

Metabolic trade-offs arising from increased free energy conservation in *Saccharomyces cerevisiae*

Schumacher, Robin

DOI

[10.4233/uuid:177e9f4c-f847-436d-9fd4-9ed97ba709d9](https://doi.org/10.4233/uuid:177e9f4c-f847-436d-9fd4-9ed97ba709d9)

Publication date

2018

Document Version

Final published version

Citation (APA)

Schumacher, R. (2018). *Metabolic trade-offs arising from increased free energy conservation in Saccharomyces cerevisiae*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:177e9f4c-f847-436d-9fd4-9ed97ba709d9>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

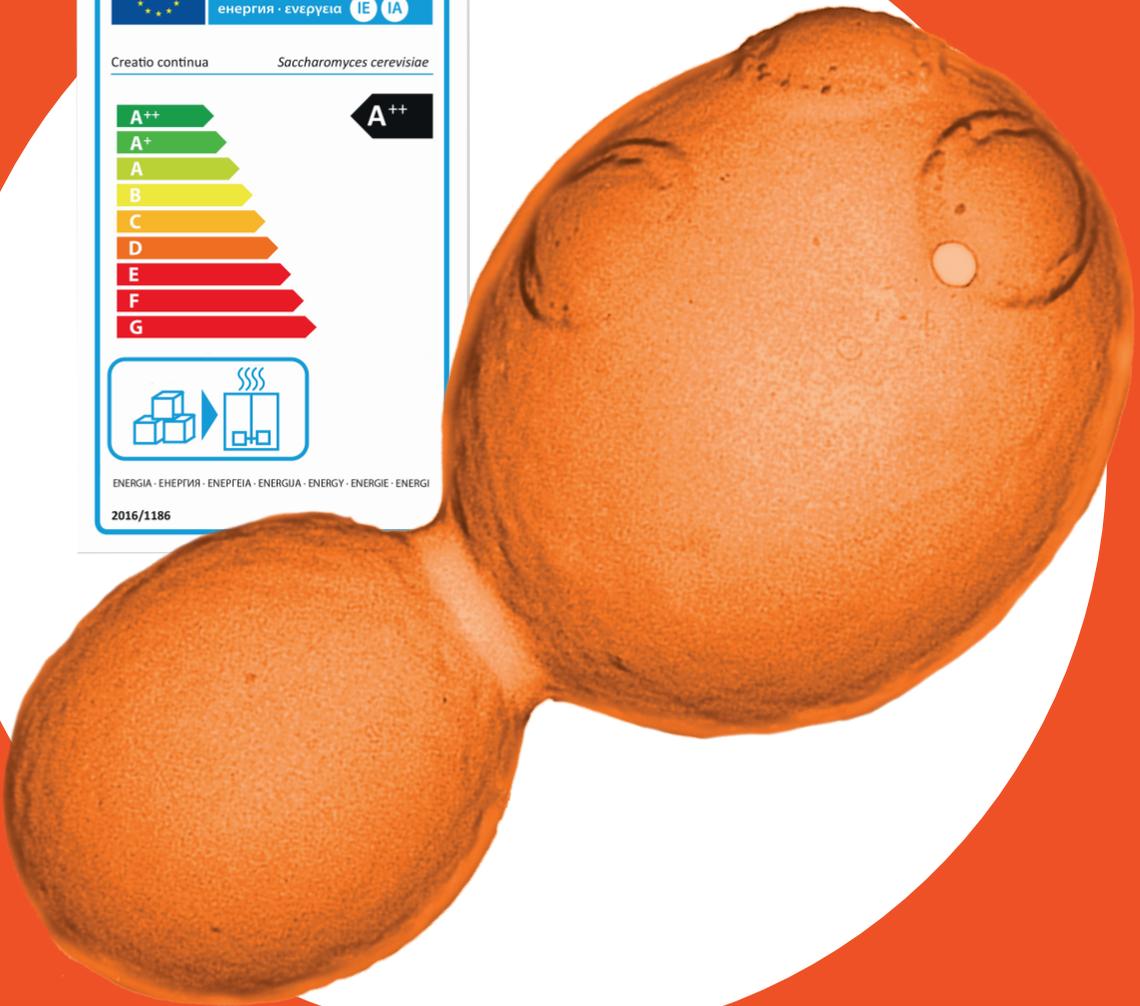
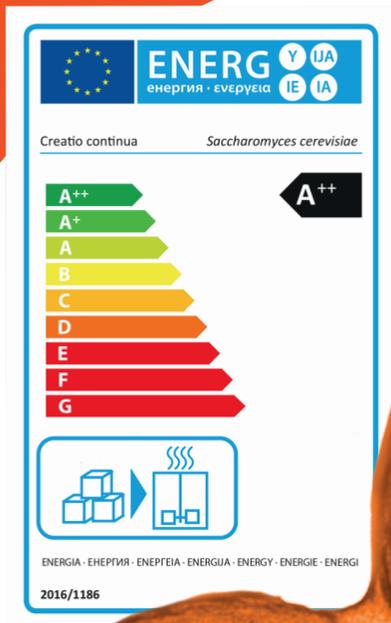
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

Metabolic trade-offs arising from increased free energy conservation in *Saccharomyces cerevisiae*

Robin Schumacher



Metabolic trade-offs arising from
increased free energy conservation
in *Saccharomyces cerevisiae*

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, Prof.dr.ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on
Tuesday 16, October 2018 at 12:30 o'clock

by

Robin SCHUMACHER
Diplom-Ingenieur im Bioingenieurwesen, TU Braunschweig, Germany
born in Albstadt, Germany

This dissertation has been approved by the promotor.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof. dr. ir. J. J. Heijnen	Delft University of Technology, promotor
Dr. S. A. Wahl	Delft University of Technology, copromotor

Independent members:

Prof. dr. I. Van Bogaert	Ghent University, Belgium
Prof. dr. H.V. Westerhoff	University of Manchester, UK / University of Amsterdam
Prof. dr. H.J. Noorman	Delft University of Technology
Prof. dr. J.T. Pronk	Delft University of Technology
Dr. H. Bachmann	VU Amsterdam

The research work was carried out in the Cell Systems Engineering group (CSE), Department of Biotechnology, Faculty of Applied Sciences, Delft University of Technology (The Netherlands).

This work was financed within the BE-Basic R&D Program (<http://www.be-basic.org/>), which was granted an FES subsidy from the Dutch Ministry of Economic Affairs, Agriculture and Innovation (EL & I).

Copyright ©2018 Robin Schumacher

ISBN 978-94-6375-156-8

*"Grey, dear friend, is all theory
and green the golden tree of life"*
Johann Wolfgang von Goethe, Faust I / Mephistopheles



Table of contents

Summary	7
Samenvatting	10
Chapter 1: General Introduction	13
Chapter 2: Modelling the physiology of <i>S. cerevisiae</i> as a function of pH and lactic acid concentration: Implications for the design of a direct lactic acid fermentation process	27
Chapter 3: Enrichment of the more efficient: Droplet based cultivation of <i>S. cerevisiae</i> for selection of phenotypes with increased free energy conservation	61
Chapter 4: Exploring the links between energy metabolism, cellular physiology and protein allocation in the evolution of <i>S. cerevisiae</i> using structured mFBA modelling	95
Chapter 5: Effective estimation of dynamic metabolic fluxes Using ¹³ C Labeling and piecewise affine approximation: From theory to practical applicability	141
Chapter 6: General Conclusions and Outlook	173
Curriculum Vitae / List of publications	178
Acknowledgements	179

Summary

This thesis deals with increasing the free energy conservation in chemotropic microorganisms with emphasis on *S. cerevisiae* and investigates a number of different aspects related to industrial fermentation processes.

Chapter 1 outlines the necessity for transduction of Gibbs free energy in metabolism and uses the concept of thermodynamic efficiency in order to derive the interdependency between the thermodynamic efficiency and the yield of an anabolic product pathway on substrate.

The importance of product yield is discussed with respect to the industrial production of commodity chemicals in cell factories and practical examples how increased free energy conservation can be achieved in the metabolic pathways of *S. cerevisiae* are illustrated. Furthermore, the chapter addresses the fundamental trade-off between thermodynamic efficiency and metabolic rate that is observed empirically, in order to line out the challenges associated with classical directed evolution for phenotypes with increased thermodynamic efficiency.

Chapter 2 studies exemplary an anaerobic lactic acid production process; a low pH process would allow extracting the weak-acid lactic acid directly from the fermentation broth, leading to the environmental advantage of mitigating the by-product formation of gypsum present in the traditional process taking place a near neutral pH.

A model is presented that allows describing the physiology of *S. cerevisiae* as a function of the relevant operating conditions of pH and lactic acid titer using a combination of stoichiometric and kinetic modelling. The model is subsequently used to infer the energetic cost as a function of the operating conditions with respect to the available Gibbs free energy from the product pathway.

Moreover the applied methodology allows predicting the possible product yield globally as a function of the operating conditions making the approach a useful tool for quantitative process design and to deduce strain improvement targets for a direct lactic acid fermentation process in *S. cerevisiae*.

With respect to increased free energy conservation, the knowledge to rationally engineer a phenotype with increased free energy conservation is not always available. In such situation a selection protocol in order to select phenotypes with increased free energy conservation is relevant.

Therefore in Chapter 3 a previously described experimental methodology consisting of repetitive incubation of lactic acid bacteria partitioned into droplets is demonstrated for *S. cerevisiae*. This approach circumvents the initially mentioned trade-off between growth rate and growth yield existent in homogenous systems. The experimental protocol is adapted to *S. cerevisiae* and furthermore a model for the experimental design of enrichment experiments is developed. In order to effectively assess the performance of such systems, the ideal performance characteristics of partitioned systems are derived. The model formulation is also extended for the relevant non-idealities observed in the system and subsequently validated experimentally by enriching a model system consisting of a wild-type population and a phenotype with increased free energy conservation. It is shown that the model can predict the experimental enrichment trajectory; this allows deducing the general enrichment characteristics of partitioned systems from the model. Therewith the model allows predicting the performance of the system *a priori* and also facilitates the quantitative design of enrichment experiments by finding appropriate experimental settings.

While Chapter 3 is a top-down approach treating the rate-yield trade-off as black-box, Chapter 4 investigates a bottom-up approach. Herein a method called structured mass constraint Flux Balance Analysis (mFBA) is used, which directly links metabolic fluxes to the mass of cellular protein in the cell. The flux solution space is constrained by extending the method to multiple protein mass constraints that are imposed by the morphology of eukaryotic cells. The model is parametrized to describe the physiology of *S. cerevisiae* with respect to the rate-yield trade-off associated with aerobic fermentation, also called Crabtree effect. This allows deriving evolutionary trajectories subject to the evolutionary pressure from the environment. This is relevant to choose the appropriate selective pressure in evolutionary experiments and also allows exploring the evolutionary origins of the Crabtree effect with respect to the prevalent environmental conditions.

Whereas so far free energy conservation was only considered in steady-state conditions, also the response to dynamic environments has been shown to lead to increased dissipation of Gibbs free energy. This is of interest, especially with respect to large scale fermentations where the non-ideal mixing continuously exposes the cells to gradients. In order to investigate the underlying intrinsic mechanisms, the metabolic fluxes have to be estimated under dynamic conditions.

Chapter 5 presents fundamental research on dynamic ^{13}C metabolic flux analysis. The presented method avoids some of the inherent conceptual disadvantages of the classical methodology of inverse kinetic modelling. However the method leads to a partly non-linear, potentially ill-posed and high dimensional optimization problem. Effective strategies for the solution of the problem are presented using constraints, shape-prescriptive modelling and a robust optimization algorithm. The performance of the method is demonstrated on experimental data and can be used to trace fluxes, e.g. futile cycles that dissipate Gibbs free energy and have an impact on the biomass yield, in dynamic conditions.

Samenvatting

Dit proefschrift betreft de verhoging van het behoud van de vrije energie in vrije energie omzettingen in chemotrofe micro-organismen, in het bijzonder *Saccharomyces cerevisiae* en onderzoekt een aantal verschillende aspecten vanuit industriële fermentatie processen.

Hoofdstuk 1 beschrijft de noodzaak voor omzetting van Gibbs vrije energie in metabolisme en gebruikt het concept van thermodynamische efficiëntie om de relatie af te leiden tussen efficiëntie en de product opbrengst van anabole producten. Het belang van product opbrengst wordt besproken in relatie tot industriële microbiële productie van bulkchemicaliën, en met voorbeelden wordt geïllustreerd hoe verbeterd behoud van vrije energie kan worden bereikt in het metabolisme van *S. cerevisiae*. Tevens wordt in dit hoofdstuk ingegaan op de fundamentele interactie tussen thermodynamische efficiëntie en snelheid die empirisch is vastgesteld, ten einde de problemen te definiëren van de klassieke gerichte evolutie van fenotypen met verhoogde thermodynamische efficiëntie.

In hoofdstuk 2 worden aspecten van een anaeroob melkzuurproductie proces bestudeerd. Met een lage pH in het proces kan het zwakke melkzuur direct uit de fermentatie vloeistof worden geëxtraheerd, met als milieuvoordeel het vermijden van gipsvorming die optreedt in het traditionele proces bij neutrale pH. Een model is ontwikkeld, gebaseerd op stoichiometrie en kinetiek welk de fysiologie beschrijft van *S. cerevisiae* als functie van de procesomstandigheden zoals pH en melkzuurconcentratie. Dit model is gebruikt om de energiekosten te berekenen als functie van procesomstandigheden in relatie tot de beschikbare Gibbs vrije energie. Met dit model kan ook de product opbrengst worden berekend als functie van proces omstandigheden en het is ook bruikbaar voor procesontwerp en voor het ontwerpen van betere stammen ten behoeve van de melkzuur fermentatie met *S. cerevisiae*.

Ten aanzien van verhoging van het behoud van Gibbs vrije energie is niet altijd de kennis aanwezig voor een rationele modificatie van het metabolisme. In zo'n situatie is een selectie protocol relevant voor het

selecteren van fenotypen met verhoogd behoud van Gibbs vrije energie. In hoofdstuk 3 wordt daarvoor een al bekende experimentele aanpak, een herhaalde kweek van melkzuur bacteriën in aparte druppels, toegepast. Deze aanpak vermijdt de al genoemde interactie tussen (groei)snelheid en groei opbrengst in homogene kweeksystemen. Het beschikbare experimentele protocol is aangepast voor *S. cerevisiae* en tevens is er een model ontwikkeld voor het ontwerpen van de selectie experimenten. Om het selectie gedrag van deze systemen in te schatten, zijn met een model de ideale eigenschappen van deze druppel systemen afgeleid. Het model beschrijft ook relevante niet-ideale situaties en is experimenteel getest met selectie experimenten tussen een wild-type en een gewenst fenotype met verhoogde Gibbs energie behoud. Het model voorspelt de experimenteel waargenomen verrijking in het fenotype. Daarmee is het model geschikt om verrijking in druppelsystemen te voorspellen en is het bruikbaar om verrijkingsexperimenten te ontwerpen.

Waar hoofdstuk 3 de snelheid-opbrengst interactie als Black Box van boven beschrijft, gaat hoofdstuk 4 uit van een benadering van onderaf. Hiervoor wordt een methode, genaamd gestructureerde mFBA, gebruikt waarmee metabole fluxen worden gekoppeld aan de eiwit massa in de cel. De flux oplossingsruimte wordt begrensd door deze methode uit te breiden met begrenzingen van meerdere eiwitmassa's in relatie tot de morfologie van eukaryote cellen. Het model wordt geparametriseerd voor de beschrijving van de fysiologie van *S. cerevisiae* ten aanzien van de snelheid-opbrengst relatie in de aerobe fermentatie, ook wel Crabtree effect genaamd. Met dit model kunnen evolutie paden worden afgeleid in relatie tot evolutie druk uit het milieu, wat relevant is voor het kiezen van een geschikte evolutie druk in evolutie experimenten en om de evolutionaire oorsprong van het Crabtree effect te onderzoeken in relatie tot voorkomende milieucondities.

Waar tot nu toe het behoud van vrije energie is beschouwd onder condities van stationaire toestand is bekend dat dynamische procescondities leiden tot meer dissipatie van Gibbs vrije energie. Dit is van belang vooral voor fermentatie op grote schaal, waar niet ideale mengcondities de cellen blootstellen aan dynamische condities. Ten einde de mechanismen die leiden tot hogere dissipatie te kunnen onderzoeken moeten de metabole snelheden onder dynamische condities worden gekwantificeerd. Hoofdstuk 5 betreft

onderzoek op basis van dynamische ^{13}C flux analyse. De voorgestelde methode vermijdt een aantal nadelen van de klassieke aanpak van omgekeerde kinetische modelvorming. Echter de methode leidt tot een gedeeltelijk niet-linear, hoog dimensioneel parameterschattingsprobleem wat zich tevens potentieel slecht gedraagt. Dit probleem kan effectief worden aangepakt met gebruik van begrenzingen, zoals voorgeschreven kinetische vormen en een robuust schattingsalgoritme. De methode wordt succesvol toegepast op dynamische intracellulaire metabolietconcentraties voor dynamische flux schattingen, in het bijzonder "futile cycles" waarmee extra Gibbs energie wordt gedissipeerd, wat een negatieve invloed heeft op de biomassa opbrengst onder dynamische procescondities.

