**Yield optimization and analysis**

*course in Advanced Technologies in Bioscience*

Hao Luo

Supervisor: Expert:

Prof. Jens Nielsen Dr. Yu Chen

* Course code: *FKBT190*
* Course credits: *2.0*
* Department: *Biology and Biological Engineering*
* Graduate school: *Bioscience*

A drawing of a cartoon character

Description automatically generatedA sign in the dark

Description automatically generated

Contents

[Abstract 2](#_Toc106883510)

[Yield introduction versus Rate 3](#_Toc106883511)

[Mathematical description of yield and rate optimization 4](#_Toc106883512)

[Current yield optimization methods 5](#_Toc106883513)

[Fixed substrate rate 5](#_Toc106883514)

[Linear fractional programming 6](#_Toc106883515)

[Elementary flux modes/vector (EFM/EFV) 6](#_Toc106883516)

[Opt-yield FBA 8](#_Toc106883517)

[The difference between the yield- and rate- optimization 9](#_Toc106883518)

[Toy model and conditions 9](#_Toc106883519)

[Yield changes when optimizing rate. 12](#_Toc106883520)

[Summary 15](#_Toc106883521)

[Some related topics 15](#_Toc106883522)

[Product envelope and Yield space 15](#_Toc106883523)

[Yield vs. Rate in financial investment 16](#_Toc106883524)

[Codes 17](#_Toc106883525)

[References 17](#_Toc106883526)

# Abstract

Yield and rate both are critical features in metabolic engineering to evaluate the productive and effect. Their concepts and differences have been discussed in many studies. For constraints-based models, the optimal rate is sometimes used to represent the optimal yield because yield optimization is a linear fractional problem and cannot be solved by linear programming straightforwardly. However, the differences of yield-optimal and rate-optimal should be clarified. In this report, we introduced current optimal yield methods and compared rate-optimal and yield-optimal. After clarifying their definitions, we illustrate their differences by example models. Finally, we provide production envelopes and yield spaces as extension comparisons. All the examples and codes can be found on GitHub (https://github.com/HaoLuoChalmers/py\_yield\_ATB).

# Yield introduction versus Rate

Table 1. Comparison of rate and yield.

|  |  |  |
| --- | --- | --- |
|  | Rate | Yield |
| Description | Productivity/speed | Efficiency of conversions |
| Units | mmol/gDW/**h**  per unit of time | One (g/g or mmol/mmol)  per amount of substrate consumed |
| Mathematical |  |  |
| Optimization problem | Linear program | Linear-fractional program |
| Optimization methods | FBA | opt-yield, FBA\*, EFM/EFV+, |

FBA\*, fixed substrate uptake rate (fixed )

EFM/EFV+, not optimization methods (strictly), return a set of pathways.

In production processes, rate and yield are key characteristics for evaluating industrial strains and they focus on productivity and efficiency with different criteria. Rate measures the speed of product formed and concerned with per unit of time. Yield is a relative value to measure the efficiency of conversions and more concerned with the amount of substrate consumed. For example, the biomass rate and yield both indicate the amount of gram dry weight of biomass formed, but rate is measured in unit of time (hour) and yield is measured in unit of amount of substrate consumed (g). Rate and yield are not independent of each other, and their relationship usually is a trade-off such as the ATP rate and ATP yield in yeast respiration and fermentation. More respiration generates ATP faster to support proliferation but with low yields [1]. Oppositely, fermentation have higher ATP yields but with low rate.

In genome-scale metabolic models (GEMs), rate- or yield- optimization are useful methods and can help us design strains according to the metabolic flux distributions. Flux-balance analysis (FBA) is a fundamental method for optimizing target reactions or predicting flux distributions in a steady state. The objective function of FBA usually is the growth rate or product rate, and the optimization could be solved by a linear programming. FBA is clearly maximizing rates in mathematics and some studies also tried to use FBA to solve it by fixing the substrate uptake rate. For example, by fixing substrate uptake rates as experimentally observed values or normalizing substrate uptake as one, optimizing target rate is equal to optimizing yield. And there are also other methods to find optimal yield by EFM/EFV or linear fractional programing.

In this report, we discuss only the forces on the rate- or yield- optimization in dry laboratory GEMs and clarify their differences by example. We also introduced current methods of yield optimization and their both advantage and disadvantages.

