Growth-coupled overproduction is theoretically possible for most metabolites even in *Saccharomyces cerevisiae* under anaerobic condition

Takeyuki Tamura[1,*]

1 Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto, Japan

*** tamura@kuicr.kyoto-u.ac.jp**

Abstract

Metabolic network analysis through flux balance has been established as a method for the computational design of production strains in metabolic engineering. A key principle often used in this method is the production of target metabolites as by-products of cell growth. Recently, the strong coupling-based method was used to demonstrate that the coupling of growth and production is possible for nearly all metabolites through knocking out reactions in genome-scale metabolic models of five major metabolite producing organisms under aerobic conditions. However it remains unknown whether this coupling is always possible using reaction knockouts under anaerobic conditions. Particularly, when growing *Saccharomyces cerevisiae* under anaerobic conditions, knockout strategies using the strong coupling-based method were possible for only 3.9% of all target metabolites. Here, we found that the coupling of growth and production is theoretically possible for 91.6% metabolites in genome-scale models of *S. cerevisiae* even under anaerobic conditions. This analysis was conducted for the worst cases using flux variability analysis. To demonstrate the feasibility of the coupling, we derived appropriate reaction knockouts using a new algorithm for target production in which the search space was divided into small cubes (CubeProd). Our results are of fundamental importance for computational metabolic engineering under anaerobic conditions. The developed software CubeProd implemented in MATLAB and obtained reaction knockouts are available on "http://sunflower.kuicr.kyoto-u.ac.jp/~tamura/CubeProd.zip".

Background

Computational modeling is gaining importance in metabolic engineering for designing and optimizing metabolite producing microbes [1, 5, 7]. One of the key principles in computational strain design is coupling cellular growth with the production of a desired metabolite [18]. This principle is known as growth coupling. However, growth coupling is not possible for every metabolite.

In metabolic engineering simulations, flux balance analysis (**FBA**) is a widely used mathematical model. FBA assumes a pseudo-steady state in which the sum of incoming fluxes must equal the sum of outgoing fluxes for each internal metabolite [12]. FBA can be formalized as linear programming (LP). In this LP, biomass production flux is maximized. This obtained maximum value is referred to as the growth rate (**GR**). For each metabolite, the production rate (**PR**) is estimated under the condition in which the GR is maximized. When both the GR and PR exceed certain criteria, growth coupling is considered as achieved.

Many efficient solvers are available for LP as LP is polynomial-time solvable. Therefore, computational simulation by FBA is efficiently possible even for genome-scale metabolic models. The fluxes calculated by FBA correspond to experimentally obtained fluxes [17].

For FBA, numerous computational methods can be used to identify optimal knockout strategies. OptKnock is a well-established method that identifies optimal reaction knockouts [2]. OptKnock derives knockout strategies through bi-level linear optimization using mixed integer linear programming. However, because mixed integer linear programming is an NP-complete problem, OptKnock cannot identify knockout strategies for many cases in genome scale networks. Many methods have been developed to overcome this problem OptGene and Genetic Design through Local Search (GDLS) identify gene deletion strategies using a genetic algorithm and local search with multiple search paths, respectively [9, 13]. EMILio and Redirector use iterative linear programs [14, 19]. Genetic Design through Branch and Bound (GDBB) uses a truncated branch and branch algorithm for bi-level optimization [3]. Fast algorithm of knockout screening for target production based on shadow price analysis (FastPros) is an iterative screening approach to discover reaction knockout strategies [11]. IdealKnock utilizes the ideal-type flux distribution and ideal point=(GR, PR) [4]. Parsimonious enzyme usage FBA (pFBA) [8] finds a subset of genes and proteins that contribute to the most efficient metabolic network topology under the given growth conditions. GridProd integrates the ideas of IdealKnock and pFBA [16].

In 2017, von Kamp and Klamt developed a strong coupling-based method and showed that growth coupling is possible for more than 90% of all metabolites by reaction knockouts in *Escherichia coli* under aerobic condition, *Saccharomyces cerevisiae* under aerobic condition, *Aspergillu niger*, *Corynebacterium glutamicum*, and *Synechocystis*. They also showed that growth coupling is possible for more than 75% of all metabolites in *E. coli* under anaerobic conditions. However, for *S.*

cerevisiae under anaerobic conditions, the strong coupling-based method could only derive reaction knockout strategies for 3.9% of all metabolites.