Chapter 1:

General Introduction

"Life is the harnessing of chemical energy in such a way that the energy-harnessing device makes a copy of itself."

Nick Lane et al. (2012)

Life as a non-equilibrium thermodynamic process

As metaphysical and ontological considerations cannot lead to an unequivocal definition for life¹, modern biology usually tries to formulate an exhaustive and exclusive systemic description of life, based on its necessary empirically observable predicates, e.g. the one of Perrett².

“Life is a potentially self-perpetuating open system of linked organic reactions, catalyzed by stepwise and almost isothermally by complex and specific organic catalysts which are themselves produced by the system”.

The sum of biochemical reactions in an organism is called metabolism and has to synthesize all building blocks for the self-replication (*autopoiesis*) of the organism, where a sequence of consecutive reactions is called a metabolic pathway. This work will primarily focus on the energetic aspects of chemotrophic metabolism with emphasis on the eukaryote *S. cerevisiae*, where it emphasizes on the implications of energy on cellular physiology and the design of industrial cell factories.

It is important to note that biochemical reactions are catalyzed by enzymes and transport proteins and the Gibbs free energy determines the direction of a chemical reaction. In order to satisfy the second law of thermodynamics a reaction can only take place in the desired direction if the change in Gibbs free energy is negative at the physiological metabolite concentrations of its educts and products.

$$\Delta_r G < 0 \quad (1.1)$$

The Gibbs free energy is also directly related to the reaction quotient Q (sometimes also referred to as mass-action ratio) of a reaction, which describes in relation to the equilibrium constant K_{eq} how far a reaction operates from the chemical equilibrium. At equilibrium the net rate of the reaction is zero, which means that life is a process that is driven by non-equilibrium conditions.

$$\Delta_r G = \Delta_r G^0 + RT \ln(Q) = RT \ln\left(\frac{Q}{K_{eq}}\right) \quad (1.2)$$

This also means that the Gibbs free energy limits the attainable extracellular product concentrations (titer) from a given reactant concentration in a product pathway, which would be maximal in a completely reversible process. Clearly this situation is purely hypothetical, as such system would have no net-flux and therewith no productivity. On the other hand, the distance of a reaction from its chemical equilibrium directly defines the amount of Gibbs free energy that is not transformed into the highest product concentration but dissipated in creating the reaction rate. This dissipated Gibbs free energy ΔG_{diss} is also referred to as thermodynamic driving force (tdf).

Energy transduction as an integral part of cellular metabolism

Current genome scale metabolic networks for *S. cerevisiae* have about 1175 annotated biochemical reactions³, where the products can be distinguished into endergonic and exergonic products with respect to the educts. As condition (1.1) has to be fulfilled for each reaction in the network that carries a flux, it is imperative to have mechanisms that can transduce Gibbs free energy from exergonic into endergonic reactions. This can for example be observed in the glycolytic *Embden-Meyerhof-Parnas* (EMP) pathway converting glucose to pyruvate; here the overall product reaction is thermodynamically favorable, however the first steps of the pathway are unfavorable, therefore requiring the coupling of free energy between reactions. Two types of coupling can be distinguished, direct and indirect.

An example for direct energetic linkage between reactions is the PTS system found in some prokaryotes; which directly couples the phosphoryl group released in the pyruvate kinase reaction to the substrate level phosphorylation. However, such direct mechanisms are inflexible and the exception. To deal with the complexity arising with the multitude of different reactions, cells have evolved a more flexible indirect coupling mechanism for energy transduction. In order to transduce Gibbs free energy between reactions the surplus from exergonic reactions gets temporarily stored into chemical bonds of a component that can subsequently be used as a coenzyme in the endergonic reactions. Here cells usually use nucleoside phosphates with the most prevalent one being Adenosine phosphates (AXP). The creation and cleavage of the phosphoanhydride bonds allows the transfer of a significant amount of Gibbs free energy and the ratio between the three different forms ATP, ADP and AMP define the so called adenylate energy charge and therewith the energy status of the cell. At physiological energy charge, pH value and Mg^{2+} concentration the hydrolysis of a phosphoanhydride bond yields a potential of about $\Delta G_p = -50\text{kJ/mol}$ of Gibbs free energy. Due to its universality ATP is often referred to as the currency of free energy in the cell.

As the *in vivo* reaction quotients for all cellular reactions are typically not known, the problem has to be simplified to a more abstract one. This can be done by modelling metabolism as a so called free energy converter⁴ (EC) (see Figure 1.1). Here only a catabolic reaction (subscript c) producing a surplus of useful Gibbs free energy (in the form of ATP) and an anabolic reaction requiring the net input of ATP (subscript a) are distinguished.

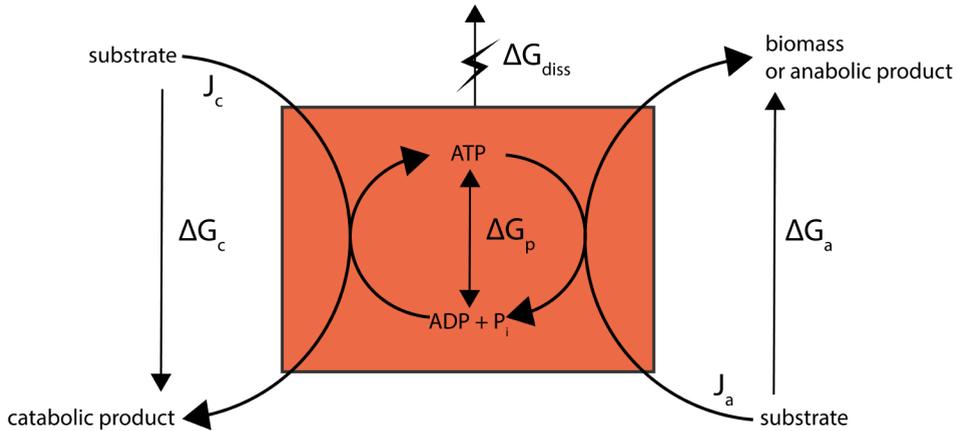


Figure 1.1: Schematic representation of metabolism as a non-ideal energy converter (adapted from Westerhoff⁵) transducing free energy between catabolism and anabolism.

By lumping all biochemical pathways into those elementary reactions, with the respective change in Gibbs free energy ΔG_c and ΔG_a with the associated specific rates J_c and J_a , metabolism can be modelled schematically as a grey box and the amount of free energy that is conserved between catabolism to anabolism can be expressed as a thermodynamic efficiency η_{EC} ; which is defined for growth on a single carbon substrate as follows:

$$\eta_{EC} = \frac{-J_a \cdot \Delta G_a}{J_c \cdot \Delta G_c} \quad (1.3)$$

This shows that there is an intrinsic coupling between the catabolic and the anabolic rate, which depends on the thermodynamic efficiency of the coupling between the two. Also a relation between the dissipated Gibbs free energy ΔG_{diss} (also called thermodynamic driving force) for the overall system can be found.

$$\Delta G_{diss} = \Delta G_c (1 - \eta_{EC}) \quad (1.4)$$

It has to be noted that the apparent thermodynamic efficiency is typically not a direct result of the theoretical ATP stoichiometry between the catabolic and the growth reaction (typically referred to as $Y_{ATP,max}$ ^{6,7}) but only gives an upper boundary on the thermodynamic efficiency. There may be other processes further increasing ΔG_{diss} (therefore called uncoupling) lowering the observed thermodynamic efficiency, e.g. weak acid cycling (see also in Chapter 2), substrate cycles or cellular maintenance requirements.

The biochemically meaningful calculation of thermodynamic efficiencies depends on the biochemically sound separation of the macrochemical growth reaction into a catabolic and an anabolic reaction, which is consequently challenging in practical

application (as also critically reviewed by Heijnen and van Dijken⁸); however the approach is methodologically useful to demonstrate the interrelation between the thermodynamic efficiency and the biomass or (ATP consuming) product yield on substrate:

$$Y_{a/S} = \frac{-J_a}{-J_a + J_c} = \frac{1}{1 + \frac{1}{\eta^{EC}} \cdot \frac{\Delta G_a}{\Delta G_c}} \quad (1.5)$$

It has been shown that it is possible to correlate the biomass yield empirically to the dissipated Gibbs free energy ΔG_{diss} , which can be directly derived from the macrochemical growth equation or measured using calorimetry⁹ and depends mainly on the degree of reduction and the chain length of the substrate^{8,10}. Moreover the ΔG_{diss} can also be correlated empirically to the (anabolic) biomass yield¹¹ or heuristically to the number of irreversible steps in a pathway¹².

Free energy conservation and Industrial Biotechnology

A high product yield on substrate is of major importance in the design of industrial fermentation processes, where a substrate (typically a sugar) is subject to microbial fermentation to a more valuable product. The advantage of such processes is that they can readily utilize renewable feedstocks and therewith have the potential to mitigate anthropogenic climate change by reduction of net CO₂ emissions.

However, this also leads to the situation that these products are usually in economic competition with building blocks derived from fossil feedstock. Especially for comparably low value commodity chemicals the cost for the feedstock has a large impact on the achievable final product price¹³, leading to the implication that the product yield on substrate is a very important design variable for a cell factory¹⁴.

A product can be produced at the (maximal) theoretical yield under two prerequisites: (1) all electrons of the substrate are conserved in the product; which is also a sufficient condition that the product can be produced in an anaerobic fermentation (i.e. without oxygen as electron acceptor); (2) the Gibbs free energy of the product reaction has to be negative (see eq.(1.1)). Therefore it is useful to express the Gibbs free energy per degree of reduction in order to analyze if an anaerobic production of a given product is thermodynamically not impossible, which is indeed the case for many chemicals that are considered relevant for the transition to a bio-based economy^{12,15}.

It has to be noted that such black-box analysis does not make any provisions about the enzymatic and transport steps of the regarded product reaction and only allows deriving the maximal possible ATP yield of a pathway subject to the limits imposed by the second law of thermodynamics, whereas the practically achievable

conservation of free energy as ATP depends on the specific pathway reactions and topology.

From eq. (1.5) it becomes clear that improving the thermodynamic efficiency can improve the yield of an ATP consuming product pathway on substrate or may even (within the thermodynamic constraints) allow transforming a hitherto ATP consuming to an ATP generating product pathway. In this case the relation between thermodynamic efficiency and product yield also changes; as the (now catabolic) product pathway delivers ATP, this will lead to a coupling between the production rate and the growth rate. This means in this case a lower ATP yield will lead to a higher product yield, as a higher production rate is required to generate the energy needed for growth. This also means that further increasing of free energy conservation in an ATP generating product pathway will lead to a lower product yield and is therefore not useful anymore. On the other hand the product pathway in this situation has to provide the ATP required to maintain and synthesize the biomass as the biocatalyst, which means a product pathway with zero ATP gain is also not desirable.

Improving the free energy conservation for net ATP consuming products has the additional effect that there is less oxygen consumption (due to a lower rate of catabolism) and less heat per consumed substrate is released in the fermentation process, therewith synergistically decreasing the utility cost associated with cooling and aeration of the fermentation broth or allow a higher productivity at the same oxygen transfer rate (OTR) in a fermenter.

However, in the design of an industrial fermentation process the ATP stoichiometry of the product pathway is not the only factor that determines the maximal possible product yield, but also the process conditions. Such physiological constraints and their associated ATP expenditure can have a significant impact on the product yield from substrate and have to be taken into account in order to maintain the viability of the host organism in a fermentation process. Those mechanisms are studied exemplary for a direct lactic acid fermentation process in *S. cerevisiae* where the organism is exposed to high lactic acid levels and osmotic stress and evaluate the impact on the attainable product yield quantitatively (see Chapter 2).

Engineering of increased free energy conservation in metabolic pathways

The principle of such engineering is fairly simple in theory:

Replace reactions with a large Gibbs energy dissipation with energy conserving reactions operating closer to the chemical equilibrium.

The most promising targets for such engineering are reactions that carry high rates and are preferably used in catabolism and anabolism. To demonstrate the principle two examples related to growth of *S. cerevisiae* on maltose are presented, which are also used later on in this study (see Chapter 3). More targets can be found in the review of de Kok et al.¹⁶ (see also Figure 1.2).

Interesting targets are active uptake mechanisms as in our example where maltose is taken up together with a proton¹⁷. This allows coupling the maltose import to the proton motive force (pmf) of the cell membrane. To restore the proton motive force and prevent intracellular acidification the imported proton has to be expelled from the cell with an investment of 1 ATP in *S. cerevisiae*¹⁸.

This allows for several approaches to improve the free energy conservation, first the substitution of the symport with a uniporter mechanism decoupling the transport from the pmf and thus ATP expenditure. It has to be taken into account that such engineering has significant influence on the maximal in/out concentration of the transported molecule, which will decrease from about 350 for a proton symporter to 1 for a uniporter¹⁹ thus requiring a higher extracellular substrate concentration to reach the same intracellular substrate concentration. It can also be seen that this strategy is not feasible for product export from the low intracellular concentrations if any substantial extracellular titer shall be produced.

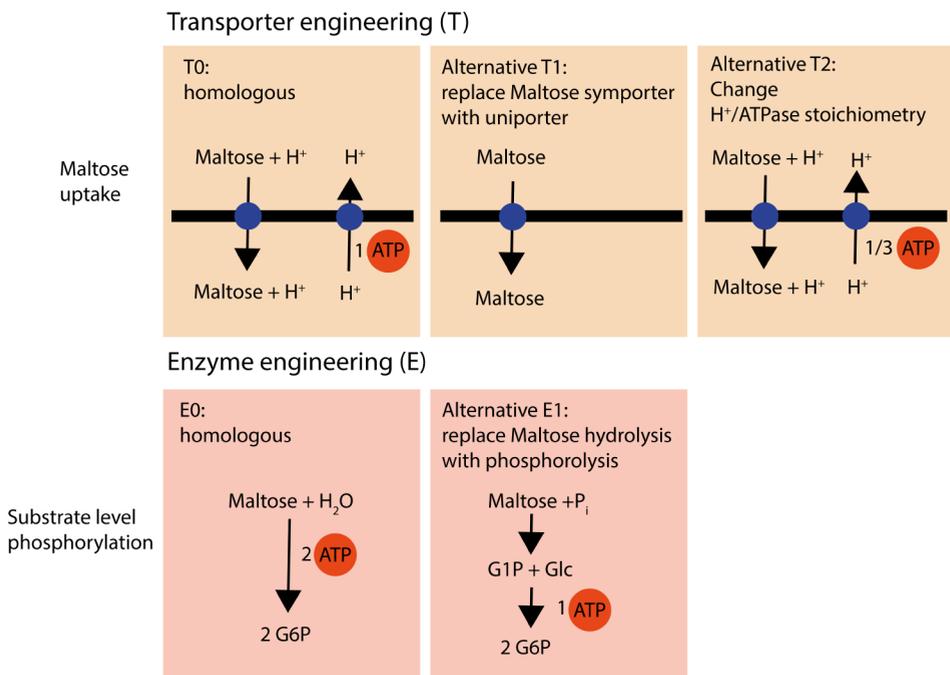


Figure 1.2: Selected opportunities¹⁶ for increasing the ATP yield in maltose grown *S. cerevisiae* on the level of maltose uptake and phosphorylation.

Secondly the proton-ATPase stoichiometry could be improved, i.e. more protons get exported with hydrolysis of one molecule of ATP. Whereas the first option seems more effective at first hand it is specific to the maltose transport, whereas the second option would decrease the ATP requirement for all reactions coupled to the pmf across the cell membrane.

A third option is the replacement of reactions using ATP as coenzyme with reactions using inorganic phosphate (P_i) as a substrate. This can in our example be done by replacing the enzyme maltase (which hydrolyses maltose to two molecules of glucose) with maltose phosphorylase (EC 2.4.1.8), saving one ATP for substrate level phosphorylation of the hexoses. Also, here it has to be taken into account that the increased free energy conservation affects the *in vivo* reaction quotient significantly, leading to changed intracellular concentrations that can also influence (positively or negatively) the rate of other enzymes in metabolism.