# Mathematical description of yield and rate optimization

A GEM usually contains all the biochemical reactions of interested organisms. Based on the coefficients of reactions and metabolites, a GEM is represented by its stoichiometric matrix S for mathematical calculation. The flux distribution through the model is represented by a rates/ flux vector of reaction which contains all the reactions rate through a GEM.

For rate optimization, the objective function is a liner combination of reactions flux like in Eq 1. Where denotes combination coefficients vector and usually contains many zeros, When the target reaction is the growth or product exchange reaction, the objective function could be written as and . FBA is a standard linear programing as Eq 1, subject to steady-state (Eq 2) and with reaction bounds (Eq 3).

Yield is a ratio of two fluxes, the produce flux of product divided by the substrate uptake flux like Eq 4. Where and are coefficients vectors for target productions and substrates, the yield formation will be written as or in the following text for specific production and substrate. Here the directions exchange reactions of product and substrate are assumed to be positive and only concern the absolute values, and . More strictly constrains for substrate because the denominator cannot be zero and yield is no mean without substrate uptake. The yield optimization is a simplified fractional linear programing Eq 5 and subject to the same constrains with FBA.

# Current yield optimization methods

## Fixed substrate rate

Some studies use FBA to optimize yield by fix the substrate uptake rate as experimental value or one. As we mentioned above , if the set as a positive constant (experimental value) then maximize Y is equal to maximize [2]. For the substrate uptake is fixed as 1, the yield is equal to product rate ( ), this method has many applications and discussed by many studies as ‘pathway normalized to the glucose uptake flux’[3][4].

The advantage of this method is it can be performed easily by FBA and the limitation is difficult to deal with unknown substrate uptake rate and related constraints presence. The return result from this method totally related with the fixed substrate value. For pathways normalization, all constraints in the model should be normalized and it is a challenge especially for ec-GEMs (many enzyme constraints)

## Linear fractional programming

Some studies treat yield optimization as simplified linear-fractional programming (LFP) and provide mathematical proofs. In the process of mathematical derivation, the LFP could be transformed into a linear problem by Charnes-Cooper transformation [5]. Under the assumption that the feasible region is non-empty and bounded, and are defined to replace as Eq8 and Eq9.

Eq 4-7 equivalent linear program:

Then the solution for and yields the solution of the original problem as:

More mathematical treatment and duality principle could be fund in Steffen’s study[2]. CellNetAnalyzer toolbox on MATLAB support yield linear fractional programming of yield [6].

## Elementary flux modes/vector (EFM/EFV)

EFMs are a set of topologically feasible, non-decomposable pathways that are underling steady state in models. This set of pathways fully characterizes the available metabolic space of a GEM. EFMs are usually used as a structural concept to identify interested targets robustness of models and a non-negative linear superposition of EFMs could express any feasible results without constraints.

EFV is similar with EFM but used to deal with inhomogeneous constraints and can be considered as catalogs of FBA pathways [7]. With inhomogeneous constraints existed model, EFV could identified disserved pathway undern complex conditions.

Both EFM and EFV are used to find optimal yield or yield space and they performed very well for medium size model. The mathematical problem is similar with the enumeration of corners and edges in a polyhedron. Some tools like efmtool [8] that based on double description method (DDM) approaches that require a huge amount of random-access memory and resources. Currently, both EFMs and EFVs limited to small-size or medium-size networks. Some other approaches like the lexicographic reverse search (lrs) based mothed have made some progress but still no many applications. MATLAB toolbox efmtool and Metatool [9], python packages efmlrs [10] (recommended, compatible with cobrapy) and efmtool (python) support EFM and EFV caculations.

Due to computational power limitations, the decomposition of GEMs into EFMs is still a challenge and the number of EFMs grows exponentially with the number of reactions in a network.

## Opt-yield FBA



Figure 1. Opt-yield FBA mathematical description and pseudo codes and illustrations. a) The objective function is a linear form with a temp value of . The initial and the object is looking for a larger until find the maximum yield. All the process subject to steady-state () capacity and irreversibility (), and substrate should be absorbed (). b) pseudo codes of opt-yield FBA. c) objective functions of meaning for yield optimization.

Opt-yield FBA is an FBA based iterative method and each iteration will return the maximum of until find the maximum yield. This method developed by us, and the code can be found in Sysbio GitHub repositories (<https://github.com/SysBioChalmers/GEM2CB_model/blob/master/Code/GEM2pathways.py> )

# The difference between the yield- and rate- optimization

## Toy model and conditions



Figure 2. Overview of the toy model and optimization conditions. The toy model contains two internal reactions: R1 (in orange arrows or text) and R2 (in green arrows or text); three internal metabolites: substrate S, production B and P. Both rate and yield of B and P will be optimized under four conditions.

