DynCubeProd: Dynamic solution space division-based methods for calculating reaction deletion strategies for constraint-based metabolic networks for substance production

Yier Ma[1], Takeyuki Tamura[1]

1. *Bioinformatics Center, Institute for Chemical Research, Kyoto University, Japan*

● Contact: mayier@kuicr.kyoto-u.ac.jp

Abstract

Flux balance analysis (FBA) is a crucial method to analyze large-scale constraint-based metabolic networks in metabolic engineering. Based on FBA, GridProd divides the solution space into squares by the cell growth rate and the target metabolite production rate to efficiently find the reaction deletion strategies and was extended to CubeProd. However, GridProd and CubeProd may naively examine places where solutions are unlikely to exist. To address this issue, we introduce dynamic solution space division methods called DynCubeProd for faster computing by avoiding searching for the places where the solutions do not exist. We applied DynCubeProd to Escherichia coli dataset iJO1366 and it significantly accelerated the calculation of the reaction deletion strategy for each target metabolite production. Under the anaerobic condition, DynCubeProd could obtain the reaction deletion strategies for almost 40% of target metabolites that the elementary flux vector-based method could not obtain.

Problem Definition

Our goal was to find reaction deletion strategies for growth coupling of target metabolite production. Let $K = \{vj | vj \in V\}$ be a set of reactions to be knocked out, where V is a set of n reactions. Then, the definition of the main problem of this study arises.

Given

$$S, LB, UB, v_{growth}, v_{target}, x_{growth}^{min}, x_{target}^{threshold}$$

Find

$$K$$

such That

$$x_{growth} \geq x_{growth}^{min} \; and \; x_{target} \geq x_{target}^{threshold}$$

maximize

$$f(x) \; (= x_{growth})$$

subject to:

$$S\boldsymbol{x} = 0$$

$$\begin{cases} x = 0 \; if, x \in K \\ LB \leq x \leq UB, otherwise. \end{cases}$$

$\boldsymbol{x} \in \mathbb{R}^n$ is an n-dimensional variable. $S \in \mathbb{R}^{m \times n}$ is the stoichiometric matrix corresponding to m metabolites and n reactions in the constraint-based models. LB and UB impose the lower and upper bounds of each $x \in \boldsymbol{x}$. When $x_{growth} \geq x_{growth}^{min}$ and $x_{target} \geq x_{target}^{threshold}$ is satisfied, we consider K achieves growth coupling, where $GR = x_{growth}$ for $v_{growth} \in V$ and $PR = x_{target}$ for $v_{target} \in V$ hold.

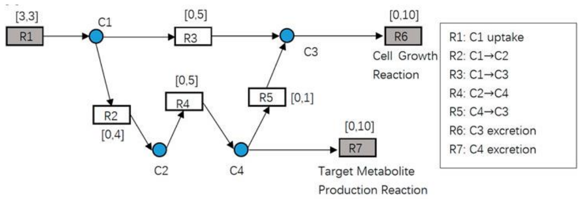


Fig. 1. A toy example of the constraint-based models. Rectangular nodes R1 to R7 are reactions, and the attached intervals represent lower and upper bounds of their reaction speeds. R1 is the nutrient uptake reaction. R6 is the cell growth reaction. R7 is the target metabolite production reaction. Circular nodes C1 to C4 are internal metabolites.

DynCubeProd Method

- DynCubeProd considers the three-dimensional solution space whose axes represent growth rate (GR), production rate (PR) and sum of absolute values of fluxes (SF). Let TMGR, TMPR, and TMSF be the theoretical maximum values of the above, respectively. Then the whole constraint space is a rectangle formed by [0, TMGR], [0, TMPR], and [0,TMSF]. According to the designated value of P, each of [0, TMGR], [0, TMPR], and [0,TMSF] are divided into P pieces. Therefore, finally, DynCubeProd considers $P3$ constraint sub-spaces. DynCubeProd will iteratively search solutions in each sub-spaces. Because the intervals on each of the three axes are equally subdivided into P sub-intervals, constraints below are added and the sum of the absolute values of fluxes is minimized for every $1 \leq i, j, k \leq P$, where i, j, k are integers.

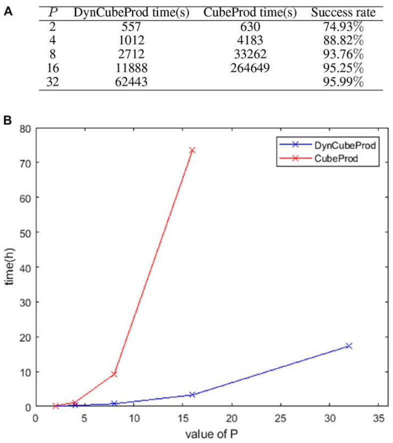
$$\frac{(i-1) \times TMGR}{P} \leq x_{growth} \leq \frac{i \times TMGR}{P},$$

$$\frac{(j-1) \times TMPR}{P} \leq x_{target} \leq \frac{j \times TMPR}{P},$$

$$\frac{(k-1) \times TMSF}{P} \leq \sum |x| \leq \frac{k \times TMSF}{P}$$

- DynCubeProd starts with $P = 1$ and doubles P if no solution is found. When applying a larger P, it refers to the result of applying the smaller P and avoids searching for places where there is no solution.

Computational Experiments



A

P	DynCubeProd time(s)	CubeProd time(s)	Success rate
2	557	630	74.93%
4	1012	4183	88.82%
8	2712	33262	93.76%
16	11888	264649	95.25%
32	62443		95.99%

Fig. 2. (A) Computation time and success ratio when DynCubeProd and CubeProd were applied to iJO1366 under aerobic conditions for different values of P. (B) Visual comparison of the computation time of DynCubeProd and CubeProd of (A).

[References]

[1] Tamura, T. (2018). Grid-based Computational Methods for the Design of Constraint-Based Parsimonious Chemical Reaction Networks to Simulate Metabolite Production: Gridprod. BMC bioinformatics 19, 325. doi:10.1186/s12859-018-2352-6[2] P. Kilpeläinen, H. Mannila: Ordered and Unordered Tree Inclusion. SIAM J. COMPUT. Vol. 24, No. 2, pp. 340-356 (1995)

[2] Tamura, T. (2021a). Efficient Reaction Deletion Algorithms for Redesign of Constraint-Based Metabolic Networks for Metabolite Production with Weak Coupling. IPSJ Trans. Bioinformatics 14, 12–21. doi:10.2197/ipsjtbio.14.12

[3] Tamura, T. (2021b). L1 Norm Minimal Mode-Based Methods for Listing Reaction Network Designs for Metabolite Production. IEICE Trans. Inf. Syst. E104.D, 679–687. doi:10.1587/transinf.2020edp7247