Whereas the maltose phosphorylase strategy has been demonstrated successfully²⁰ the rational engineering of a maltose uniporter²¹ is very challenging in practice. The same holds for engineering the stoichiometry of the proton/ATPase¹⁸. Because of these challenges evolutionary approaches are considered useful to engineer phenotypes with increased free energy conservation in this work. When there is a lack of rational knowledge, evolutionary engineering (also called directed evolution or adaptive laboratory evolution) is a versatile tool to obtain a desired phenotype by creating a selective environment, where the desired phenotype having an advantage in growth rate can be enriched and later isolated and reverse engineered.

The relation between dissipated Gibbs free energy and rate

So far the focus was almost exclusively on the relation between thermodynamic efficiency and the catabolic yield, neglecting another fundamental relationship.

This is the inverse relationship between the thermodynamic driving force (ΔG_{diss}) and the rate of a reaction, which can be observed phenomenologically as a trade-off between substrate uptake rate and biomass yield on substrate²²⁻²⁵. The exact shape of this rate-yield trade-off can only be observed empirically as current non-equilibrium thermodynamics do not allow deriving a general relationship *ab initio* from thermodynamic fundamentals, but depend on the complex interaction between the mechanism of the enzymes, their specific kinetic factors and the resulting metabolite concentrations in the pathway^{26,27}. This holds in particular for reactions that are operating far from chemical equilibrium²⁸ and are therefore favorable targets for engineering towards increased free energy conservation.

Under the premise that unicellular organisms are selected by evolution subject to a maximized growth rate^{29,30}, it is expected to find an optimal compromise between substrate uptake rate and biomass yield on substrate maximizing the growth rate. This could explain the typically low thermodynamic efficiencies⁵ found in such systems. Coming back to directed evolution, the existence of such a correlation leads to the challenge that a desired phenotype with an increased thermodynamic efficiency cannot be linked inherently to an increased growth rate. This renders the standard approaches of directed evolution e.g. sequential batch reactors useless for the chosen objective and new experimental approaches have to be considered^{29,31} and validated. To address this challenge this work investigates a top-down approach comprised of partitioning cells into droplets allowing circumventing the rate-yield

trade-off for enrichment of phenotypes with increased free energy conservation (see Chapter 3).

The unknown shape of the rate-yield trade off also impedes the prediction of evolutionary outcomes and the rational design of evolution experiments e.g. choosing the appropriate evolutionary pressure. There is a lack of modelling tools that allow predictions of evolutionary trajectories quantitatively. This work presents a bottom-up approach, where a model that can describe the physiology of yeast cells based on protein constraints derived from the morphology of cells is developed. This allows describing the fluxes as a direct function of cellular protein allocation and is further demonstrated to be a useful tool in understanding the evolutionary history of yeast (see Chapter 4). Understanding those fundamental mechanisms of evolution will help to design better evolutionary experiments and in consequence to create more efficient industrial fermentation processes.

Understanding the link between free energy conservation and dynamic environments

While so far always quasi steady-state conditions were considered and a single optimization criterion (i.e. maximal growth rate) was applied, in nature microorganisms are also exposed to dynamic conditions. Such fluctuations require metabolic flexibility in the pathways to maintain the cellular homeostasis in transient environments. Exposing cells to dynamic condition therefore coincides with an increased dissipation of free energy⁹ and consequently a decrease in biomass yield compared to steady-state conditions^{32,33}.

This is also of major interest in large scale industrial fermentations, where cells are continuously exposed to substrate gradients^{34,35} in the fermenter which will e.g. lead to rapid fluctuations in the substrate uptake rate and effects on the product yield. It is imperative to understand the linkage between energy dissipation, in particular with respect to energy dissipating futile cycles³⁶ (also called substrate cycles), and the underlying regulatory mechanisms^{37,38} contributing to the phenotypical robustness in large scale fermentations. This would also facilitate the scale-up from bench to industrial scale and may even allow to rationally engineering more efficient or more robust phenotypes.

However these fundamental connections are largely not understood due to a lack of direct observables^{39,40} but are also hampered by the unavailability of unbiased methods that allow a quantitative estimation of metabolic fluxes under dynamic conditions. Especially futile cycles can only be observed using labelling techniques⁴¹ as they lead to an underdetermined stoichiometry matrix. Although proven methods are available for flux analysis in metabolic steady-state, useful methods that allow flux estimation in dynamic conditions have to be developed, a challenge that is addressed in this work (see Chapter 5).

Scope of the thesis

In short, this thesis focusses on the following aspects arising from increased free energy conservation.

- (1) The relation between the dissipated Gibbs energy (ΔG_{diss}) and the achievable product yield with respect to the process conditions in a fermentation process.
- (2) The additional relation between the dissipated Gibbs energy (ΔG_{diss}) and the growth rate leading to the rate-yield trade-off is investigated. First with respect to a directed evolution strategies that allowing for selection of phenotypes with increased free energy conservation. Secondly, a bottom-up modelling approach is presented allowing to predict adequate selective pressures.
- (3) The relation between the dissipated Gibbs energy (ΔG_{diss}) and phenotypical robustness under dynamic conditions as prevalent in large-scale industrial fermentations is analyzed. Here fundamental research on dynamic flux estimation is presented that enables identifying flux functions and therewith facilitates understanding of the underlying dissipation mechanisms.

References

- 1 Luisi, P. L. About Various Definitions of Life. *Origins of life and evolution of the biosphere* 28, 613-622 (1998).
- 2 Perret, J. Biochemistry and bacteria. *New Biology* 12, 68-69 (1952).
- 3 Förster, J., Famili, I., Fu, P., Palsson, B. Ø. & Nielsen, J. Genome-Scale Reconstruction of the *Saccharomyces cerevisiae* Metabolic Network. *Genome research* 13, 244-253 (2003).
- 4 Westerhoff, H. V., Hellingwerf, K. J. & Van Dam, K. Thermodynamic efficiency of microbial growth is low but optimal for maximal growth rate. *Proceedings of the National Academy of Sciences of the United States of America* 80, 305-309 (1983).
- 5 Westerhoff, H. V., Hellingwerf, K. J. & Van Dam, K. Thermodynamic efficiency of microbial growth is low but optimal for maximal growth rate. *Proceedings of the National Academy of Sciences of the United States of America* 80, 305-309 (1983).
- 6 Verduyn, C., Stouthamer, A. H., Scheffers, W. A. & Dijken, J. P. A theoretical evaluation of growth yields of yeasts. *Antonie Van Leeuwenhoek* 59 (1991).
- 7 Neidhardt, F. C., Ingraham, J. L. & Schaechter, M. *Physiology of the Bacterial Cell: A Molecular Approach*. (Sinauer Associates, 1990).
- 8 Heijnen, J. J. A new thermodynamically based correlation of chemotrophic biomass yields. *Antonie van Leeuwenhoek* 60, 235-256 (1991).
- 9 Van Kleeff, B. H. A., Kuenen, J. G. & Heijnen, J. J. Heat flux measurements for the fast monitoring of dynamic responses to glucose additions by yeasts that were subjected to different feeding regimes in continuous culture. *Biotechnology Progress* 12, 510-518 (1996).
- 10 Liu, J.-S., Vojinović, V., Patiño, R., Maskow, T. & von Stockar, U. A comparison of various Gibbs energy dissipation correlations for predicting microbial growth yields. *Thermochimica Acta* 458, 38-46 (2007).
- 11 Heijnen, J. J. & Van Dijken, J. P. In search of a thermodynamic description of biomass yields for the chemotrophic growth of microorganisms. *Biotechnology and Bioengineering* 39, 833-858 (1992).
- 12 Cueto-Rojas, H. F., van Maris, A. J., Wahl, S. A. & Heijnen, J. J. Thermodynamics-based design of microbial cell factories for anaerobic product formation. *Trends Biotechnol* 33, 534-546 (2015).
- 13 Straathof, A. J. J. in *Comprehensive Biotechnology (Second Edition)* 811-814 (Academic Press, 2011).
- 14 Straathof, A. J. J. & Bampouli, A. Potential of commodity chemicals to become bio-based according to maximum yields and petrochemical prices. *Biofuels, Bioproducts and Biorefining* (2017).

- 15 Ataman, M. & Hatzimanikatis, V. Heading in the right direction: thermodynamics-based network analysis and pathway engineering. *Current Opinion in Biotechnology* 36, 176-182 (2015).
- 16 de Kok, S., Kozak, B. U., Pronk, J. T. & van Maris, A. J. A. Energy coupling in *Saccharomyces cerevisiae*: selected opportunities for metabolic engineering. *FEMS Yeast Research* 12, 387-397 (2012).
- 17 Weusthuis, R. A., Adams, H., Scheffers, W. A. & van Dijken, J. P. Energetics and kinetics of maltose transport in *Saccharomyces cerevisiae*: a continuous culture study. *Appl Environ Microbiol* 59, 3102-3109 (1993).
- 18 de Kok, S., Yilmaz, D., Daran, J.-M., Pronk, J. T. & van Maris, A. J. A. In vivo analysis of *Saccharomyces cerevisiae* plasma membrane ATPase Pma1p isoforms with increased in vitro H⁺/ATP stoichiometry. *Antonie van Leeuwenhoek* 102, 401-406 (2012).
- 19 van Maris, A. J., Konings, W. N., van Dijken, J. P. & Pronk, J. T. Microbial export of lactic and 3-hydroxypropanoic acid: implications for industrial fermentation processes. *Metab Eng* 6, 245-255 (2004).
- 20 de Kok, S. *et al.* Increasing free-energy (ATP) conservation in maltose-grown *Saccharomyces cerevisiae* by expression of a heterologous maltose phosphorylase. *Metab Eng* 13, 518-526 (2011).
- 21 Henderson, R. & Poolman, B. Proton-solute coupling mechanism of the maltose transporter from *Saccharomyces cerevisiae*. *Scientific reports* 7, 14375 (2017).
- 22 Pfeiffer, T., Schuster, S. & Bonhoeffer, S. Cooperation and competition in the evolution of ATP-producing pathways. *Science* 292, 504-507 (2001).
- 23 Postma, E., Verduyn, C., Scheffers, W. A. & Van Dijken, J. P. Enzymic analysis of the crabtree effect in glucose-limited chemostat cultures of *Saccharomyces cerevisiae*. *Appl Environ Microbiol* 55, 468-477 (1989).
- 24 Maja Novak, Thomas Pfeiffer, Richard E. Lenski, Uwe Sauer & Sebastian Bonhoeffer. Experimental Tests for an Evolutionary Trade-Off between Growth Rate and Yield in *E. coli*. *The American Naturalist* 168, 242-251 (2006).
- 25 MacLean, R. C. The tragedy of the commons in microbial populations: insights from theoretical, comparative and experimental studies. *Heredity* 100, 233-239 (2008).
- 26 Pekar, M. The thermodynamic driving force for kinetics in general and enzyme kinetics in particular. *Chemphyschem : a European journal of chemical physics and physical chemistry* 16, 884-885 (2015).
- 27 Wu, L., Wang, W., van Winden, W. A., van Gulik, W. M. & Heijnen, J. J. A new framework for the estimation of control parameters in metabolic pathways using lin-log kinetics. *European Journal of Biochemistry* 271, 3348-3359 (2004).

- 28 Canelas, A. B., Ras, C., ten Pierick, A., van Gulik, W. M. & Heijnen, J. J. An in vivo data-driven framework for classification and quantification of enzyme kinetics and determination of apparent thermodynamic data. *Metabolic Engineering* 13, 294-306 (2011).
- 29 Bachmann, H. *et al.* Availability of public goods shapes the evolution of competing metabolic strategies. *Proceedings of the National Academy of Sciences* 110, 14302-14307 (2013).
- 30 Molenaar, D., Van Berlo, R., De Ridder, D. & Teusink, B. Shifts in growth strategies reflect tradeoffs in cellular economics. *Molecular systems biology* 5, 323 (2009).
- 31 Bachmann, H., Bruggeman, F. J., Molenaar, D., Branco dos Santos, F. & Teusink, B. Public goods and metabolic strategies. *Current Opinion in Microbiology* 31, 109-115 (2016).
- 32 Suarez-Mendez, C., Ras, C. & Wahl, S. Metabolic adjustment upon repetitive substrate perturbations using dynamic ¹³C-tracing in yeast. *Microbial Cell Factories* 16, 161 (2017).
- 33 Suarez-Mendez, C., Sousa, A., Heijnen, J. & Wahl, A. Fast "Feast/Famine" Cycles for Studying Microbial Physiology Under Dynamic Conditions: A Case Study with *Saccharomyces cerevisiae*. *Metabolites* 4, 347 (2014).
- 34 Delvigne, F. & Noorman, H. Scale-up/Scale-down of microbial bioprocesses: a modern light on an old issue. *Microbial Biotechnology* 10, 685-687 (2017).
- 35 Noorman, H. J. & Heijnen, J. J. Biochemical engineering's grand adventure. *Chemical Engineering Science* 170, 677-693 (2017).
- 36 Zhao, Z., ten Pierick, A., de Jonge, L., Heijnen, J. J. & Wahl, S. A. Substrate cycles in *Penicillium chrysogenum* quantified by isotopic non-stationary flux analysis. *Microbial Cell Factories* 11, 140 (2012).
- 37 Newsholme, E. A., Challiss, R. A. J. & Crabtree, B. Substrate cycles: their role in improving sensitivity in metabolic control. *Trends Biochem Sci* 9 (1984).
- 38 Newsholme, E. A. & Crabtree, B. Substrate cycles in metabolic regulation and in heat generation. *Biochem Soc Symp*, 61-109 (1976).
- 39 Mumby, M. & Brekken, D. Phosphoproteomics: new insights into cellular signaling. *Genome Biology* 6, 230-230 (2005).
- 40 Schulz, J. C., Zampieri, M., Wanka, S., von Mering, C. & Sauer, U. Large-scale functional analysis of the roles of phosphorylation in yeast metabolic pathways. *Science Signaling* 7 (2014).
- 41 Wahl, S., Noh, K. & Wiechert, W. ¹³C labeling experiments at metabolic nonstationary conditions: An exploratory study. *BMC Bioinformatics* 9, 152 (2008).

Chapter 2:

Modelling the physiology of *S. cerevisiae* as a function of pH and lactic acid concentration: Implications for the design of a direct lactic acid fermentation process

in collaboration with S.A. Wahl and J.J Heijnen

“The essence of engineering is to find the best compromises subject to the given constraints”

Abstract

Lactic acid production processes at low pH have been suggested as a more sustainable alternative to bacterial processes at near-neutral pH, as they allow the direct removal of lactic acid from the broth and therewith eliminate the formation of gypsum as byproduct.

In this work the physiology of the industrial workhorse *S. cerevisiae* at high titers of lactic acid and lactate is analyzed. A stoichiometric metabolic model coupled to kinetic mechanisms is developed to describe the relevant phenomena with respect to the cellular physiology in a mechanistic way; (1) energetic uncoupling due to diffusion of lactic acid, acetic acid, succinic acid and protons through the cell membrane and (2) the response to osmotic pressure leading to elevated glycerol and acetate production in anaerobic conditions. For model calibration anaerobic chemostat experiments were performed at a variety of pH and nominal lactic acid concentrations to estimate the apparent permeability coefficients for lactic acid, as well as for glycerol in the mathematical model.

A thermodynamic analysis is conducted to assess whether a low pH direct lactic acid fermentation process could be anaerobic. The analysis shows that the product pathway cannot deliver enough Gibbs free energy to generate ATP at pH values below the pK_a of lactic acid, therewith rendering an anaerobic process at such pH impossible. With the parametrized model the maximal lactic acid yield on substrate is predicted globally for an aerobic direct lactic acid production process in *S. cerevisiae* as a function of the pH, the lactic acid titer and the osmolarity of the broth. It is found that the prevalent effect decreasing the product yield at low pH is the uncoupling caused by back diffusion of lactic acid from the broth whereas at near neutral pH the effect of glycerol production due to osmotic response is predominant. The decrease caused by two effects has approximately the same magnitude leading to an optimal pH at around 2.5. The model also allows identifying metabolic engineering targets on a quantitative basis for improvement of the product yield e.g. reducing the membrane permeability for lactic acid and will be useful in assessing the economic feasibility of a direct weak-acid production processes, also with respect to the downstream processing.

Introduction

Lactic acid is considered a renewable and decomposable building block for polylactic acid (PLA), a polymer that can substitute the mostly oil derived plastic PET^{1,2}. Traditional workhorses for industrial lactic acid production are lactic acid bacteria (LAB). These bacteria are mostly incapable of oxidative phosphorylation³, where homolactic species produce lactate anaerobically from sugars under near neutral pH conditions⁴. In production scale fermentation the near neutral pH of the culture is usually maintained by addition of calcium hydroxide⁵.