To study the rate and yield optmization, a very simple model is introduced. As showen in Figure 2 the toy model contains two internal reactions (R1, R2) and three exchange reactions (Rs, Rb and Rp). The exchange reactions are used to detect the uptake/output rate of three internal metabolites S, B and P. Where S represent the subastrate and B and P represent two products. The model is use to find out the optimal rate and yiedl of B and P under 4 sudo conditions/ scenarios.

* Condition 1, a initial free condition. There are no extra constraints, only subject to steady-state and reaction reversibility.
* Condition 2, a regular condition to limit substrate uptake rate. With this constraint the result will not be infinite.
* Condition 3, add a extra constraint for R1. The R1 flux is limited under 5.
  + Actually, under this condition, we asummed the R1 represent aerobic respiration pathway and R2 represent anaerobic fermentation pathway. R1 usually will affect by oxygen and we limited the oxygen flux under 5. The R2 could produce B (Biomass) and P (could be asummed as ethanol or others ). In R1, all resoure are used to contribuate growth, and in R2 resoure are used to contribuate both product and product.
* Condition 4, add a extra constraint for R2. The R2 flux is limited upper 1.
  + Under this condition, we asummed the R2 represent ATPM, the P asummed as maintain enerage (ATPm). The strain need separate resoucre to gengertae ATP to maintain growth.

Before caculations, we need to note that yield is a property of a pathway and it is not affected by how much fluxes through that pathway. For example the B yields through R1 or R2 is setteled, the B yield via R1 is 1 and via R2 is 0.5. So in this model, the B yield range is from 0.5 to one, and middles yields come from a combination R1 and R2.

Table 2. The theoretical results of rate-optmial and yield-optmial of B and P.



Since there are only two internal reactions/ pathways, all optimizations and flux distributions can be caculated by theoretical derivation. First let us foces on product B as the object. R1 contribute all resourse to produce B and R2 separate resourse to P, so R1 is priority when optimazing B rate or yield. Under Condition 1 and 2, there are no constrans for R1 and R2, R1 takes over all fluxes. Under Condition 3 and 4, when R1 overload, R2 takes over the rest fluxes. For product P, the optimization is simple because R2 has no competing paths. All optimizations are shown in Table 2.

## Yield changes when optimizing rate.



Figure 3. Relationships of - and - when optimizing B rate under Condition 2-3, The Orange background indicate the fluxes through R1, and grey background indicate the fluxes through both R1 and R2. a) under condition 2, the yield is a constant when and increase and fluxes through R1. b) under condition 3, the yield will decrease when the is bounded and fluxes distributions change. c) under condition 4, the yield increases nonlinearly with and fluxes through R1&R2.

To find out whether rate-optimal results can represent yield-optimal, we first optimize B rate at different S uptake rate by FBA and the relationships of - and - are shown in Figure 3. The maximum B rate under Condition 1 is infinite but the trend of yield shoud the same as Condition 2 (Figure 3a). The yield is a constant because the R1 take all the fluxes and the yield from R1 is one. For Condition 2 (Figure 3b), maximum B yield is still one via R1, but in the final statues, the maximum B rate will via both R1 and R2. The rate-optimal cannot represent yield-optimal under Condition 3. For Condition 4, when R2 flux does not meet its lower bound the optimization is infeasible and the is not start from zero points. The yield increases nonlinearly with but reach the maximum when rate maximized.



Figure 4. Relationships of - and - when optimizing B rate under Condition 2-3, The green background indicate the fluxes through R2. a-c) under condition 2-4, the yield is a constant and increases nonlinearly with , all fluxes through 2.

For product P, because only R2 can produce P and none of the conditions are limiting R2 lower bound, all the fluxes through R2 and P rate increases nonlinearly with substrate. As we mentioned yield is a a property of single pathway and P yield from R2 is 0.5 under all conditions.



Figure 5. Relationship of product rate and yield with substrate rate in *E. coli* core model. a-b) Distributions of - and -, when optimizing biomass rate; c-d) Distributions and - and - when optimizing acetate rate.