Therefore, how to develop a method that achieves growth coupling for *S. cerevisiae* under anaerobic conditions remains unclear. In this study, we developed a novel algorithm, CubeProd, to demonstrate that growth coupling is possible for most metabolites in *S.cerevisiae* under anaerobic conditions.

Results

To determine whether growth coupling is possible for *S. cerevisiae* under anaerobic conditions, we developed the novel algorithm, CubeProd, to derive reaction knockouts that achieve growth coupling. The details about CubeProd and model configurations are described in the Methods section. The associated FBA model for *S. cerevisiae* was iMM904. iMM904 contains 1228 internal metabolites. However, the theoretical maximum yield is zero for 541 of the 1228 metabolites. Therefore, we applied CubeProd for 687 target metabolites in iMM904 which are the same as those evaluated by von Kamp and Klamt [18].

Computations were conducted for three different levels of demanded minimum product yield, 10%, 30%, and 50% of the theoretical maximum yield for each target metabolite as in [18]. When the strong coupling-based method was applied, suitable reaction knockouts were obtained only for 3.9%, 1.7% and 1.3% of target metabolites, respectively [18]. However, when CubeProd was applied, suitable reaction knockouts were found for 91.6%, 69.6%, and 47.6% of all target metabolites for the 10%, 30% and 50% minimum yield level, respectively, as shown in Fig. 1. While the sizes of the reaction knockouts derived through the strong coupling-based method were less than 3% of all reactions, those derived through CubeProd were more than 60%.

All procedures for CubeProd were implemented on a workstation with CPLEX, COBRA Toolbox [15], and MATLAB on a CentOS 7 machine with an Intel Xeon Processor with 2.30 GHz 18C/36T, and 128 GB memory.

Below, we explain the main ideas used in the development of CubeProd. CubeProd calculates reaction knockouts that achieve growth coupling for genome-scale metabolic networks by extending the concept of GridProd [16], which was developed by the author. The main idea of GridProd is to divide the solution space of optKnock into P^{-2} small grids and conduct LP twice for each grid, where P^{-1} is an integer parameter. The first LP identifies the reactions to be knocked out, while the second LP calculates the PR of the target metabolite under the condition in which the determined reactions by the first LP are knocked out and GR is maximized for each grid. The knockout strategy of the grid whose PR is highest is then adopted as the GridProd solution. The area size of each grid in GridProd is $(P \times TMGR) \times (P \times TMPR)$, where TMGR and TMPR are the theoretical maximum growth rate and theoretical maximum production rate, respectively.

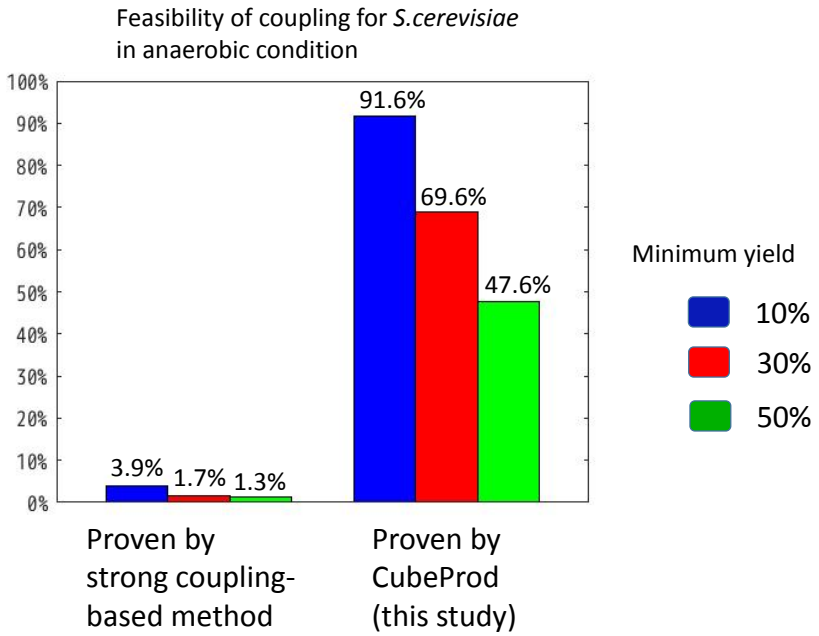

Figure 1. Feasibility of coupling for *S. cerevisiae* under anaerobic conditions demonstrated by strong coupling-based method and CubeProd. The feasibility was demonstrated by calculating reaction knockouts. Percentage of producible organic metabolites is shown for each minimum yield level.