A neutral fermentation pH has the advantage that the resulting product calcium lactate forms a solid phase with a solubility limit of $\sim 50\text{g/l}$ at 30°C ⁶, which can be separated readily from the broth. This allows high titers in batch cultivation without introducing additional osmotic pressure or product inhibition. However, in order to gain the polymerizable lactic acid (HLac) from the calcium lactate (CaLac_2), this has to be treated with a stronger acid, usually sulfuric acid, leading to stoichiometric formation of gypsum as byproduct ($0.5\text{mol CaSO}_4/\text{mol HLac}$).

The use of the titrants (calcium hydroxide and sulfuric acid) and the resulting by-product gypsum not only pose a significant cost factor on the process⁷, but are also considered unacceptable with respect to an ambition for a circular economy.

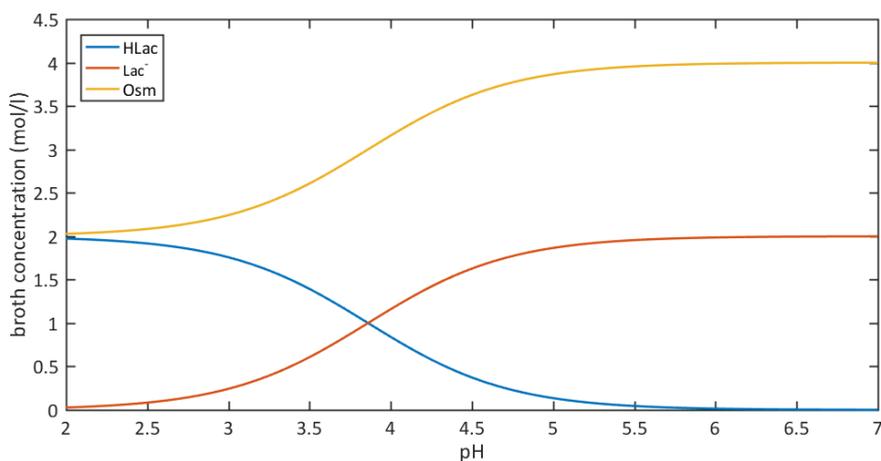


Figure 2.1. Weak-acid equilibrium of lactic acid and resulting osmolarity using a monoprotic base as titrant.

Noting that the lactic acid is a weak acid with a pK_a of 3.68, one seemingly straightforward strategy to circumvent this use of titrant and consequently the by-product formation of gypsum is to perform the fermentation at a $\text{pH} < \text{pK}_a$, where predominantly lactic acid exists in the broth, allowing its direct extraction from the

broth⁸⁻¹⁰. Another advantage of low pH fermentation is that the osmolarity of the broth at equivalent lactic acid titer is much lower, as no counter-ion is present (see Figure 2.1). Clearly, such direct lactic acid fermentation will require an acid tolerant production host; therefore in this work the physiology of the widely used yeast *S. cerevisiae* at high lactic acid titers and low pH is investigated.

Physiological response of *S. cerevisiae* to osmotic pressure and weak acid stress

One phenomenon regularly observed in fermentations is the so called weak-acid cycling¹¹ leading to ATP dissipation. It is caused by the ability of small undissociated weak-acids to diffuse through the cell membrane into the cytosol. As the cytosolic pH has to be maintained at around neutral pH¹², the acid dissociates in the cytosol to its anion and a proton. In order to maintain the intracellular pH and therewith the cellular state, the proton as well as the respective anion have to be exported back to the extracellular space. This creates a futile cycle, where the export against the diffusion gradient requires cellular energy in the form of ATP (e.g. H⁺-ATPase, ABC transporter) or indirectly by dissipating the proton motive force over the cellular membrane when the anion is exported (e.g. uniport or H⁺-antiport of the anion). This means the magnitude of weak acid cycling is expected to be proportional to the concentration of the undissociated acid in the broth and therewith increase with more acidic pH at the same nominal titer. Since the ATP consumed in this futile cycle has to be provided by the catabolism of substrate, this phenomenon has a direct influence on the achievable product and biomass yield.

There is a second mechanism that has to be taken into consideration, the osmotic stress response of *S. cerevisiae*. The cell membrane has a limited mechanical strength, so in order to counteract elevated extracellular concentrations leading to osmotic pressure, the cell has to maintain an elevated intracellular concentration of an osmolyte (also called compatible solute or osmoprotectant). In *S. cerevisiae* the major osmolyte is glycerol¹³. However glycerol is also able to diffuse through the cell membrane¹⁴ leading to glycerol leakage and consequently a loss in carbon due to osmotic response¹⁵. Furthermore, since the degree of reduction per carbon of glycerol is higher compared to the substrate glucose, the cell has to balance its redox levels by producing another molecule with a lower degree of reduction per carbon e.g. excretion of acetic acid or succinic acid (see Figure 2.2), whereas the production of CO₂ is not an option in the anaerobic experimental conditions. This by-product formation is undesirable for anaerobic fermentation processes with high

titer as it leads to additional carbon loss and moreover the excreted metabolites are weak-acids, which themselves also contribute to the futile cycling.

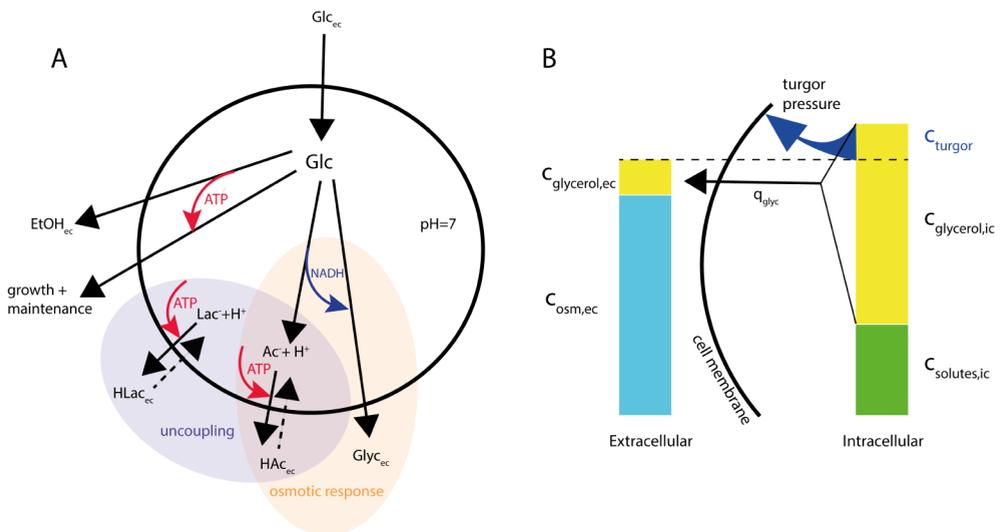


Figure 2.2. A: Schematic representation of the physiology of *S. cerevisiae*: under lactic acid and osmotic stress under anaerobic conditions. B: Schematic representation of the osmotic response in *S. cerevisiae*.

The aim of the presented work is to develop a structured metabolic model capable of describing these aspects (futile cycling, osmotic response) of cellular physiology under the relevant cultivation conditions for a direct lactic acid fermentation process. The model allows predicting the performance of *S. cerevisiae* and therewith allows exploring the design space of relevant conditions for an industrial direct lactic acid production process. Of special interest is the prediction of a maximum possible product yield, as the cost of substrate is the main cost driver in commodity chemical processes⁷. The model is useful for bioprocess engineers to evaluate different process options and for genetic engineers to deduce targets for metabolic engineering to improve the host organism.

Materials and Methods

Experimental methods

To experimentally quantify the effect of futile cycling and the osmotic response, carbon limited chemostat conditions at a constant dilution rate were chosen. The rationale for this setup is that the biomass yield on substrate is directly related to the steady-state biomass concentration in these conditions. This allows inferring the amount of ATP dissipation and carbon loss from the decrease in biomass

concentration. In order to have a sensitive system for resolving the change in biomass yield due to ATP dissipation, anaerobic conditions were chosen. Additional advantages of anaerobic conditions are that the redox requirements of the pathways can be quantified readily and that lactate cannot be catabolized.

As there is no anaerobic lactic acid producing *S. cerevisiae* strain available¹⁶, the wild-type *S. cerevisiae* strain CEN.PK113-7D was used and lactic acid production was mimicked by adding L(+)-lactic acid (80% (w/w), Sigma Aldrich) at different flowrates to the medium feed using a precision peristaltic pump (ISMATEC, Switzerland) before feeding to the bioreactor.

Three sets of chemostat experiments were performed, where the total lactic acid concentration in the broth was increased stepwise by adding lactic acid to the feed. The first experiment was performed at a pH of 3, below the pK_a of lactic acid, the second experiment was performed at a pH above the pK_a of lactic acid, at pH 5 and the third experiment was performed with adding KCl to the medium to solely investigate the effect of osmotic pressure.

Chemostat cultivations were performed at a dilution rate of 0.1 h^{-1} ; the broth volume was controlled at 1 liter by a level sensor actuating the effluent pump. The pH in the bioreactor was controlled to the respective setpoint by addition of 2mol/l KOH (at pH 3), and 4mol/l KOH (at pH 5). The temperature of the broth was controlled at 30°C. The vessel was sparged with nitrogen at a flowrate of 0.5 vvm. Norprene tubing (Cole-Parmer, USA) with little oxygen permeability was used; moreover oxygen was stripped from the medium by sparging nitrogen gas through the medium vessel. The offgas concentration of carbon dioxide was measured after drying the gas flow using an infrared analyzer (NGA 2000, Rosemount, USA).

The samples were taken from (quasi) steady-state conditions. A steady-state was assumed when the biomass dry weight and carbon dioxide production rates changed by less than 4% over at least 2 volume changes.

For the experiments with addition of lactic acid to the feed, synthetic medium with 27.5g/l D-(+)-glucose as the limiting carbon and energy source was used as described previously¹¹. The medium was also supplemented with the anaerobic growth factors ergosterol (10 mg/l) and Tween 80 (420 mg/l).

The addition of lactic acid and base, as well as the outflow of the chemostat were measured by recording the weight of the individual vessels on-line, this also allowed to calculate the respective steady-state concentrations. The mass flowrates were estimated by linear regression of all recorded data points in the steady-state. For the experiments performed with addition of KCl to induce osmotic pressure, the KCl

was directly dissolved in the medium and a final glucose concentration of 20g/l was chosen.

Biomass dry weights were determined via filtration of 20g weighed broth sample over pre-dried and weighed nitrocellulose filters with a pore size of 0.45 μ m. After filtration of the broth, the filters were washed twice with demineralized water, dried in an oven at 70°C for 72h and weighed.

Culture supernatants were obtained after sterile filtration of broth with a syringe filter with 0.22 μ m pore size (Merck Millipore) and stored at -80°C until further processing. Supernatants and media were analyzed (glucose, glycerol, acetate) via HPLC using an Aminex HPX-87H column operated at 60°C with 10 mmol/l H_2SO_4 as mobile phase at a flow rate of 0.6 ml/min. A quantitative succinate analysis was not possible as the lactate peak overlapped the succinate peak.

The osmolarity of the broth was calculated by summing up the salts of the medium, the added KOH or KCl respectively and the added lactic acid.

Metabolic core model

In order to describe the metabolic stoichiometry of *S. cerevisiae*, the following reactions are modelled (see Table 2.1), where all intracellular metabolites together with protons (charge conservation) and the cofactors ATP and NADH are balanced. As ATP/ADP and NADH/NAD⁺ are conserved moieties, ADP and NAD⁺ have been removed from the stoichiometry. For simplicity the glucose needed to balance the NADPH demand for growth was lumped into the growth reaction as described in literature^{17,18}. The model is comprised of the major pathways, glucose catabolism to ethanol to provide ATP for the biomass reaction.

The glycerol pathway has been included to balance the redox of growth under the anaerobic experimental conditions and moreover to describe the osmotic response. The redox of the osmotic response is on the other hand balanced by the pathways that produce acetic acid and succinic acid. Moreover the model considers that lactic acid, acetic acid, succinic acid and protons can diffuse through the cell membrane where they dissociate. The anions have to be exported under the investment of ATP (see also Figure 2.S4) and same holds for the protons that are exported by the proton-ATPase.

Table 2.1. Reactions for metabolic core model, where extracellular specific rates q and intracellular fluxes v are distinguished.

Name		Reaction	Mechanism
Glucose uptake rate	$q_{\text{glc,upt}}$	$\text{Glc}_{\text{ec}} \rightarrow \text{Glc}$	Facilitated diffusion
Ethanol production rate	v_{etoh}	$0.5 \text{Glc} \rightarrow 1 \text{EtOH} + 1 \text{CO}_2 + 1 \text{ATP}$	
Ethanol excretion	$q_{\text{etoh,diff}}$	$\text{EtOH} \rightarrow \text{EtOH}_{\text{ec}}$	Diffusion
Glycerol production	v_{glyc}	$0.5 \text{Glc} + 1 \text{ATP} + 1 \text{NADH} \rightarrow 1 \text{Glyc}$	
Glycerol excretion	$q_{\text{glyc,diff}}$	$\text{Glyc} \rightarrow \text{Glyc}_{\text{ec}}$	Diffusion
Acetate production	v_{ac}	$0.5 \text{Glc} \rightarrow 1 \text{Ac}^- + 1 \text{H}^+ + 1 \text{ATP} + 2 \text{NADH} + 1 \text{CO}_2$	
Acetate excretion	$q_{\text{ac,exp}}$	$\text{Ac}^- + 1 \text{ATP} \rightarrow \text{Ac}^-_{\text{ec}}$	ABC export of Ac^-
Acetic acid back-diffusion	$q_{\text{ac,diff}}$	$1 \text{HAc}_{\text{ec}} \rightarrow 1 \text{Ac}^- + 1 \text{H}^+$	Back-diffusion of HAc
Succinate production	v_{succ}	$1 \text{Glc} \rightarrow 1 \text{Succ}^{2-} + 2 \text{CO}_2 + 2 \text{ATP} + 5 \text{NADH} + 2 \text{H}^+$	Oxidative part of TCA cycle
Succinate excretion	$q_{\text{succ,exp}}$	$1 \text{Succ}^{2-} + 1 \text{ATP} \rightarrow 1 \text{Succ}^{2-}_{\text{ec}}$	ABC export of Succ^{2-}
Succinic acid back-diffusion	$q_{\text{succ,diff}}$	$1 \text{H}_2\text{Succ}_{\text{ec}} \rightarrow 1 \text{Succ}^{2-} + 2 \text{H}^+$	Diffusion of H_2Succ
Carbon dioxide excretion	q_{CO_2}	$1 \text{CO}_2 \rightarrow 1 \text{CO}_{2,\text{ec}}$	Diffusion
Biomass equation	μ	$0.1865 \text{Glc} + 1.34 \text{ATP} + 0.17 \text{NH}_4^+ \rightarrow 1 \text{C}_{1\text{H}_{1.87}\text{O}_{0.62}\text{N}_{0.17}} + 0.119 \text{CO}_2 + 0.178 \text{NADH} + 0.17 \text{H}^+$	¹¹
Lactic acid back-diffusion	$q_{\text{lac,diff}}$	$1 \text{HLac}_{\text{ec}} \rightarrow 1 \text{Lac}^- + 1 \text{H}^+$	Diffusion of HLac
Lactate excretion	$q_{\text{lac,exp}}$	$1 \text{Lac}^- + 1 \text{ATP} \rightarrow 1 \text{Lac}^-_{\text{ec}}$	ABC transport of Lac^-
Proton back-diffusion	$q_{\text{H,diff}}$	$1 \text{H}^+_{\text{ec}} \rightarrow 1 \text{H}^+$	Diffusion of protons
Proton excretion	$q_{\text{H,exp}}$	$1 \text{H}^+ + 1 \text{ATP} \rightarrow 1 \text{H}^+_{\text{ec}}$	Proton export by H^+ -ATPase
Cellular maintenance	q_{m}	$1 \text{ATP} \rightarrow$	

Selected free rates and kinetic expressions

The stoichiometry matrix of the balanced intracellular metabolites has 8 degrees of freedom; this means 8 linearly independent rates (q_f) have to be chosen in order to find a unique solution for all rates in the system. Kinetic expressions are introduced

(see Table 2.2) that link the operating conditions to the stoichiometry. The same set of kinetic equations and parameters is used to describe all experimental conditions; this has the advantage that the model with the estimated parameters can be extrapolated for other conditions and it also leads to a reduction of parameters to be estimated from the observables.

