



HAL
open science

Logic and linear programming for seed identification in metabolic networks

Chabname Ghassemi Nedjad, Clémence Frioux, Loïc Paulevé

► To cite this version:

Chabname Ghassemi Nedjad, Clémence Frioux, Loïc Paulevé. Logic and linear programming for seed identification in metabolic networks. Biorégul 2023 - Modélisation Formelle de Réseaux de Régulation Biologique, Jun 2023, Porquerolles (Hyères), France. pp.1-1, 2023. hal-04328778

HAL Id: hal-04328778

<https://inria.hal.science/hal-04328778>

Submitted on 7 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Logic and linear programming for seed identification in metabolic networks

Chabname GHASSEMI NEDJAD^{1,2} and Clémence FRIOUX¹ and Loïc PAULEVÉ²

¹ Inria Université de Bordeaux, 33400 Talence, France

² LaBRI, CNRS, Université de Bordeaux, 34300 Talence, France

From Genome-Scale Metabolic Network to behaviour prediction

Metabolic modelling:

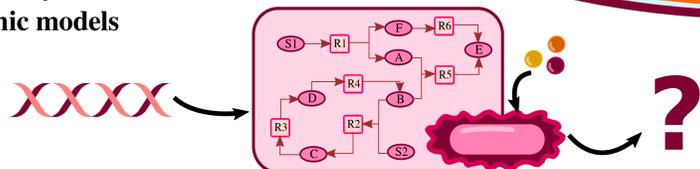
- Genome-Scale metabolic Networks (GSMNs) [1]
- Producibility modelling : **network expansion** [2]

Prediction of behaviour of an organism in a particular environment:

- Environmental response data
- Mathematical modeling

Linear programming problems to solve:

- Flux Balance Analysis** (standard)
- Discrete dynamic models**



Reverse ecology [3]: Identificaton of environmental nutrients

What ?

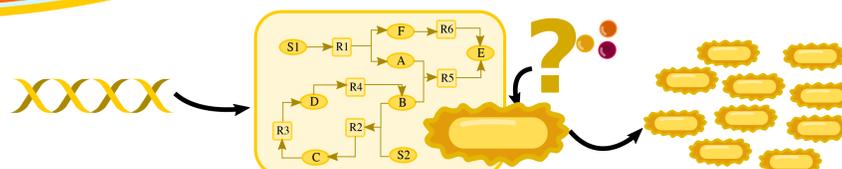
- Seeds:** precursor compounds to produce essential metabolites

Why ?

- Ensure the **growth of a species**, represented by the **biomass reaction**

Purpose ?

- Understand the link between a system and its environment [4]
- Design of culture** for uncultivated species



ASP

Answer Set Programming

Declarative programming paradigm.
Solve **combinatorial** problems by designing a **set of rules** which represent the **knowledge** of the system

FBA Method

Assumption:

- Steady State

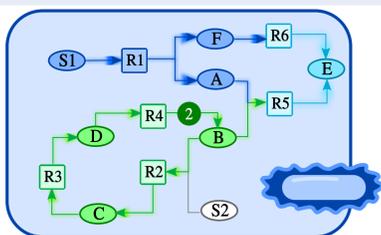
Approach:

- Linear Programming

Formula:

No Accumulations
 $Input - Output = 0$

S : Stoichiometry Matrix
 v : Vector of fluxes
 $S \cdot v = 0$



Solutions



OR representation:
The solution contains one of the seeds inside the bubble

AND representation:
The solution contains as many as there is bubbles

Metabolite	Reaction	Flux	Stoichiometry
Green circle	Green square	Green arrow	Green circle
Blue circle	Blue square	Blue arrow	Blue circle
Light blue circle	Light blue square	Light blue arrow	Light blue circle
White circle	White square	White arrow	White circle

Result

- 1 to 20 subset minimal models found for few species (≤ 20 min)
- 1 minimize model for 9 species

Hybrid Method

Seeds detection rules

```
% Create import reactions for seeds
ireaction(M) :- seed(M).
iproduct(M,I,ireaction(M)) :- seed(M).
bounds(ireaction(M),0,1000) :- seed(M).

&dom(L,U) = R :- bounds(R,L,U), scopeR(R).
&sum R | = 0 :- reaction(R), not scopeR(R).

intern(M) :- metabolite(M), reactant(M,_R), not seed(M).

&sum IS*IR : product(M,IS,IR);
-OS*OR : reactant(M,OS,OR);
IOS*IOR : iproduct(M,IOS,IOR)
| = 0 :- intern(M), fluxbalance.

&sum R : objective(R) | >= "0.0001"
```

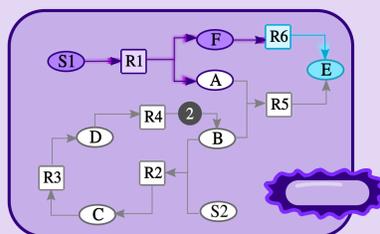
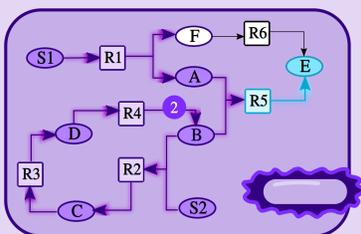
LP-ASP

Assumption:

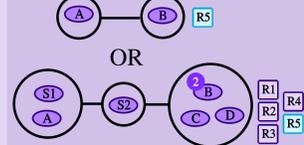
- Transient dynamic

Approach :

- Discrete Programming
- Linear Programming



Solutions



Metabolite	Reaction	Flux	Stoichiometry
Purple circle	Purple square	Purple arrow	Purple circle
Light purple circle	Light purple square	Light purple arrow	Light purple circle
White circle	White square	White arrow	White circle

Result

- 10 to 20 subset minimal models found for few species (≤ 20 min)
- 1 minimize model found for 2 species

Discrete Method

Assumption:

- Transient dynamic

Approach :

- Discrete Programming

Seeds detection rules

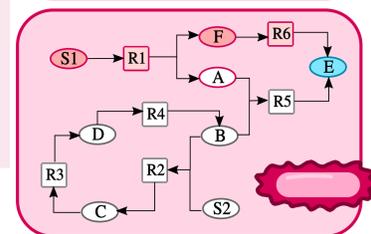
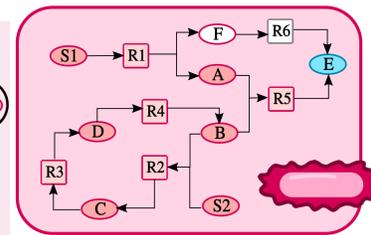
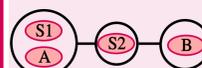
```
can_reach(M) :- target(M).
can_reach(M) :- reactant(M,R) ; product(P,R) ; can_reach(P).

% Choose a set of seed from the possible seeds
{ seed(M) : can_reach(M) }.

activated(S) :- seed(S).
scopeR(R) :- reaction(R) ; activated(M) : reactant(M,_R).
activated(M) :- product(M,_R) ; scopeR(R).

:- target(M) ; not activated(M).
```

Solutions



Metabolite	Reaction	Flux
Red circle	Red square	Red arrow
Light red circle	Light red square	Light red arrow
White circle	White square	White arrow

Result

- 20 subset minimal models found for the whole benchmark (≤ 20 min)
- 20 minimize models found for 3 species

Data and additional approaches

Data:

- 108 GSMNs from database BIGG
- Up to **thousands of reactions** for a species
- Import Reactions removed
- Validation** : fba with cobra.py

Objectives:

- Full Network:** All metabolites are produced (fba test only on Biomass Reaction described on GSMNs)
- Target:** Focus on producing only the metabolites needed for the Biomass Reaction described on GSMNs

Optimization:

\mathcal{P} : set of solutions $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D} \neq \emptyset$

- Subset Minimal:** $\mathcal{A}, \mathcal{B} \in \mathcal{P}$ such as $\mathcal{A} \cap \mathcal{B} = \mathcal{A}$, $S = \{\mathcal{A}\}$
 $\mathcal{A}, \mathcal{C} \in \mathcal{P}$ such as $\mathcal{A} \cap \mathcal{C} = \emptyset$, $S = \{\mathcal{A}, \mathcal{C}\}$
 $\mathcal{A}, \mathcal{D} \in \mathcal{P}$ such as $\mathcal{A} \cap \mathcal{D} \neq \emptyset$ and $\mathcal{A} \cap \mathcal{D} \neq \mathcal{A}$, $S = \{\mathcal{A} \cap \mathcal{D}\}$

- Minimize:** $\mathcal{A} \in \mathcal{P}$ such as $\forall \chi, \gamma \in \mathcal{A}$ size(χ) = size(γ) and $\mathcal{B} \in \mathcal{P}$ such as $\forall z \in \mathcal{B}$, $\forall \chi \in \mathcal{A}$, size(z) > size(χ)
 $S = \{\mathcal{A}\}$

Discussion and Perspective

Work in Progress:

- Improve hybrid solution (limited tractability)
- Evaluation of solutions of different methods for steady growth
- Evaluate diversity and size of identified seeds compared to other methods

Discussion:

- Improve discrete method to avoid hybrid methods

Perspective:

- Optimization for minimize in order to obtain solutions in less than 20 min

References:

- C. Francke, R. J. Siezen, and B. Teusink, Reconstructing the metabolic network of a bacterium from its genome, Trends in Microbiology, vol. 13, no.11, pp. 550–558, Nov. 2005. doi: 10.1016/j.tim.2005.09.001
- T. Handorf, O. Ebenhööh, R. Heinrich, Expanding Metabolic Networks: Scopes of Compounds, Robustness, and Evolution, Journal of Molecular Evolution, vol. 61, no. 4, pp. 498–512, Jan. 2005. doi: 10.1007/s00239-005-0027-1
- Levy, R., Borenstein, E.: Reverse Ecology: From Systems to Environments and Back. pp. 329–345. Springer, New York, NY (2012). https://doi.org/10.1007/978-1-4614-3567-9_15, http://link.springer.com/10.1007/978-1-4614-3567-9_15
- i, Y.F., Costello, J.C., Holloway, A.K., Hahn, M.W.: Reverse ecology and the power of population genomics. Evolution; international journal of organic evolution 62(12), 2984–94 (dec 2008). https://doi.org/10.1111/j.1558-5646.2008.00486.x, http://www.ncbi.nlm.nih.gov/pubmed/18752601