**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](#_Toc106972414)

[Yield introduction versus Rate 3](#_Toc106972415)

[Mathematical description of yield and rate optimization 4](#_Toc106972416)

[Current yield optimization methods 5](#_Toc106972417)

[Fixed substrate rate 5](#_Toc106972418)

[Linear fractional programming 6](#_Toc106972419)

[Elementary flux modes/vector (EFM/EFV) 6](#_Toc106972420)

[Opt-yield FBA 8](#_Toc106972421)

[The difference between the yield- and rate- optimization 9](#_Toc106972422)

[Toy model and conditions 9](#_Toc106972423)

[Yield changes when optimizing rate. 12](#_Toc106972424)

[Summary 15](#_Toc106972425)

[Production envelope and Yield space 15](#_Toc106972426)

[Yield vs. Rate in financial investment 17](#_Toc106972427)

[Codes 17](#_Toc106972428)

[References 17](#_Toc106972429)

# Abstract

Both yield and rate are critical features to evaluate productivity and efficiency in metabolic engineering. 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 between yield-optimal and rate-optimal should be clarified. In this report, we introduce the current optimal yield methods and compare 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 concerns with per unit of time. Yield is a relative value to measure the efficiency of conversions and more concerns 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 the rate is measured in a unit of time (hour) and yield is measured in a unit of amount of substrate consumed (g). Rate and yield are not independent of each other, and their relationship is usually 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 has a higher ATP yield but with a lower 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 linear programming. FBA is clearly maximizing rates in mathematics, and it is also used tried to solve yield problems by fixing the substrate uptake rate in some studies. For example, by fixing substrate uptake rates as experimentally observed values or normalizing substrate uptake as one, optimizing the target rate is equal to optimizing yield. And there are also other methods to find optimal yield by EFM/EFV or linear fractional programming.

In this report, we only forces on the rate- or yield- optimization in dry laboratory GEMs and clarify their differences by example models. 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 the interested organism. 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 rate/flux vector of reaction which contains all the reactions rate through a GEM.

For rate optimization, the objective function is a linear combination of reaction fluxes such as 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 form of 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 or only concern the absolute values, and . More strictly constraints for substrate because the denominator cannot be zero and yield is no mean without substrate uptake. The yield optimization is a simplified fractional linear programming Eq 5 and is subject to the same constraints with FBA.

# Current yield optimization methods

## Fixed substrate rate

Some studies use FBA to optimize yield by fixing the substrate uptake rate as an 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 the 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 that it can be performed easily by FBA and the limitation is that it is difficult to deal with unknown substrate uptake rates and the presence of related constraints. The return result from this method is related to 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 a mathematical proof. In the process of the mathematical derivation, the LFP could be transformed into a linear problem by the 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 principles could be found in Steffen's study [2]. CellNetAnalyzer toolbox on MATLAB supports 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 underlying 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 to EFM but is used to deal with inhomogeneous constraints and can be considered as a catalogue of FBA pathways [7]. With inhomogeneous constraints existing model, EFV could identify disserved pathway under complex conditions.

Both EFM and EFV are used to find optimal yield or yield space and they performed very well for medium size models. 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 calculations.

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 finding the maximum yield. All the processes subject to steady-state () capacity and irreversibility (), and the 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 finding the maximum yield. This method was 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 optimization, a very simple model is introduced. As shown 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 substrate and B and P represent two products. The model is used to find out the optimal rate and yield of B and P under 4 pseudo conditions/ scenarios.

* Condition 1, an 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 an extra constraint for R1. The R1 flux is limited under 5.
  + Actually, under this condition, we assumed that R1 represents the aerobic respiration pathway and R2 represents the anaerobic fermentation pathway. R1 is usually will affect by oxygen and we limited the oxygen flux under 5. The R2 could produce B (Biomass) and P (which could be assumed as ethanol or others ). In R1, all resources are used to contribute to growth, and in R2 resources are used to contribute both product and product.
* Condition 4, add an extra constraint for R2. The R2 flux is limited to upper 1.
  + Under this condition, we assumed the R2 represent ATPM, the P assumed as maintain energy (ATPm). The strain needs separate resources to generate ATP to maintain growth.

Before calculations, 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 and R2 are settled, 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 of R1 and R2.

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



Since there are only two internal reactions/ pathways, all optimizations and flux distributions can be calculated by theoretical derivation. First, let us focus on product B as the object. R1 contribute all resources to produce B and R2 separates resource to P, so R1 is a priority when optimizing B rate or yield. Under Condition 1 and 2, there are no constraints for R1 and R2, R1 takes over all fluxes. Under Condition 3 and 4, when R1 overloads, 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 the B rate under Condition 2-3, The Orange background indicates the fluxes through R1, and the grey background indicates 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 the B rate at different S uptake rates 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 should be the same as Condition 2 (Figure 3a). The yield is a constant because the R1 takes all the fluxes and the yield from R1 is one. For Condition 2 (Figure 3b), the maximum B yield is still one via R1, but in the final statutes, the maximum B rate will be via both R1 and R2. The rate-optimal cannot represent yield-optimal under Condition 3. For Condition 4, when R2 flux does not meet the lower bound, the optimization is infeasible and the is not starting from zero points. The yield increases nonlinearly with but reach the maximum when the rate is maximized.



Figure 4. Relationships of - and - when optimizing B rate under Condition 2-3, The green background indicates 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 is limiting R2 lower bound, all the fluxes through R2 and P rate increase nonlinearly with the substrate. As we mentioned yield is a property of a 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 the core *E. coli* 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 an *E. coli* core model. We limited the glucose and oxygen uptake rate to simulate a practical fermentation environment, and the ATPM lower bound of the initial model is kept. The biomass rate and acetate rate are optimized at different glucose uptake rates (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 is reached in the middle of the glucose uptake range and the maximum rate is reached when the maximum glucose uptake rate.

## Summary

The relationship between rate optimization and yield optimization is uncertain and is influenced by constraints. Sometimes rate optimization can represent yield optimization like Figure 3c, 3d and Figure 4a-c, but rate optimization is clearly different from yield optimization like Figure 3b and Figure 5a-d. For general GEMs with many constraints, yield optimization should be calculated separately from rate optimization.

# Production envelope and Yield space



Figure 6. Production envelopes and yield spaces of P-B in toy model. Because there are only two internal pathways (R1 R2), the yield spaces are lines. Orange points indicate the flux only through R1 and green points indicate only through R2.



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

In geometry, all feasible reaction flux vectors form a bounded polyhedron and both rate and yield optimization could be considered as searching the optimal pathways in it. If we focus on two target products variability, the solution space could be present as 2-dimensional production envelopes or yield space. The production envelope and yield space could help us to understand two products rate and yield relationships. 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 flexible as fluxes in our toy modle. In Figure 7, we mapped acetate and biomass rate optimal pathways to yield space based on flux distributions and it is clear that the rate optimal pathway can not represent the yield optimal pathway. The rate-optimal pathway is not unique (underdetermined system) and the yield maybe not be unique either, the Figure 7 only shows the results returned from the Cobrapy FBA function.

## Yield vs. Rate in financial investment

There are similar terms for yield and rate in financial models and mature optimization methods. Their yield concept forces the return from an investment and typically used to evaluate long-term investments. Even though financial models have many differences with ours, their optimization methods are a good resource for biologists.

# Codes

Model manipulations and plotting were performed in the Python programming language with cobrapy package [11]. All the codes with Jupyter notebooks are available 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.