In addition to $(P \times TMGR) \times (P \times TMPR)$, CubeProd divides the solution space by $(P \times TMGR) \times (P \times TMPR) \times (P \times TMTF)$, where TMTF is the theoretical maximum total flux.

Fig. 2 shows the percentages of reaction knockouts obtained by CubeProd that achieved growth coupling for the 687 target metabolites of iMM904 for different cube sizes and demanded minimum yield level. The horizontal axis corresponds to different resolutions (P^{-1}) of CubeProd. The blue, red, black, and green lines correspond to demanded minimum yield levels of 1%, 10%, 30%, and 50%, respectively. When the demanded minimum yield was 1% of the TMY, the percentage was 88.9% with $P^{-1} = 5$ and 93.2% with $P^{-1} \leq 15$. When the demanded minimum yield was 10% of the TMY, the percentage was 76.0% with $P^{-1} = 5$, 90.8% with $P^{-1} \leq 15$, and 91.6% with $P^{-1} \leq 30$. When the demanded minimum yield was 1% of the TMY, the percentage was 51.1% with $P^{-1} = 5$ and 76.6% with $P^{-1} \leq 15$. When the demanded minimum yield was 1% of the TMY, the percentage was 25.0% with $P^{-1} = 5$ and 59.7% with $P^{-1} \leq 15$. Each line is shown with a monotone increase because the experiments were conducted for every $5 \leq c \leq P^{-1}$ when $P^{-1} = c$.

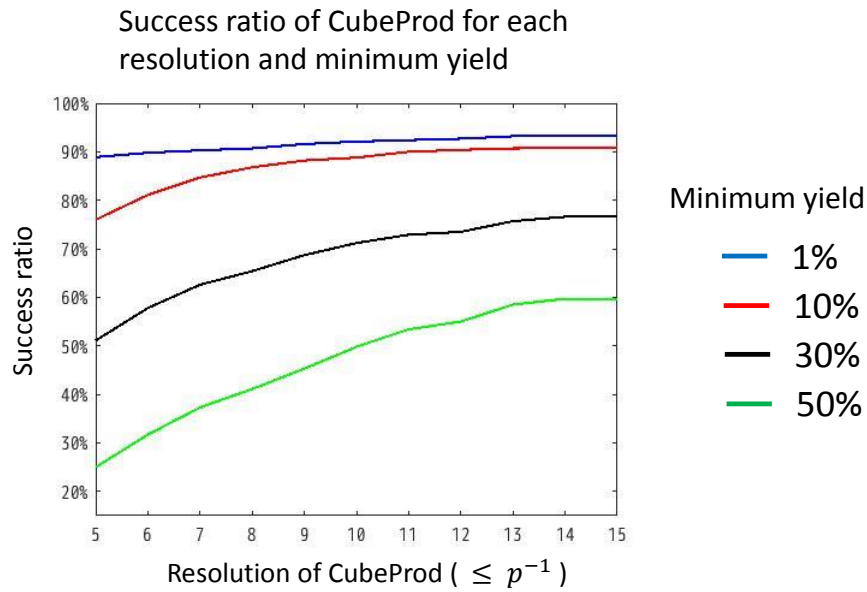

Figure 2. Success ratio of CubeProd in identifying reaction knockouts that could achieve growth coupling for 687 target metabolites of iMM904 at each resolution ($\leq p^{-1}$) and demanded minimum yield level.

Discussion

Fig. 1 shows that CubeProd identified reaction knockouts that can achieve growth coupling for 91.6%, 69.6%, and 47.6% of the 687 target metabolites in iMM904 for the demanded minimum yield levels of 10%, 30%, and 50%, respectively, while the strong coupling-based method could identify the reaction knockouts for 3.9%, 1.7%, and 1.3% of metabolites, respectively [18]. CubeProd substantially improved the results of the strong coupling-based method. Computational experiments using CubeProd showed that growth coupling is possible for most metabolites, even in *S. cerevisiae* under anaerobic conditions.