Table 2.2. Mathematical expressions used for the chosen set of linearly independent rates

Name	Expression
$q_{glyc,diff}$	$q_{glyc,diff} = P_{glyc} \cdot a \cdot (c_{osm,ec} + c_{turgor} - c_{solute,ic})$
$q_{ac,exp}$	$q_{ac,exp} = (q_{glyc,diff} - q_{glyc,\mu}) \cdot \lambda_{ac/glyc}$
μ	$\mu = const.$
$q_{lac,diff}$	$q_{lac,diff} = P_{HLac} \cdot a \cdot c_{HLac,ec}$ $with\ c_{HLac,ec} = \frac{c_{lac,ec}}{10^{pH-3.86} + 1}$
$q_{ac,diff}$	$q_{ac,diff} = P_{Hac} \cdot a \cdot c_{Hac,ec}$ $with\ c_{Hac,ec} = \frac{c_{ac,ec}}{10^{pH-4.76} + 1}$
$q_{succ,diff}$	$q_{succ,diff} = P_{H_2succ} \cdot a \cdot c_{H_2succ,ec}$ $with\ c_{H_2succ} = \frac{c_{succ,ec}}{10^{pH-4.2} + 10^{(4.2-5.6)-pH^2} + 1}$
$q_{H,diff}$	$q_{H^+,diff} = P_{H^+} \cdot a \cdot c_{H^+,ec}$
q_m	$q_m = m_{ATP}$

For weak-acids, it is assumed that only the uncharged species can diffuse through the cell membrane and that the intracellular concentrations of these species are negligible compared to the extracellular concentrations (the mole fractions are well below 0.001) at the near neutral pH conditions inside the cell. The concentrations of undissociated weak organic acids are calculated from the total concentration and the pH in the broth using the Henderson-Hasselbalch equation (see Table 2.2).

Effect of osmotic pressure on the specific cell surface

The cell volume and therewith the specific cell surface a , are a function of the osmolarity. An experimental relation as previously published¹⁹ is used here, with an initial biomass specific surface area of 2.6791 m²/gDW²⁰. Same holds for the turgor pressure, which can be derived using the Boyle-Van't Hoff relation²¹ that describes

the dependence between the turgor pressure and the observed cell volume²². The collected data has been approximated using polynomial regressions (see Appendix and Figure 2.S1 therein). The turgor pressure has been converted to concentrations using the Van't Hoff equation for convenience.

Simulation and parameter estimation

With the set of parameters θ , the independent rates q_f are defined and the remaining dependent rates q_d and fluxes v have been calculated assuming steady-state conditions for the balanced intracellular metabolites.

$$\begin{bmatrix} q_d \\ v_i \end{bmatrix} = -N_d^{-1} \cdot N_f \cdot q_f(\theta) \quad (2.1)$$

Using all rates, the set of mass balances for the broth as well as the gas phase balances were solved for the extracellular steady-state concentrations in the fermenter using the Levenberg-Marquardt algorithm in the MATLAB *fso/vc* optimizer.

$$\begin{aligned} \frac{dc_{i,ec} V_{liquid}}{dt} = 0 &= F_{in} \cdot c_{i,feed} - F_{out} \cdot c_{i,ec} + q_i c_x V_{liquid} \\ \left(\frac{pV_{gas}}{RT} \right) \left(\frac{dy_{CO_2}}{dt} \right) = 0 &= q_{CO_2} c_x V_{liquid} - F_{gas} y_{CO_2, out} \end{aligned} \quad (2.2)$$

With this forward simulation on hand, the kinetic parameters have been estimated by minimizing the residual sum of squares between experimental observables and model prediction with the MATLAB *fmincon* solver, using the sequential quadratic programming algorithm therein.

The directionality of all dependent fluxes has been constrained to the intended direction; also strictly positive steady state concentrations have been enforced using non-linear inequality constraints in the optimizer.

$$\begin{aligned}
\hat{\theta} &= \arg \min_{\theta} (\text{RSS}(\theta)) \\
&\text{such that} \\
&\begin{cases} c_i > 0 \\ q_i > 0 \end{cases} \\
&\text{with} \\
\text{RSS} &= \sum_{\text{exp}} \sum_i \left(\frac{c_{i,ec} - c_{i,meas}}{\sigma_i} \right)^2
\end{aligned} \tag{2.3}$$

The used measurement errors σ_i can be found in the Appendix. The comparably large error for ethanol reflects the uncertainty associated with ethanol stripping from the fermenter^{23,24}. Asymptotic confidence intervals have been computed by linearizing the system at the parameter optimum using finite differences and performing error propagation.

Prediction of the maximal yield for an aerobic direct lactic acid production process

To predict the maximum possible lactic acid yield on glucose, a flux balance analysis (FBA) was performed. Three reactions for aerobic glucose catabolism and oxidative phosphorylation have been added to the stoichiometric network where a P/O ratio of 1.2²⁵ was assumed.

Table 2.3. Added reactions for prediction of an aerobic lactic acid production process.

Name		Reaction
Aerobic glucose catabolism	$v_{\text{glc,cat}}$	$\text{Glc} \rightarrow 6\text{CO}_2 + 2\text{ATP} + 12\text{NADH}$
Oxidative phosphorylation	q_{resp}	$\text{NADH} + 0.5\text{O}_2 \rightarrow 1.2\text{ATP}$
Lactate production	v_{lac}	$\text{Glc} \rightarrow 2\text{Lac}^- + 2\text{H}^+ + 2\text{ATP}$

The lactic acid excretion rate was maximized, where the glucose uptake rate is set to the maximum value. The critical oxygen uptake rate was constrained to its experimentally observed maximum value²⁶ to reflect the maximal respiratory capacity of the system.

$$\begin{bmatrix} q_{i,opt} \\ v_{i,opt} \end{bmatrix} = \arg \max_{q_i, v_i} (q_{lac,exp})$$

such that

$$\left\{ \begin{array}{l} N \cdot \begin{bmatrix} q_i \\ v_i \end{bmatrix} = 0 \\ q_{glc,upt} = q_{glc,upt,max} \\ v_i \geq 0 \\ q_i \geq 0 \\ q_{o_2} \leq q_{o_2,crit} \end{array} \right. \quad (2.4)$$

Where the achieved product yield then calculates to:

$$Y_{p/S} = \frac{q_{lac,exp} - q_{lac,diff}}{q_{glc,upt}} \quad (2.5)$$

The FBA has been performed using the MATLAB *linprog* optimizer, where the intended directionality of the fluxes was enforced using bound constraints.

Results and Discussion

Experimental results

The lactic acid experiments show a comparable decrease in steady-state biomass concentration for both the experiment with increasing lactic acid concentration in the broth at a pH of 3, but also the pH 5 conditions (see Figure 2.3).

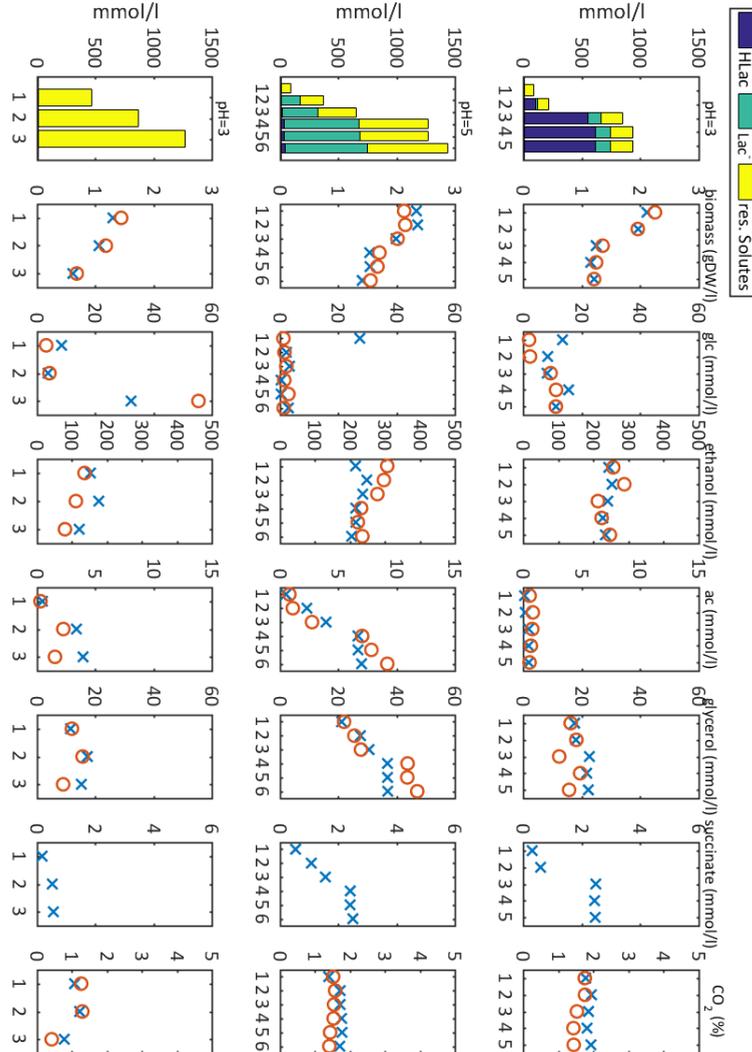


Figure 2.3. Experimental observables and best model fit. Circles are measurements, crosses are model simulations (best fit). Top row: pH3 with exposure to lactic acid; middle row: pH5 with exposure to lactic acid; bottom row: pH3 exposure to KCl. The abscissa shows the different experimental steady-states consecutively. For succinate no measurements were obtained.

This indicates that weak-acid cycling of lactic acid is not the only relevant effect influencing the biomass yield in the system. Importantly the model predicts the steady-state biomass concentrations consistently within 10% of the measurements.

Osmotic response and redox metabolism

Looking at the concentrations of glycerol it becomes clear that the conditions at pH 5 with an osmolarity up to 1430 mOsm results in the highest glycerol production leading to a steady state concentration up to 45 mmol/l at a pH of 5 (corresponding to 14.9% of the carbon uptake), which is slightly underpredicted by the model as the optimizer has to find a trade-off for P_{glyc} across all conditions, whereas the effect is less pronounced at a pH of 3 with a maximal osmolarity is 930 mOsm (20 mmol/l glycerol, corresponding to 8.4% of the carbon uptake). The effect is also observed in a comparable magnitude when exposing cells to KCl only (maximum of 1265 mOsm, corresponding to 11.7% of the carbon uptake). It is known that glycerol production is induced by osmotic pressure²⁷, where the metabolite is a so called compatible solute²⁸ that can counteract the osmotic pressure. Therefore the glycerol production rate was modelled as a function of the overall osmolarity in the medium. This osmolarity is comprised of the salts in the medium, the added lactic acid and the added base for pH control. However, a slightly lower glycerol rate in the KCl experiment was observed compared to the lactic acid conditions; it seems that *S. cerevisiae* can mitigate osmotic stress towards KCl better than towards lactate (see Figure 2.4 left). This is reflected in a higher estimated value for the concentration of intracellular solutes c_{solute} . The observed difference is within the variance of previously observed intracellular accumulation of solutes under osmotic stress²⁹. The basal glycerol production rate $q_{glyc,\mu}$ observed across all conditions is an effect of balancing the redox of the biomass equation. Side products such as acetate or succinate provide the additional NADH needed to synthesize the glycerol lost in the osmotic response.

Therefore, it is not surprising that the production rate of acetate correlates with the observed glycerol production (see Figure 2.4, right). However, when the specific rate of glycerol formation is plotted as a function of acetate formation, a significant variance in the coupling depending on the imposed conditions can be observed. It can be seen that the observed rate of acetate production is not sufficient to balance the redox under the lactic acid conditions, where the effect is more pronounced at a pH of 3. The KCl condition, on the other hand, shows a good

agreement. To reflect this redox deficiency, the model was extended by another reaction providing NADH, succinic acid production through the oxidative part of the TCA cycle³⁰, where further a linear correlation between the glycerol and the acetate production rate is assumed using the empirical proportionality factor $\lambda_{ac/glyc}$ (see Table 2.2). The secreted weak organic acids lead to additional futile cycling therewith affecting the energy metabolism of the cells. Therefore it is crucial that the model is able to predict the steady state concentrations as good as possible to get a representative picture of the ATP sinks.

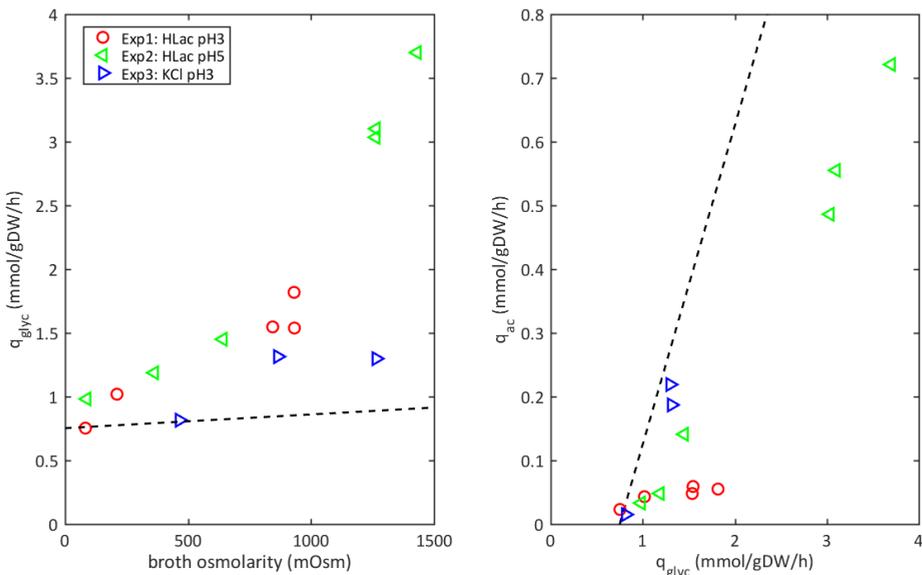


Figure 4. Left: Specific glycerol production rate as a function of broth osmolarity, dashed line indicates the stoichiometric glycerol production rate caused by biomass formation also taking into account the slight increase in dilution rate caused by the addition of base with increasing osmolarity. Right: Redox balancing between specific glycerol and specific acetate production rate under the different experimental conditions. Dashed line indicates a redox balanced ratio between acetate and glycerol formation of $\lambda_{ac/glyc} = 0.5$ with the root $q_{glyc,\mu}$.

Estimated model parameters and energy metabolism

In contrast to the osmotic and redox related effects, the energetic effects, such as lactic acid, acetic acid, and proton uncoupling cannot be discriminated from the observable rates but have to be inferred from the estimated model parameters.

From the best fit one can see that the model can reproduce the experimental observations sufficiently to investigate the underlying effects on cellular physiology. Moreover, all unknown free parameters of the model can be estimated

and also identified from the observables. The model also gives a reconciled set of rates (see Figure 2.S2).

Table 2.4. Model parameters with asymptotic confidence interval for parameters estimated from the experimental observables.

	Value ($\pm \sigma$, if applicable)	Remark
a	m^2/gDW	Function of broth osmolarity c_{osm}^{19} ; see also Figure S1
c_{turgor}		Function of broth osmolarity c_{osm}^{19} ; see also Figure S1
P_{glyc}	$1.584 \cdot 10^{-6} \pm 0.07 \cdot 10^{-6} \text{ m/h}$	This work
$c_{\text{solute},\text{lac}}$	$330.5 \pm 18.64 \text{ mmol/l}$	This work; intracellular solute concentration for HLac experiments
$c_{\text{solute},\text{KCl}}$	$461.4 \pm 25.01 \text{ mmol/l}$	This work; intracellular solute concentration for KCl experiments
P_{Hlac}	$2.882 \cdot 10^{-6} \pm 0.306 \cdot 10^{-6} \text{ m/h}$	This work
P_{Hac}	$2.6 \cdot 10^{-4} \text{ m/h}$	Taken from literature ³¹
$P_{\text{H}_2,\text{succ}}$	$3.96 \cdot 10^{-6} \text{ m/h}$	Taken from literature ³¹
P_{H^+}	10^{-3} m/h	Taken from literature ³¹
m_{ATP}	$0.1440 \pm 0.3506 \text{ mmolATP/gDW/h}$	This work
$q_{\text{o}_2,\text{crit}}$	$20 \text{ mmol oxygen/gDW/h}$	Taken from literature ²⁶
$q_{\text{glyc},\mu}$	$0.7903 \text{ mmol glycerol/gDW/h}$	Estimated from growth stoichiometry
$\lambda_{\text{ac/glyc}}^{\text{Exp1}}$	0.0288	Estimated by linear regression from Figure 2.4 ($R^2=0.80$), pH=3 addition of HLac
$\lambda_{\text{ac/glyc}}^{\text{Exp2}}$	0.2506	Estimated by linear regression from Figure 2.4 ($R^2=0.99$), pH=5, Addition of HLac
$\lambda_{\text{ac/glyc}}^{\text{Exp3}}$	0.381	Estimated by linear regression from Figure 2.4 ($R^2=0.97$), pH=3, Addition of KCl

With the estimated parameters at hand, the contributions of the respective mechanisms to the ATP turnover can be calculated. The estimated maintenance coefficient is comparable with values estimated previously^{32,33}, this indicates that the model reflects the major ATP consuming reactions in *S. cerevisiae*.