We also tested the rate and yield on a *E. coli* core model. We limited the glucose and oxygen uptake rate to simulate a practical fermentation environment, the ATPM lower bound of initial model is kept. The biomass rate and acetate rate are optimized at different glucose uptake rate (Figure 5). due to the constraints, rate-rate relationships have some change points (Figure 5a, 5c) and yield-rate relationships are more variable (Figure 5b,5d). The yield maximum of biomass and acetate occurred in the middle of glucose uptake range and rate maximum occurred maximum glucose uptake rate.

## Summary

The relationship between rate optimization and yield optimization is uncertain and influenced by constraints. Sometimes rate optimization can represent yield optimization like Figure 3c, 3d and Figure 4a-c, but rate optimization is totally different from yield optimization like Figure 3b and Figure 5a-d.

# Some related topics

## Product envelope and Yield space



Figure 6. Production envelops and yield space of P-B in toy model.



Figure 7. Production envelops and yield space of acetate-biomass in E. coli core GEM and the location of acetate and biomass rate optimal pathways.

Both rate and yield optimization could be considered as looking for the optimal pathways form a bounded polyhedron. If we force on two targets, the solution space could be present as 2-dimensional production envelops or yield space. The production envelops and yield space could help us to understand the available rate and yield. From Figure 6, the difference of rate and yield solution spaces can be highlighted. The yield is a property of a pathway and not as flaxbility as fluxes in our toy modle. In Figure 7, we mapped acetate and biomass rate optimal pathways to yield space by flux distributions and it is clearly that the rate optimal path can not represent the yield optimal pathways. Note: the rate optimal pathways are not unique and Figure 7 showing the Cobrapy returend FBA resuts.

## Yield vs. Rate in financial investment

There are similar terms about yield and rate in financial and meture caculation methods. Basically, yield forces on the return from an investment and not concern the time as much as rate. Even there are many differences in our biological model, the financial investment methds are good resourse for biologists.

# Codes

The case study is performed by the Python program language and models produces using Cobrapy[11]. All the Python Functions and the Jupyter version of code can be found on GitHub <https://github.com/HaoLuoChalmers/py_yield_ATB> .

# References

1. Pfeiffer T, Schuster S, Bonhoeffer S. Cooperation and competition in the evolution of ATP-producing pathways. Science (80- ). 2001;292:504–7. doi:10.1126/SCIENCE.1058079/SUPPL\_FILE/EQ7.GIF.

2. Klamt S, Müller S, Regensburger G, Zanghellini J. A mathematical framework for yield (vs. rate) optimization in constraint-based modeling and applications in metabolic engineering. Metab Eng. 2018;47:153-169results.

3. Liao JC, Oh MK. Toward Predicting Metabolic Fluxes in Metabolically Engineered Strains. Metab Eng. 1999;1:214–23.

4. Schuster S, Pfeiffer T, Fell DA. Is maximization of molar yield in metabolic networks favoured by evolution? J Theor Biol. 2008;252:497–504.

5. Charnes A, Cooper WW. Programming with linear fractional functionals. Nav Res Logist Q. 1962;9:181–6. doi:10.1002/NAV.3800090303.

6. von Kamp A, Thiele S, Hädicke O, Klamt S. Use of CellNetAnalyzer in biotechnology and metabolic engineering. J Biotechnol. 2017;261:221–8.

7. Klamt S, Regensburger G, Gerstl MP, Jungreuthmayer C, Schuster S, Mahadevan R, et al. From elementary flux modes to elementary flux vectors: Metabolic pathway analysis with arbitrary linear flux constraints. PLOS Comput Biol. 2017;13:e1005409. https://dx.plos.org/10.1371/journal.pcbi.1005409. Accessed 5 Feb 2021.

8. Terzer M, Stelling J. Large-scale computation of elementary flux modes with bit pattern trees. Bioinformatics. 2008;24:2229–35. doi:10.1093/bioinformatics/btn401.

9. von Kamp A, Schuster S. Metatool 5.0: Fast and flexible elementary modes analysis. Bioinformatics. 2006;22:1930–1.

10. Buchner BA, Zanghellini J. EFMlrs: a Python package for elementary flux mode enumeration via lexicographic reverse search. BMC Bioinformatics. 2021;22:1–21. doi:10.1186/S12859-021-04417-9/FIGURES/9.

11. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Syst Biol. 2013;7:1–6. doi:10.1186/1752-0509-7-74/FIGURES/2.