One reason why existing methods cannot identify reaction knockouts for many target metabolites in some models may be time constraints. Because identifying an optimal subnetwork that achieves the maximum PR under the condition in which GR is maximized is an NP-hard problem, it is often impossible to determine this information for genome-scale models in a realistic time. To overcome this problem, the author previously developed GridProd, which does not guarantee identification of the optimal subnetwork, but can reveal appropriate knockout strategies for many target metabolites for *E.coli* under microaerobic conditions [16]. However, GridProd was not efficient for *S.cerevisiae* under anaerobic conditions in our preliminary experiments.

When GridProd could not derive the appropriate reaction knockouts, the flux obtained by the first LP was often quite different from that obtained by the second LP. This is because the optimal solution of LP is not always uniquely determined. One of promising methods for overcoming this problem is to decrease

the number of optimal solutions by adding other constraints for each LP.

To this end, while GridProd imposes the following two constraints

$$\begin{aligned} TMGR \times P \times i &\leq GR \leq TMGR \times P \times (i+1) \\ TMPR \times P \times j &\leq PR \leq TMPR \times P \times (j+1), \end{aligned}$$

CubeProd imposes the following three constraints

$$\begin{aligned} TMGR \times P \times i &\leq GR \leq TMGR \times P \times (i+1) \\ TMPR \times P \times j &\leq PR \leq TMPR \times P \times (j+1), \\ TMTF \times P \times k &\leq PR \leq TMTF \times P \times (k+1), \end{aligned}$$

for all integers $1 \leq i, j, k \leq P^{-1}$, and then minimizes the sum of absolute values of all fluxes.

Finding knockout strategies with minimum sets of genes or reactions to produce valuable metabolites has been an important problem computational biology. Because large amounts of time and effort are required to knock out several genes, a smaller number of knockouts is preferred in knockout strategies. However, the technologies for DNA synthesis are currently being improved [6]. Although the ability to read DNA is better than the ability to write DNA, designing synthetic DNA may soon become important for producing metabolites rather than knocking out genes in the original genome. In this case, shorter DNA is preferable. In contrast to traditional knockout strategies, the number of genes included in the design of synthetic DNA should be as small as possible because of the experimental effort and time required.

List of abbreviation

Flux Balance Analysis (FBA)
Flux Variability Analysis (FVA)
Linear Programming (LP)
Growth Rate (GR)
Production Rate (PR)
Theoretical Maximum Growth Rate (TMGR)
Theoretical Maximum Production Rate (TMPR)
Theoretical Maximum Total Flux (TMTF)
Glucose Uptake Rate (GUR)
Oxygen Uptake Rate (OUR)

Materials and Methods

The pseudo-code of CubeProd is as follows.

Procedure $CubeProd(target, P)$

$TMGR =$ **max** v_{growth}

s.t. $\Sigma \ S_{i,j} \cdot v_j = 0$

$LB_j \leq v_j \leq UB_j$, $v_{glc_uptake} \geq -GUR$

$v_{o2_uptake} \geq -OUR$, $v_{atp_main} \geq NGAM$

$TMPR =$ **max** v_{target}

s.t. $\Sigma \ S_{i,j} \cdot v_j = 0$

$LB_j \leq v_j \leq UB_j$, $v_{glc_uptake} \geq -GUR$

$v_{o2_uptake} \geq -OUR$, $v_{atp_main} \geq NGAM$

$v_{growth} \geq v_{growth}^{min}$

$TMTF =$ **max** $\Sigma \ |v_j|$

s.t. $\Sigma \ S_{i,j} \cdot v_j = 0$

$LB_j \leq v_j \leq UB_j$, $\mathrm{v}_{glc_uptake} \geq -GUR$

$v_{o2_uptake} \geq -OUR$, $v_{atp_main} \geq NGAM$

$v_{growth} \geq v_{growth}^{min}$

for $i = 1$ to P **do**

$biomassLB = TMGR \times P \times (i-1)$

$biomassUB = TMGR \times P \times i$

for $j = 1$ to P **do**

$targetLB = TMPR \times P \times (j-1)$

$targetUB = TMPR \times P \times j$

for $k = 1$ to P **do**

$totalFluxLB = TMTF \times P \times (k-1)$

$totalFluxUB = TMTF \times P \times k$

% The first LP for (i, j).