At pH 3, the ATP dissipation increases with the concentration of lactic acid in the broth and is clearly higher compared to the same nominal titer at lower pH (see Figure 2.5). The excreted acetate is coupled to the osmotic response (glycerol) as discussed before, where acetic acid is a much more potent uncoupler compared to lactic acid with a permeability coefficient 2 orders of magnitude higher (see Table 2.4). Additionally acetic acid has a higher pK_a of 4.76, which leads to substantial acetate based uncoupling at pH 5, whereas the uncoupling caused by succinic acid was found to be negligible under all experimental conditions. Besides the already discussed loss in carbon, the production of glycerol is also the largest sink of ATP at pH 5.

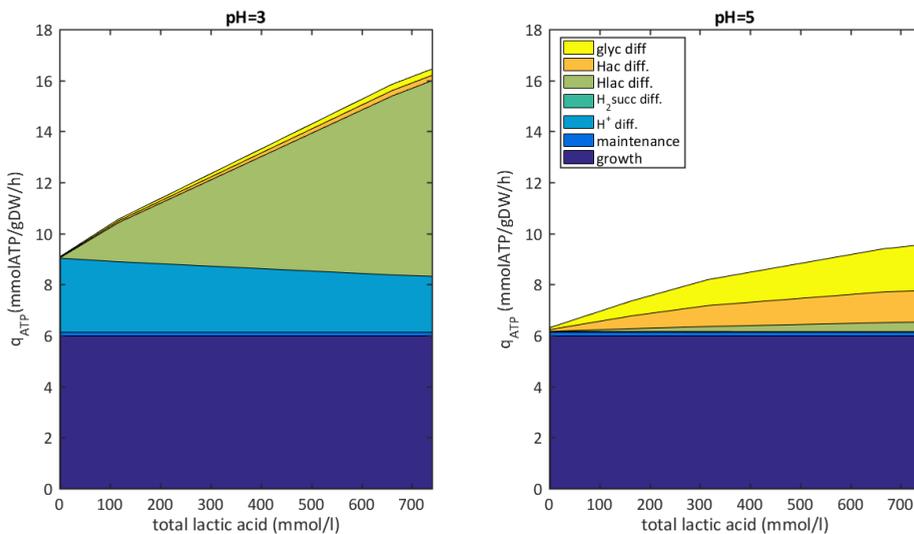


Figure 2.5. ATP demand by growth ($\mu = 0.1$ 1/h), maintenance, diffusion of the weak-acids (acetic, lactic, succinic acid) and protons through the cell membrane and the diffusion of glycerol as a function of the total lactic acid concentration.

Analysis of a direct lactic acid production process

From a process design perspective, anaerobic processes are favored as the fermenters do not need aeration; this saves cost for the compression of air as well as for agitation (mass-transfer) and also makes it easier to scale-up the process³⁴. Therefore the feasibility of an anaerobic microbial production process was evaluated by calculating the available Gibbs free energy of the product pathway at conditions representative for a production process (see Figure 2.6; for the calculations refer to the Appendix).

The cost for downstream processing will primarily depend on the concentration of lactic acid in the broth and therewith favor a process at a pH below the pK_a of lactic acid. However, the disadvantage of a low pH fermentation is that the change in Gibbs free energy obtained in the pathway from glucose to lactic acid is decreasing with decreasing pH, whereas the amount of Gibbs free energy needed for the export of lactic acid from the pH neutral cytosol through the cell membrane increases³⁵. Considering that available Gibbs free energy, it can be seen that the pathway can deliver 2 ATP (corresponding to a change in Gibbs free energy of about 50 kJ/mol under physiological conditions) per glucose at neutral pH. However, due to the decrease in Gibbs free energy, this ATP production reduces with decreasing pH to only 1 ATP below a pH of 6.4 and 0 ATP below a pH of 4.8. This also offers the explanation why homofermentative lactic acid bacteria cannot grow anymore at low pH⁴.

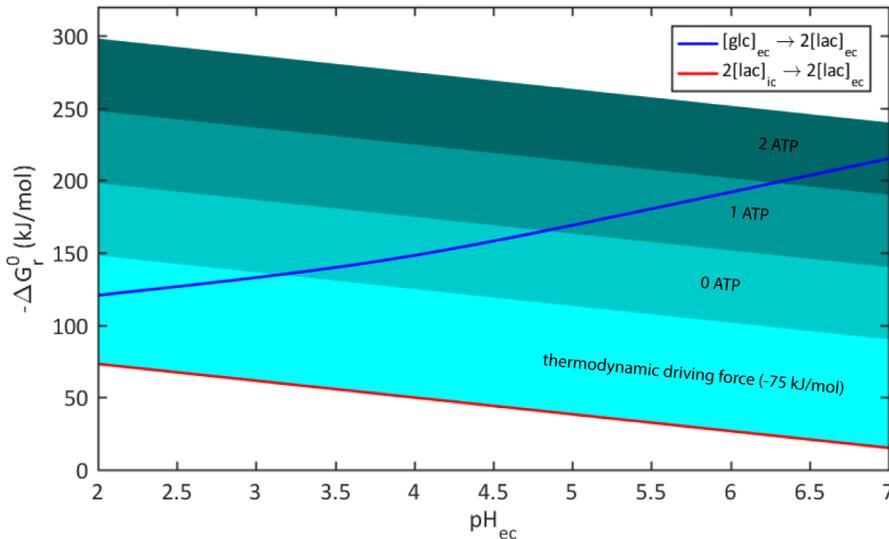


Figure 2.6. Black-box thermodynamic considerations for a direct lactic acid production process at an extracellular Lac titer of 2 mol/l as a function of the broth pH. Further an intracellular Lac concentration of 0.1 mol/l and an intracellular pH of 7 have been assumed. The thermodynamic driving force of the pathway has been calculated to 55kJ/mol as presented in literature³⁶ with an addition of 20 kJ/mol for the export of the two Lac molecules from the cytosol (see also Appendix).

This means that an anaerobic low pH process is thermodynamically impossible under the assumed conditions, even using strategies to increase the free energy conservation³⁷ will only change this conclusion marginally. Assuming that it would be possible to gain another 0.5 ATP from the product pathway³⁸, this would correspond

to a decrease of the thermodynamic driving force of one third and reduce the feasible pH from 4.8 to 4.15, which is still significantly above the pK_a of lactic acid. This means the only viable process alternative for a low pH process is an aerobic process, where oxidative phosphorylation can provide the ATP for growth and maintenance of the system. This may be considered a drawback but at the same time it solves another challenge that has been observed in the anaerobic experiment, the balancing of redox. In an aerobic process the oxidative TCA cycle can provide the NADH needed to balance the glycerol pathway, therewith eliminating the by-product formation of acetate and succinate. This has multiple advantages for a process; first there is no additional by-product which complicates the downstream processing, especially when considering a selective continuous process such as reactive extraction^{5,8,10,39,40} which could lead to accumulation of the byproduct in the broth. Secondly, aerobic conditions reduce the uncoupling induced especially by acetic acid but also succinic acid which can lead to significant ATP loss at elevated concentrations⁴¹. Oxidative phosphorylation will lead to CO_2 as a by-product that undergoes phase separation and can be readily removed from the broth. The third advantage is that the carbon in CO_2 is fully oxidized and will therewith lead to a lower carbon loss compared to acetic or succinic acid produced to compensate for the redox lost in glycerol production.

Prediction of the maximal product yield on substrate

Having found thermodynamically feasible operating conditions, the parametrized model allows extrapolating the pH and osmolarity effects to a wide range of pH and titer conditions for a producing system. To predict the yield in an aerobic process, the stoichiometry matrix is extended with the reactions for the TCA cycle and the oxidative phosphorylation as well as the lactate production pathway. This then allows calculating the maximum possible yield as a function of the respective pH and lactic acid titer conditions.

In the optimization the availability of the electron acceptor oxygen leads consistently to the result that acetate secretion is absent, as it would result in a larger carbon and energy loss. The considered mechanisms that decrease the product yield are the glucose oxidation to cover up for the ATP demand for cellular maintenance and lactate export, the uncoupling due to the lactic acid in the broth and the glycerol synthesis for the osmotic response.

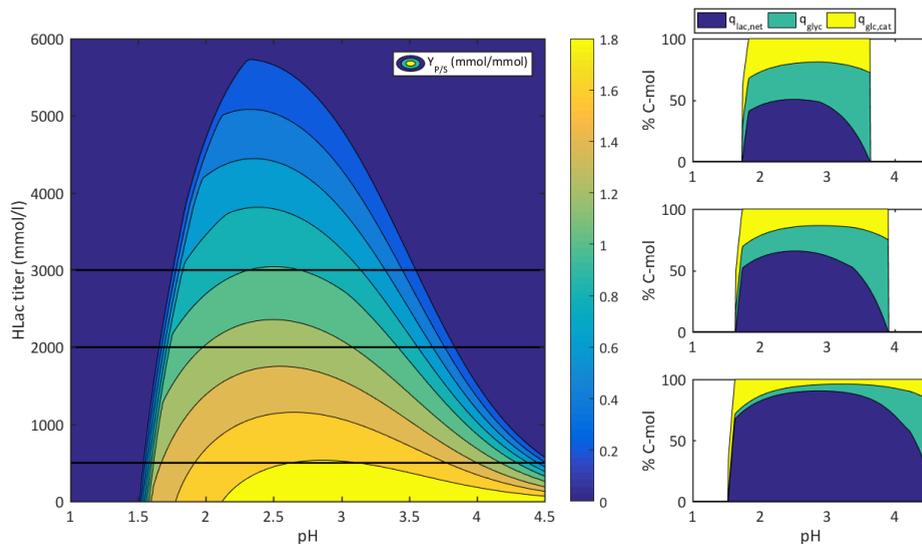


Figure 2.7. Left: Prediction of the maximum possible lactic acid yield on glucose as a function of the lactic acid titer in the broth and the pH in the broth. Right: Estimated glucose distributions for lactic acid concentrations of 500 mmol/l (low), 2000 mmol/l (middle) and 3000 mmol/l (upper).

In Figure 2.7, it can be seen that the yield isoclines show a clear optimum, where the optimal pH is around 2.5. This shows that the carbon and ATP loss induced by osmolarity, which increases with pH, leads to higher losses in product yield than uncoupling due to lactic acid diffusion. It can be clearly seen that a higher lactic acid titer will coincide with a loss in product yield due to increasing uncoupling. This means a downstream process e.g. a reactive extraction of lactic acid should best be operated with a titer as low as possible from a product yield point of view. The overall process economics also depend on other factors e.g. for a direct extraction of lactic acid a high titer in order to achieve a high mass transfer to the extractant will be desirable, which can only be achieved on the cost of product yield. Here, the model can help finding the best compromise between the cost for fermentation and downstream processing. It can also be seen that at a pH of about 1.6 the system is rapidly decaying, the reason for this is that the respiratory capacity cannot provide sufficient ATP for the export of the protons. This limit is in agreement with experimental observations⁴².

Metabolic engineering targets

From the model a number of strategies that could improve the product yield on substrate can be deduced.

The model shows that the energetic capacity of *S. cerevisiae* is sufficient to counteract lactic acid diffusion even at high lactic acid titers up to 5.5 mol/l.

However, high titers can only be achieved on the cost of ATP and therewith product yield, leading to a target conflict with respect to downstream processing where a high titer will be desirable for minimizing cost. This raises the question how the cells could be engineered to increase the product yield at elevated lactic acid titers in order to alleviate this conflict.

The P/O ratio mainly determines the ATP yield on substrate, *S. cerevisiae* exhibits a low ratio due to absence of a proton-translocating NADH dehydrogenase compared to other yeasts e.g. *Candida utilis*¹⁷; Here a yeast with a higher P/O ratio would have a lower rate of glucose catabolism necessary for the same rate of ATP therewith showing an increased product yield.

S. cerevisiae shows a remarkable osmotolerance²⁹, which comes at the price of glycerol production. In contrast to the growth associated glycerol production^{23,43} this flux cannot be eliminated e.g. by knockout of the genes encoding the glycerol-3-phosphate dehydrogenase without obstructing the cellular physiology⁴⁴. Here a possible strategy would be to engineer the cell membrane towards a lower permeability for glycerol, therewith reducing the loss by diffusive outflow at elevated osmotic pressure.

Another strategy could be to (over)express a pathway producing an alternative compatible solute that does not diffuse out of the cell, e.g. other polyols like arabitol, mannitol or sorbitol. Research⁴⁵ has shown that 6-carbon polyols are retained more compared to glycerol and that these components can partly complement the lack of glycerol. However the role of glycerol is more complicated as it is also involved in the dynamic adaptation to osmotic shock via the HOG⁴⁶ pathway using the FPS1 transporter, which may make replacing it a tedious task.

To reduce the ATP loss from for weak-acid cycling, the same metabolic strategy as proposed for glycerol could be applied, reducing the membrane permeability for lactic acid⁴⁷. This has for example been successfully achieved for acetic acid by use of directed evolution³¹. Assuming that the diffusion of lactic acid as well as the carbon loss through glycerol diffusion is reduced by 80% the solution space for a high product yield of 1.8 mol lactic acid/mol glucose can be extended to a titer of 2.2 mol/l lactic acid (see Figure 2.S3).

Low pH lactic acid production processes have been successfully implemented using non-*Saccharomyces* yeast; whereas efforts in *S. cerevisiae* to our best knowledge have been not successful. Conversely, the results indicate that the seemingly low suitability of *S. cerevisiae* for the production of lactic acid at low pH is not caused by energetic limitations or a limited capacity for the active export of lactic acid in the applied conditions. It has been shown that a process at low pH would have to be aerobic in order to deliver the necessary energy for the product pathway and export. However, there are also reports of adverse effect of oxygen in presence of lactic acid^{48,49} leading to a low lactic acid tolerance. The relations causing this effect are poorly understood and should be researched, best in comparison to yeast strains showing higher robustness to high lactic acid concentrations at low pH in the presence of oxygen. This may lead to the situation that the dissolved oxygen concentration has to be minimized i.e. a microaerobic process. On the other hand, the productivity is dependent on the availability of oxygen to provide the ATP for the product pathway and oxygen limitation may have detrimental effects on PDC negative cells^{50,51}.

Independently of the implementation, the modeling approach gives a clear ranking of metabolic engineering targets as a function of the defined process targets and can help finding an optimal compromise for a direct lactic acid process by evaluating the impact of the operating conditions on the cellular physiology and therewith the maximum possible product yield in the fermentation.

