Starting from CO_2, the photosynthesis in the chloroplast creates glucose that enters the mitochondrion to create ATP*

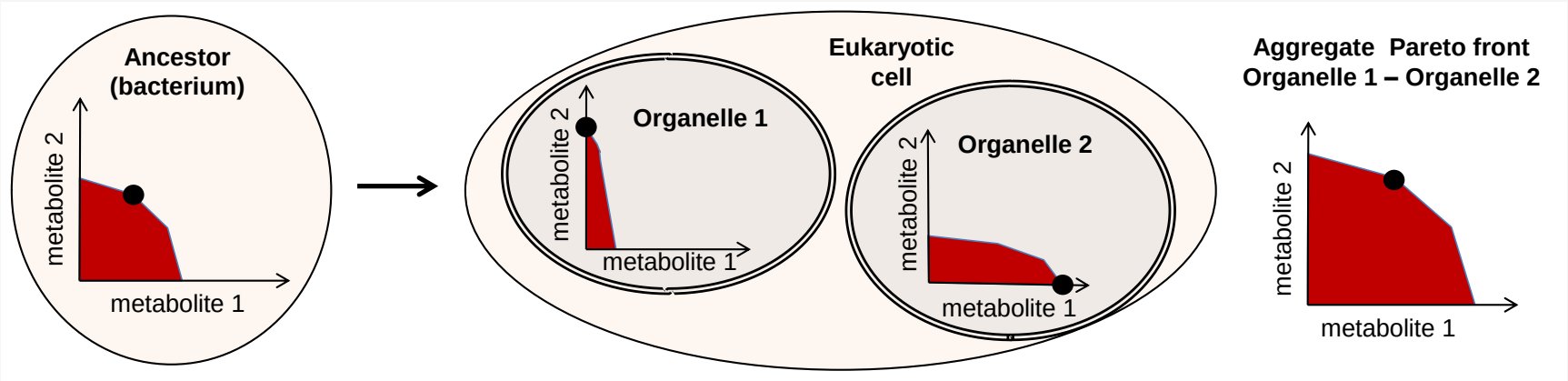
Network of Organelles

- Systems composed of different species living and interacting in the same organism
- Design, analyse and optimise the "global" metabolism
- Optimise two or more objectives in different organelles simultaneously
- Highlight the complementarity of different metabolisms

Example

- *Mitochondria and chloroplasts are (usually) both found in plants*
- *Part of the same functional pipeline*
- *Starting from CO_2, the photosynthesis in the chloroplast creates glucose that enters the mitochondrion to create ATP*

Evolution through Pareto fronts



- The evolution of a Pareto-front can highlight the benefits of an engulfment and subsequent compartmentalisation
- Pareto-optimal point before the engulfments outperformed by the aggregate Pareto-optimal point.

Conclusion

1. Pareto fronts combined with sensitivity, robustness and identifiability
2. Understand the steps of the cellular evolution and the engulfments and specialisation of organelles
3. In silico design to explore the reaction network for the solutions that optimise two or more objectives simultaneously
4. Comprehensive insight into the energy balance in the cell
5. Possible explanation of evolution and compartmentalisation

Acknowledgements

- *Giovanni Carapezza, Dept. of Maths and Computer Science, University of Catania*
- *Jole Costanza, Dept. of Maths and Computer Science, University of Catania*
- *Dr. Pietro Lió, Computer Laboratory, University of Cambridge*
- *Dr. Giuseppe Nicosia, Dept. of Maths and Computer Science, University of Catania*