$R_{not_used}(i, j)$ is such that

min $\Sigma \ t_j$

s.t. $\Sigma \ S_{i,j} \cdot v_j = 0$

$LB_j \leq v_j \leq UB_j$, $-t_j \leq v_j \leq t_j$

$v_{glc_uptake} \geq -GUR$, $v_{o2_uptake} \geq -OUR$

$v_{atp_main} \geq NGAM$, $biomassLB \leq v_{growth} \leq biomassUB$

$targetLB \leq v_{target} \leq targetUB$

$R_{not_used} = \{v_j | v_j < 10^{-5}\}$

if the first LP is not feasible

$R_{not_used}(i, j, k) = \phi$

v_{target} is such that

max v_{growth}

s.t. $\Sigma \ S_{i,j} \cdot v_j = 0$

$LB_j \leq v_j \leq UB_j$ for $\{j | v_j \notin R_{not_used}(i, j, k)\}$

$v_j = 0$ for $\{j | v_j \in R_{not_used}(i, j, k)\}$

$v_{glc_uptake} \geq -GUR$, $v_{o2_uptake} \geq -OUR$

$v_{atp_main} \geq NGAM$

if $v_{growth} \geq v_{growth}^{min}$

$PR(i,j,k) = v_{target}$
else
$PR(i,j,k) = 0$
$(i,j,k) = argmax(PR(i,j,k))$
return $R_{not_used}(i,j,k)$, $PR(i,j,k)$, FVA$min(i,j,k)$

In the above pseudo-code, the TMGR, TMPR, and TMTF are calculated first. $S_{i,j}$ is the stoichiometric matrix. LB_j and UB_j are the lower and upper bounds of v_j, respectively, which represent the flux of the jth reaction.

v_{glc_uptake}, v_{o2_uptake}, and v_{atp_main} are the lower bounds of the uptake rate of glucose (GUR), oxygen uptake rate (OUR), and non-growth-associated APR maintenance requirement (NGAM), respectively. v_{growth}^{min} is the minimum cell growth rate.

In each cube, LP is conducted twice. "*biomassLB*" and "*biomassUB*" represent the lower and upper bounds of GR, respectively. "*targetLB*" and "*targetUB*" represent the lower and upper bounds of PR, respectively. Similarly, "*totalFluxLB*" and "*totalFluxUB*" represent the lower and upper bounds of total flux, respectively. These values are used as the constraints in the first LP. Each cube is represented by the three constraints, "$biomassLB \leq v_{growth} \leq biomassUB$", "$targetLB \leq v_{target} \leq targetUB$", and "$totalFluxLB \leq \Sigma\ v_j \leq totalFluxUB$". $TMPR \times P$, $TMGR \times P$, and $TMTF \times P$, represent length of the first, second and third for each cube, respectively.

In the solution of the first LP, a set of reactions whose fluxes are nearly 0 (less than 10^{-5}) are represented as R_{not_used}, which is used as a set of unused reactions in the second LP. In the second LP, the fluxes of the reactions included in R_{not_used} were forced to be 0. If the obtained PR is more than or equal to v_{growth}^{min} in the solution of the second LP, the value of PR is stored to $PR(i,j,k)$. Otherwise 0 is stored. Finally, the (i,j,k) yielding the maximum value in $PR(i,j,k)$ is searched, and the corresponding $R_{not_used}(i,j,k)$ and $PR(i,j,k)$ are obtained. The minimum PR from FVA for $R_{not_used}(i,j,k)$ are also calculated.