Symbols and Abbreviations

a	Specific surface area
c	Concentration
F	Flow rate
m_{ATP}	Maintenance coefficient
N	Stoichiometry matrix
P	Permeability coefficient
q	Specific rate
R	Gas constant
RSS	Residual sum of squares
T	Temperature
t	Time
v	Flux
V	Volume
y	Mole fraction (gas phase)
$Y_{P/S}$	Product yield on substrate
θ	Model parameters
$\hat{\theta}$	Optimal set of parameters with respect to minimal RSS
$\lambda_{ac/glyc}$	Stoichiometric coefficient
σ	Standard deviation

Subscript/Superscript

ac	Nominal acetic acid
ac^-	Acetate
ATP	Adenosine triphosphate
cat	Catabolic
CO_2	Carbon dioxide
$crit$	Critical uptake rate
d	Dependent rate/flux
$diff$	Diffusion through cell membrane
$etoh$	Ethanol
Exp	Experiment
ec	Extracellular
exp	Export from cytosol
f	Free rate
gas	Gaseous control volume
glc	Glucose
$glyc$	Glycerol

H^+	Proton
<i>Hac</i>	Acetic acid
<i>Hlac</i>	Lactic acid
H_2succ	Succinic acid
<i>i</i>	Generic variable for metabolites
<i>in</i>	Inflow to control volume
<i>lac</i>	Nominal lactic acid or refers to experiments using lactic acid in feed
lac^-	Lactate
<i>liquid</i>	Liquid control volume
<i>m</i>	Cellular maintenance
O_2	Oxygen
<i>osm</i>	Osmotic pressure
<i>out</i>	Outflow of control volume
<i>resp</i>	Respiratory
<i>solutes</i>	Intracellular solutes
<i>succ</i>	Nominal succinic acid
$succ^{2-}$	Succinate
<i>turgor</i>	Turgor pressure
<i>upt</i>	Uptake rate
<i>upt, max</i>	Maximal uptake rate
<i>x</i>	Biomass
μ	Growth/Biomass

References

- 1 Papong, S. *et al.* Comparative assessment of the environmental profile of PLA and PET drinking water bottles from a life cycle perspective. *Journal of Cleaner Production* 65, 539-550 (2014).
- 2 Singh, J., Auras, R. & Singh, S. P. Performance evaluation of PLA against existing PET and PS containers. *Journal of Testing and Evaluation* 34, 1-7 (2006).
- 3 Brooijmans, R. *et al.* Heme and menaquinone induced electron transport in lactic acid bacteria. *Microbial Cell Factories* 8, 28 (2009).
- 4 Adamberg, K., Kask, S., Laht, T.-M. & Paalme, T. The effect of temperature and pH on the growth of lactic acid bacteria: a pH-auxostat study. *International Journal of Food Microbiology* 85, 171-183 (2003).
- 5 Ghaffar, T. *et al.* Recent trends in lactic acid biotechnology: A brief review on production to purification. *Journal of Radiation Research and Applied Sciences* 7, 222-229 (2014).
- 6 Kubantseva, N., Hartel, R. W. & Swearingen, P. A. Factors affecting solubility of calcium lactate in aqueous solutions. *Journal of dairy science* 87, 863-867 (2004).
- 7 Straathof, A. J. J. in *Comprehensive Biotechnology (Second Edition)* 811-814 (Academic Press, 2011).
- 8 Yabannavar, V. & Wang, D. Extractive fermentation for lactic acid production. *Biotechnology and bioengineering* 37, 1095-1100 (1991).
- 9 Woodley, J. M., Bisschops, M., Straathof, A. J. J. & Ottens, M. Future directions for in-situ product removal (ISPR). *Journal of Chemical Technology & Biotechnology* 83, 121-123 (2008).
- 10 Joglekar, H. G., Rahman, I., Babu, S., Kulkarni, B. D. & Joshi, A. Comparative assessment of downstream processing options for lactic acid. *Separation and Purification Technology* 52, 1-17 (2006).
- 11 Verduyn, C., Postma, E., Scheffers, W. A. & van Dijken, J. P. Physiology of *Saccharomyces cerevisiae* in anaerobic glucose-limited chemostat cultures. *J Gen Microbiol* 136, 395-403 (1990).
- 12 Orij, R., Brul, S. & Smits, G. J. Intracellular pH is a tightly controlled signal in yeast. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1810, 933-944 (2011).
- 13 Hohmann, S. Osmotic stress signaling and osmoadaptation in yeasts. *Microbiol Mol Biol Rev* 66 (2002).
- 14 De Gier, J., Mandersloot, J. G. & Van Deenen, L. L. M. Lipid composition and permeability of liposomes. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 150, 666-675 (1968).
- 15 Nevoigt, E. & Stahl, U. Osmoregulation and glycerol metabolism in the yeast *Saccharomyces cerevisiae*. *FEMS Microbiology Reviews* 21, 231-241 (1997).

- 16 Abbott, D. A., van den Brink, J., Minneboo, I. M. K., Pronk, J. T. & van Maris, A. J. A. Anaerobic homolactate fermentation with *Saccharomyces cerevisiae* results in depletion of ATP and impaired metabolic activity. *FEMS Yeast Research* 9, 349-357 (2009).
- 17 Verduyn, C., Stouthamer, A. H., Scheffers, W. A. & van Dijken, J. P. A theoretical evaluation of growth yields of yeasts. *Antonie Van Leeuwenhoek* 59, 49-63 (1991).
- 18 Verduyn, C., Postma, E., Scheffers, W. A. & Dijken, J. P. Physiology of *Saccharomyces cerevisiae* in anaerobic glucose-limited chemostat cultures. *J Gen Microbiol* 136 (1990).
- 19 Meikle, A. J., Reed, R. H. & Gadd, G. M. Osmotic adjustment and the accumulation of organic solutes in whole cells and protoplasts of *Saccharomyces cerevisiae*. *J Gen Microbiol* 134, 3049-3060 (1988).
- 20 Kresnowati, M. T. A. P., Suarez-Mendez, C., Groothuizen, M. K., van Winden, W. A. & Heijnen, J. J. Measurement of fast dynamic intracellular pH in *Saccharomyces cerevisiae* using benzoic acid pulse. *Biotechnology and Bioengineering* 97, 86-98 (2007).
- 21 Nobel, P. S. The Boyle-Van't Hoff relation. *Journal of Theoretical Biology* 23, 375-379 (1969).
- 22 Dainty, J. 9.—Plant Cell—Water Relations: The Elasticity of the Cell Wall. *Proceedings of the Royal Society of Edinburgh: Section A Mathematics* 70, 89-93 (1972).
- 23 Guadalupe-Medina, V. *et al.* Carbon dioxide fixation by Calvin-Cycle enzymes improves ethanol yield in yeast. *Biotechnology for Biofuels* 6, 125 (2013).
- 24 Duboc, P. & von Stockar, U. Systematic errors in data evaluation due to ethanol stripping and water vaporization. *Biotechnology and Bioengineering* 58, 428-439 (1998).
- 25 Gulik, W. M. v. & Heijnen, J. J. A metabolic network stoichiometry analysis of microbial growth and product formation. *Biotechnology and Bioengineering* 48, 681-698 (1995).
- 26 Verduyn, C., Postma, E., Scheffers, W. A. & Van Dijken, J. P. Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast* 8, 501-517 (1992).
- 27 Petrovska, B., Winkelhausen, E. & Kuzmanova, S. Glycerol production by yeasts under osmotic and sulfite stress. *Can J Microbiol* 45, 695-699 (1999).
- 28 Brown, A. D. Compatible solutes and extreme water stress in eukaryotic micro-organisms. *Advances in microbial physiology* 17, 181-242 (1978).
- 29 García, M. J., Rios, G., Ali, R., Bellés, J. M. & Serrano, R. Comparative physiology of salt tolerance in *Candida tropicalis* and *Saccharomyces cerevisiae*. *Microbiology* 143, 1125-1131 (1997).

- 30 Modig, T., Granath, K., Adler, L. & Lidén, G. Anaerobic glycerol production by
Saccharomyces cerevisiae strains under hyperosmotic stress. *Applied*
31 *Microbiology and Biotechnology* 75, 289-296 (2007).
- 32 Jamalzadeh, E. Transport of Dicarboxylic Acids in Saccharomyces cerevisiae
(2013).
- 33 Boender, L. G. M. *et al.* Extreme calorie restriction and energy source
starvation in Saccharomyces cerevisiae represent distinct physiological
states. *Biochimica et biophysica acta* 1813 (2011).
- 34 Vos, T. *et al.* Maintenance-energy requirements and robustness of
Saccharomyces cerevisiae at aerobic near-zero specific growth rates.
Microbial Cell Factories 15, 111 (2016).
- 35 Noorman, H. J. & Heijnen, J. J. Biochemical engineering's grand adventure.
Chemical Engineering Science 170, 677-693 (2017).
- 36 van Maris, A. J., Konings, W. N., van Dijken, J. P. & Pronk, J. T. Microbial
export of lactic and 3-hydroxypropanoic acid: implications for industrial
fermentation processes. *Metab Eng* 6, 245-255 (2004).
- 37 Cueto-Rojas, H. F., van Maris, A. J., Wahl, S. A. & Heijnen, J. J.
Thermodynamics-based design of microbial cell factories for anaerobic
product formation. *Trends Biotechnol* 33, 534-546 (2015).
- 38 de Kok, S., Kozak, B. U., Pronk, J. T. & van Maris, A. J. A. Energy coupling in
Saccharomyces cerevisiae: selected opportunities for metabolic engineering.
FEMS Yeast Research 12, 387-397 (2012).
- 39 de Kok, S. *et al.* Increasing free-energy (ATP) conservation in maltose-grown
Saccharomyces cerevisiae by expression of a heterologous maltose
phosphorylase. *Metab Eng* 13, 518-526 (2011).
- 40 Vaidya, A. N. *et al.* Production and Recovery of Lactic Acid for Polylactide—
An Overview. *Critical Reviews in Environmental Science and Technology* 35,
429-467 (2005).
- 41 López-Garzón, C. S. & Straathof, A. J. J. Recovery of carboxylic acids
produced by fermentation. *Biotechnology Advances* 32, 873-904 (2014).
- 42 Wahl, S. A., Bernal Martinez, C., Zhao, Z., van Gulik, W. M. & Jansen, M. L. A.
Intracellular product recycling in high succinic acid producing yeast at low
pH. *Microb Cell Fact* 16, 90 (2017).
- 43 Buzás, Z., Dallmann, K. & Szajáni, B. Influenc of pH on the growth and ethanol
production of free and immobilized Saccharomyces cerevisiae cells.
Biotechnology and Bioengineering 34, 882-884, (1989).
- 44 Guadalupe Medina, V., Almering, M. J. H., van Maris, A. J. A. & Pronk, J. T.
Elimination of glycerol production in anaerobic cultures of a Saccharomyces
cerevisiae strain engineered to use acetic acid as an electron acceptor. *Appl*
Environ Microbiol 76 (2010).
- 45 Albertyn, J., Hohmann, S., Thevelein, J. M. & Prior, B. A. GPD1, which encodes
glycerol-3-phosphate dehydrogenase, is essential for growth under osmotic

- stress in *Saccharomyces cerevisiae*, and its expression is regulated by the high-osmolarity glycerol response pathway. *Molecular and Cellular Biology* 14, 4135-4144 (1994).
- 45 Shen, B., Hohmann, S., Jensen, R. G., Bohnert & Hans, J. Roles of Sugar Alcohols in Osmotic Stress Adaptation. Replacement of Glycerol by Mannitol and Sorbitol in Yeast. *Plant Physiology* 121, 45-52 (1999).
- 46 Tamás, M. J., Rep, M., Thevelein, J. M. & Hohmann, S. Stimulation of the yeast high osmolarity glycerol (HOG) pathway: evidence for a signal generated by a change in turgor rather than by water stress. *FEBS Letters* 472, 159-165 (2000).
- 47 Lee, J. J., Crook, N., Sun, J. & Alper, H. S. Improvement of lactic acid production in *Saccharomyces cerevisiae* by a deletion of *ssb1*. *Journal of Industrial Microbiology & Biotechnology* 43, 87-96 (2016).
- 48 Abbott, D. A. *et al.* Catalase overexpression reduces lactic acid-induced oxidative stress in *Saccharomyces cerevisiae*. *Appl Environ Microbiol* 75, 2320-2325 (2009).
- 49 Bonnet, J. A., de Kok, H. E. & Roels, J. A. The growth of *Saccharomyces cerevisiae* CBS 426 on mixtures of glucose and ethanol: a model. *Antonie Van Leeuwenhoek* 46, 565-576 (1980).
- 50 Abbott, D. A., van den Brink, J., Minneboo, I. M., Pronk, J. T. & van Maris, A. J. Anaerobic homolactate fermentation with *Saccharomyces cerevisiae* results in depletion of ATP and impaired metabolic activity. *FEMS Yeast Res* 9, 349-357 (2009).
- 51 van Maris, A. J. A., Winkler, A. A., Porro, D., van Dijken, J. P. & Pronk, J. T. Homofermentative Lactate Production Cannot Sustain Anaerobic Growth of Engineered *Saccharomyces cerevisiae*: Possible Consequence of Energy-Dependent Lactate Export. *Applied and Environmental Microbiology* 70, 2898-2905 (2004).

Supporting Appendix

Polynomial regressions for specific surface area a and c_{turgor}

$$a = 5.533e-15c_{osm,ec}^4 - 7.887e-11c_{osm,ec}^3 + 4.314e-07c_{osm,ec}^2 - 0.0012c_{osm,ec} + 2,996$$

$$c_{turgor} = \begin{cases} -0.518c_{osm,ec} + 470.79; & 0 < c_{osm,ec} \leq 908.4 \\ 0; & c_{osm,ec} < 908.4 \end{cases}$$

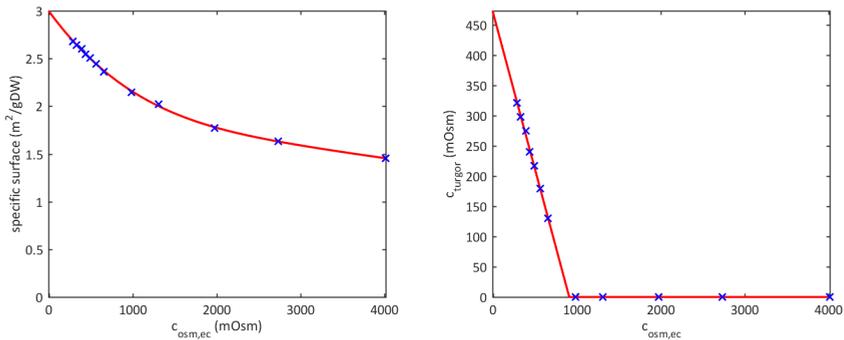


Figure 2.S1. Left: Specific surface area as a function of broth osmolarity with polynomial approximation. Right: c_{turgor} as a function of broth osmolarity with piecewise linear approximation.

Table 2.S1. Standard deviations for weighting during parameter estimation and best objective function

	σ
C_x	0.1 gDW/l
C_{glc}	1.5 mmol/l
C_{eth}	50 mmol/l
C_{ac}	0.2 mmol/l
C_{glyc}	1mmol/l
Y_{CO_2}	0.2%
RSS_{min}	1317.92

Model prediction for specific rates

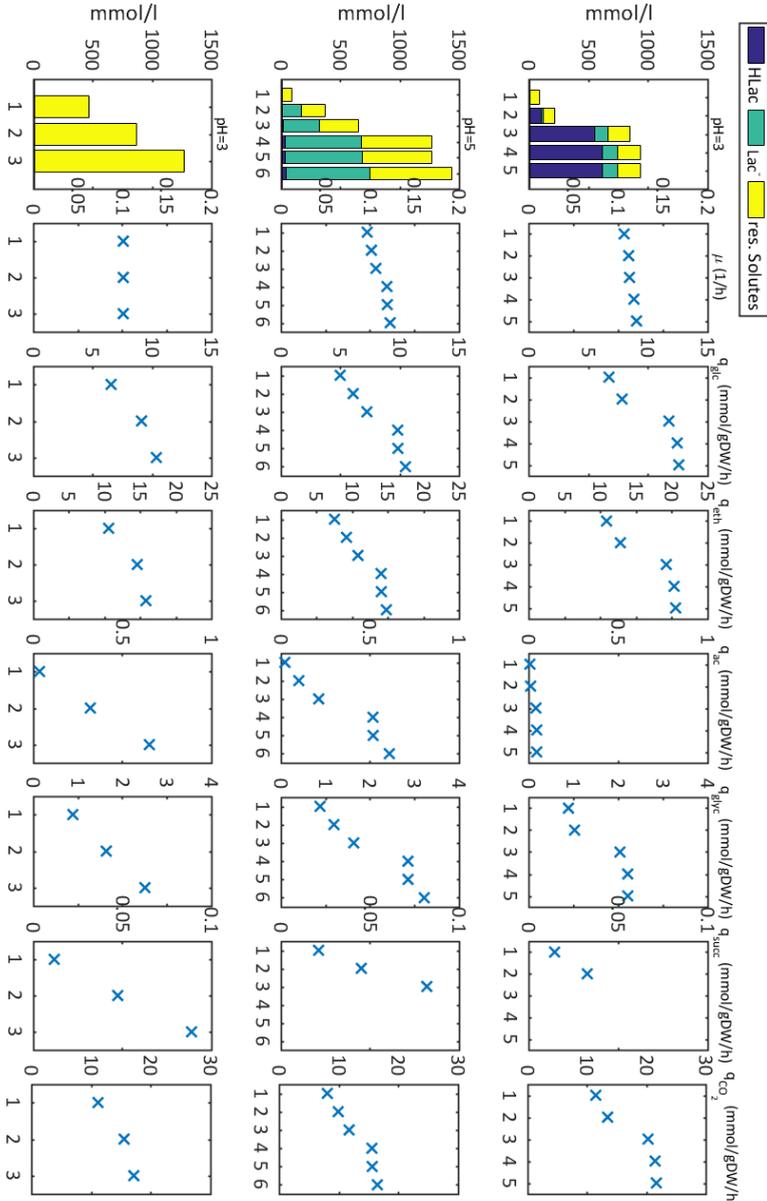


Figure 2.S2. Specific rates for the best model fit for all experimental conditions.