Genome-scale metabolic model of *S. cerevisiae*

We used iMM904 as an original mathematical model of metabolic networks [10]. iMM904 is a genome-scale reconstruction of the metabolic network in *S. cerevisiae*. This model includes 1228 internal metabolites, 1577 reactions with 1020 repressible reactions and 557 irrepressible reactions. The upper bound of glucose uptake is 10 mmol $gDW^{-1}h^{-1}$, and ATP maintenance demand is 1 mmol $gDW^{-1}h^{-1}$. To simulate anaerobic conditions, cytochrome *c* oxidase was removed as in [18]. To simulate the production potential of each target metabolite in this model, an exchange reaction was temporarily added to the model as in [18].

References

1. J. Becker and C. Wittmann. Advanced biotechnology: Metabolically engineered cells for the bio-based production of chemicals and fuels, materials, and health-care products. *Angewandte Chemie International Edition*, 54(11):3328–3350, 2015.

2. A. P. Burgard, P. Pharkya, and C. D. Maranas. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and bioengineering*, 84(6):647–657, 2003.

3. D. Egen and D. S. Lun. Truncated branch and bound achieves efficient constraint-based genetic design. *Bioinformatics*, 28(12):1619–1623, 2012.

4. D. Gu, C. Zhang, S. Zhou, L. Wei, and Q. Hua. Idealknock: a framework for efficiently identifying knockout strategies leading to targeted overproduction. *Computational biology and chemistry*, 61:229–237, 2016.

5. J. D. Keasling. Manufacturing molecules through metabolic engineering. *Science*, 330(6009):1355–1358, 2010.

6. S. Kosuri and G. M. Church. Large-scale de novo dna synthesis: technologies and applications. *Nature methods*, 11(5):499–507, 2014.

7. S. Y. Lee and H. U. Kim. Systems strategies for developing industrial microbial strains. *Nature biotechnology*, 33(10):1061, 2015.

8. N. E. Lewis, K. K. Hixson, T. M. Conrad, J. A. Lerman, P. Charusanti, A. D. Polpitiya, J. N. Adkins, G. Schramm, S. O. Purvine, D. Lopez-Ferrer, et al. Omic data from evolved e. coli are consistent with computed optimal growth from genome-scale models. *Molecular systems biology*, 6(1):390, 2010.

9. D. S. Lun, G. Rockwell, N. J. Guido, M. Baym, J. A. Kelner, B. Berger, J. E. Galagan, and G. M. Church. Large-scale identification of genetic design strategies using local search. *molecular systems biology*, 5(1):296, 2009.

10. M. L. Mo, B. Ø. Palsson, and M. J. Herrgård. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC systems biology*, 3(1):37, 2009.

11. S. Ohno, H. Shimizu, and C. Furusawa. Fastpros: screening of reaction knockout strategies for metabolic engineering. *Bioinformatics*, 30(7):981–987, 2014.

12. J. D. Orth, I. Thiele, and B. Ø. Palsson. What is flux balance analysis? *Nature biotechnology*, 28(3):245–248, 2010.

13. K. R. Patil, I. Rocha, J. Förster, and J. Nielsen. Evolutionary programming as a platform for in silico metabolic engineering. *BMC bioinformatics*, 6(1):308, 2005.

14. G. Rockwell, N. J. Guido, and G. M. Church. Redirector: designing cell factories by reconstructing the metabolic objective. *PLoS Comput Biol*, 9(1):e1002882, 2013.

15. J. Schellenberger, R. Que, R. M. Fleming, I. Thiele, J. D. Orth, A. M. Feist, D. C. Zielinski, A. Bordbar, N. E. Lewis, S. Rahmanian, et al. Quantitative prediction of cellular metabolism with constraint-based models: the cobra toolbox v2. 0. *Nature protocols*, 6(9):1290–1307, 2011.

16. T. Tamura. Grid-based computational methods for the design of constraint-based parsimonious chemical reaction networks to simulate metabolite production: Gridprod. *BMC bioinformatics*, 19(1):325, 2018.

17. A. Varma and B. O. Palsson. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type escherichia coli w3110. *Applied and environmental microbiology*, 60(10):3724–3731, 1994.

18. A. von Kamp and S. Klamt. Growth-coupled overproduction is feasible for almost all metabolites in five major production organisms. *Nature communications*, 8:15956, 2017.

19. L. Yang, W. R. Cluett, and R. Mahadevan. Emilio: a fast algorithm for genome-scale strain design. *Metabolic engineering*, 13(3):272–281, 2011.