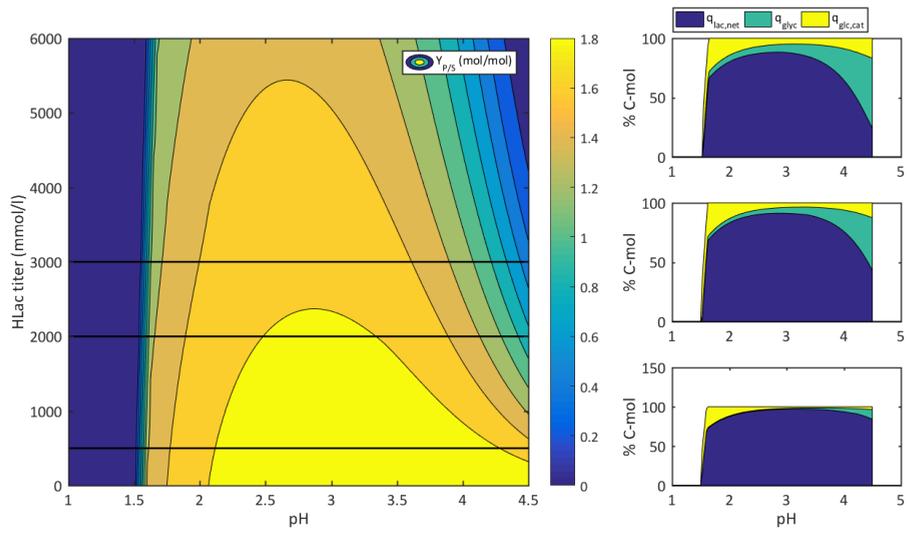


Figure 3.S3. Yield predictions for 80% reduced lactic acid diffusion and glycerol loss.

Thermodynamic calculations

$$\Delta_r G^0 = 2 \left[\Delta_f G_{Lac^-}^0 + RT \ln \frac{[Lac^-]_{ec}}{1} \right] + 2 \left[\Delta_f G_{H^+}^0 + RT \ln \frac{[H^+]_{ec}}{1} \right] - \left[\Delta_f G_{Glc}^0 + RT \ln \frac{[Glc]_{ec}}{1} \right]$$

with

$$[Glc]_{ec} = 10^{-3} \text{ mol/l}$$

$$[H^+]_{ec} = 10^{-pH}$$

$$[H^+]_{ic} = 10^{-7}$$

$$[Lac]_{ic} = 0.1 \text{ mol/l}$$

$$[Lac]_{ec} = 2 \text{ mol/l}$$

$$[Lac^-]_{ec} = \frac{[Lac]_{ec}}{1 + 10^{(pH - pKa)}}$$

$$T = 303.15 \text{ K}$$

$$\Delta_f G_{Lac^-}^0 = -517.81 \text{ kJ/mol}$$

$$\Delta_f G_{H^+}^0 = 0 \text{ kJ/mol}$$

$$\Delta_f G_{Glc}^0 = -917.33 \text{ kJ/mol}$$

$$TDF_{pathway} = 0.25 \cdot n \cdot 20 \text{ kJ/mol} \text{ (Cueto-Rojas et al., 2015)}$$

$$n = 11$$

$$\Delta G_{export} = RT \ln \frac{[HLac]_{ic}}{[HLac]_{ec}}$$

$$TDF = TDF_{pathway} + TDF_{export} = 20 \text{ kJ/mol}$$

Transporter equilibria

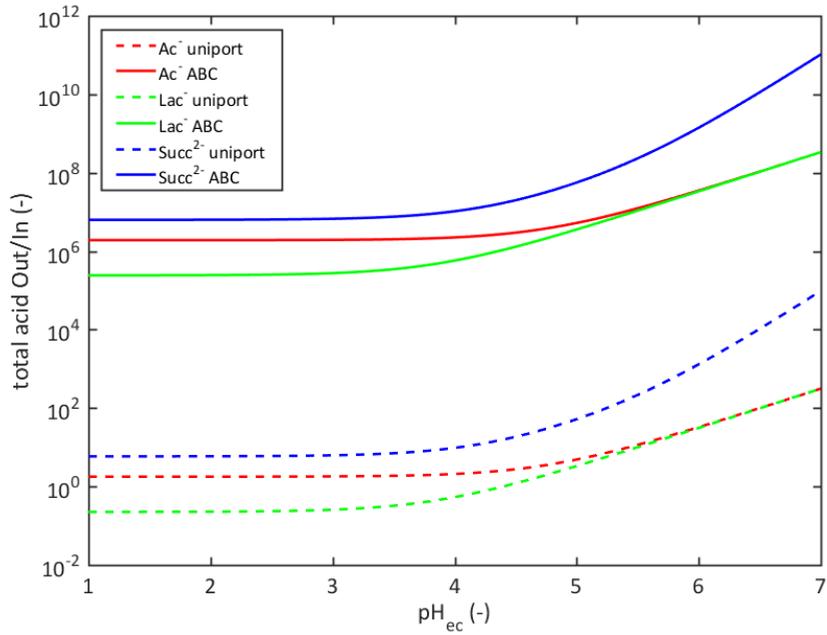


Figure 3.S4. Out/In ratios in thermodynamic equilibrium of the transport from the cytosol for uniport and ABC transport with a stoichiometry of ATP for acetate, lactate and succinate assuming a constant proton motive force of $-0.15V$ and an intracellular pH of 7.

Chapter 3:

Enrichment of the more efficient:
Droplet based cultivation of *S. cerevisiae*
for selection of phenotypes with
increased free energy conservation

in collaboration with R. Mans, T. van Maris, B. Teusink and S.A.
Wahl

“Competition for access to limited resources inevitably leads to a conflict between self-interest and the efficient usage of the common good.”

Adapted from ‘The tragedy of the commons’ of Garrett Hardin

Abstract

Increasing the free energy conservation in anabolic product pathways can increase the product yield on substrate in cell factories. However, the desired phenotypes cannot always be engineered rationally. Adaptive laboratory evolution has been shown to be a useful tool to select desired phenotypes. Usually, the desired trait of the desired phenotype can be linked to an increased growth rate, which allows enriching and isolating the desired mutants. This strategy does not work for the enrichment of more ATP efficient mutants as there seems to be a trade-off between the growth rate and the ATP efficiency i.e. more efficient strains are likely to grow slower. To circumvent this trade-off an experimental methodology consisting of repetitive incubation of cells partitioned into droplets is demonstrated for *S. cerevisiae*. A previously described experimental protocol for lactic acid bacteria is adapted to *S. cerevisiae* and furthermore a model for the experimental design of enrichment experiments is developed. In order to effectively assess the performance of such systems the ideal performance characteristics of partitioned systems are derived. The model is then extended for the relevant non-idealities observed in the system and subsequently validated experimentally by enriching a phenotype with increased free energy conservation and decreased maximal growth rate in presence of the wild-type population. It is shown that the model can predict the experimental enrichment trajectory which allows deducing the general enrichment characteristics of partitioned systems. The model facilitates the design of enrichment experiments by finding appropriate experimental settings *a priori*.

Introduction

Without political incentive, e.g. subsidies, microbial production of bulk chemicals from renewable feedstocks is in economic competition with building blocks derived from fossil resources. Especially, for low value commodity chemicals the price of the feedstock determines to a large extent the price of the final product. This means high product yields on substrate are often crucial for an economically viable production process^{1,2}.

The maximum theoretical yield for a product reaction can be derived with a degree of reduction balance, while also taking the change in Gibbs free energy into consideration^{3,4}. However, the mentioned methodology treats the product reaction as a black box and does not consider the actual available metabolic reactions. The pathway structure in turn determines how much of the released Gibbs free energy is converted to usable cellular energy in the form of ATP.

In native pathways, the available thermodynamic driving force of the product pathway reaction is usually only converted to a small extent to ATP. This can lead to the consequence that e.g. a product reaction is ATP neutral or even ATP consuming, although the change in Gibbs free energy would be theoretically sufficient to produce a surplus of ATP³. In those cases, the ATP requirement of the product reaction has to be balanced by another pathway providing ATP, e.g. complete catabolism of the substrate. This leads to the formation of side-products like CO₂, therewith decreasing the achievable product yield on substrate.

Consequently, strategies to improve the ATP yield in metabolic pathways have been proposed⁵ (see also Chapter 2), i.e. more of the Gibbs free energy of a reaction shall be converted to ATP. Still, the rational engineering results often to be very challenging. For example, de Kok et al. attempted to increase the ATP efficiency of the proton ATPase in *S. cerevisiae*⁶. Increasing the export from one to two or three protons per ATP a broad variety of yeast bioprocesses could be improved significantly⁷. Unfortunately, the rational engineering failed, urging for alternative approaches.

Evolutionary strategies can be a very powerful tool when the rational knowledge to engineer a strain with desired phenotype is not available. Also termed directed evolution, this method comprises the creation of a selective environment for the desired phenotype, here for a phenotype with an increased ATP efficiency.

Common evolutionary engineering strategies rely on selection of phenotypes with the highest growth rate under the applied experimental conditions⁸⁻¹⁰. Introducing a coupling between growth rate and the desired trait e.g. a high rate of product

formation with growth then leads to an enrichment of the desired phenotype in the system.

To demonstrate this a simplified equation for the growth rate of an unicellular organism exhibiting exponential growth can be written, where the (energy limited) growth rate is a function of the specific substrate uptake rate q_s , the apparent ATP stoichiometry of the catabolic pathway(s) n_{ATP} and the biomass yield on ATP $Y_{X/ATP}$.

$$\mu = q_s n_{ATP} Y_{X/ATP} \quad (3.1)$$

Assuming that the experimental setup selects for the fastest growing phenotype, a maximization problem can be derived. Furthermore it is assumed that the ATP requirement for biomass synthesis ($Y_{X/ATP}$) is constant for all phenotypes and independent of the growth rate.

$$\begin{aligned} \mu_{\max} &= \arg \max_{q_s, n_{ATP}} (q_s n_{ATP} Y_{X/ATP}) \\ &\text{with} \\ Y_{X/ATP} &= \text{const.} \end{aligned} \quad (3.2)$$

At first glance it seems like selecting for the most efficient and at the same time highest substrate uptake rate is straightforward, as this will lead to the fastest growing phenotype. However, this line of thought assumes that the rate and the ATP-stoichiometry in a pathway can be increased independent of each other. But the loss of thermodynamic driving force by an increased ATP stoichiometry affects the rate of a pathway e.g. for irreversible reactions that operate far from chemical equilibrium.

Such correlation can be observed in nature between the yield and rate of cellular pathways, usually referred to as rate-yield trade-off hypothesis^{11,12}. One well-known example of a rate-yield tradeoff is alcoholic fermentation in the presence of oxygen in some yeast species, the so called Crabtree effect¹³ that leads to a high rate of ATP formation, but a low ATP yield compared to respiration.

This essentially means that nature has to make a compromise between ATP stoichiometry and rate in order to achieve a maximal growth rate. The rate-yield trade-off at a given (here the maximal) growth rate can be represented as isocline in the rate-yield space (Figure 3.1A, black line).

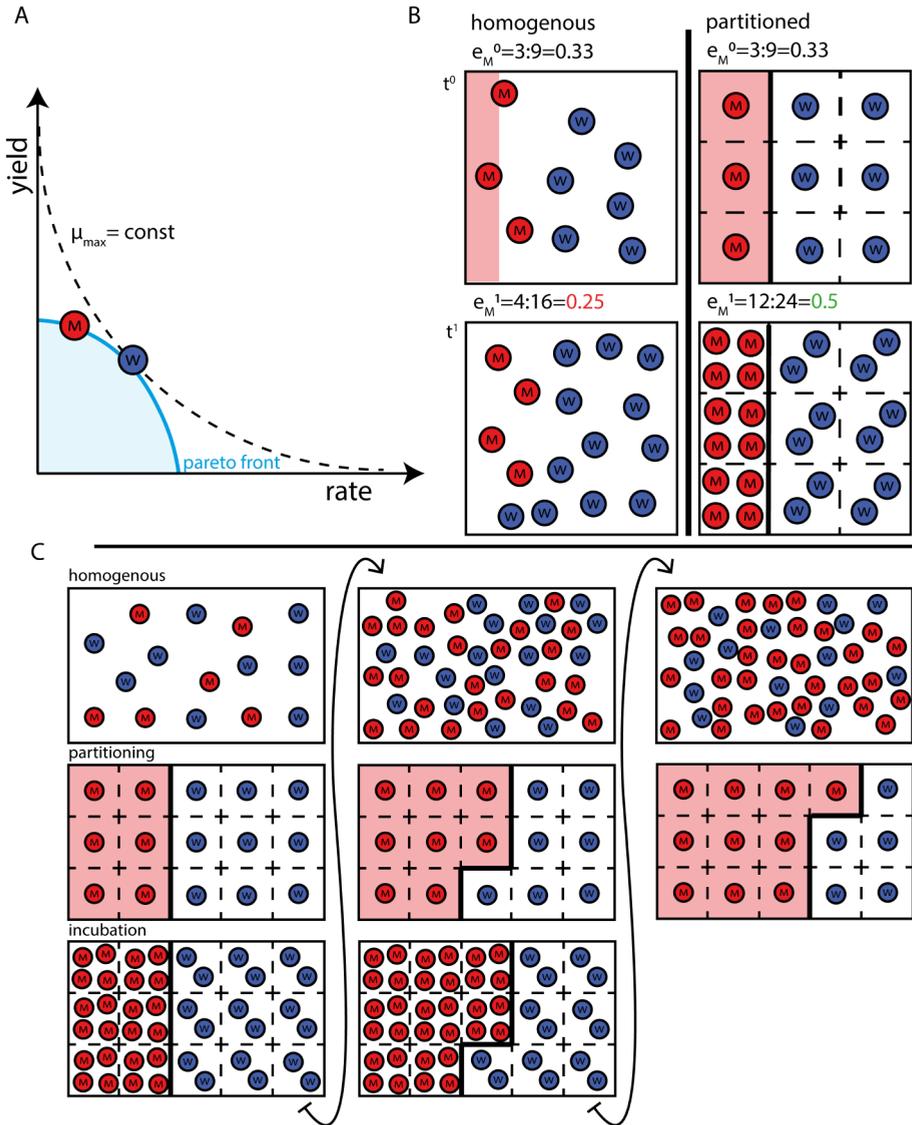


Figure 3.1. A) Pareto front with respect to rate and yield and its isocline. B) Left: Homogeneous cultivation using a common substrate – the wild-type strain (W) reaches a higher growth rate than the mutant (M) and therefore can claim more of the substrate (white compared to light red area) and produce more offspring despite lower biomass yield. Right: Partitioned cultivation – each strain has its own resource and efficient (high yield) mutants (M) generate more offspring compared to the W phenotype. C) Repetitive transfer of the partitioned culture upon depletion of the substrate can enrich the mutant population.

The actual shape of the organism's rate-yield trade-off (Figure 3.1A, red line) is not known *a priori* and can also be discontinuous. The point(s) of contact with the μ_{\max}

isocline represent the pareto-optimal point(s), where the strain reaches its maximum possible growth rate. This representation highlights that evolution that selects for maximal growth rate will result in such pareto-optimal phenotypes (here called W) subject to the growth rate and will therefore likely not result in strains with a maximal ATP-yield (here called M)¹⁴.

This poses an inherent limitation to select for a more efficient phenotype using classical laboratory evolutionary engineering approaches.

To resolve this issue, an approach that allows to partition substrate to different cell populations independent of their respective growth rate has been proposed¹⁵. This eliminates the direct competition for substrate and therewith eliminates the selective pressure for the highest growth rate. In contrast cells with a higher n_{ATP} will (albeit slower growth) produce a higher number of offspring from the same amount of available substrate¹⁶ (Figure 3.1B). Repeated compartmentalized growth (Figure 3.1C) will thus lead to a successive enrichment of more efficient phenotypes. The method has successfully shown that *Lactococcus lactis* populations can be enriched under aerobic, partitioned cultivation conditions. Here, this approach is established for anaerobic cultures of *S. cerevisiae*.

A mathematical model for partitioned cultivation of strains is developed, which is used to investigate the ideal properties of the approach and predict its efficiency under ideal conditions. Based on rational considerations and observations different mechanism that have an influence on the enrichment characteristics under experimentally relevant conditions like cell clustering and multiple partitioning of droplets are introduced into a rigorous model for selection. The model is validated experimentally using two strains with known difference in ATP-yield. Furthermore, the applicability of the approach is studied for enrichments from very low mutation frequencies using the model.

Model description

Enrichment efficiency for ideal partitioned cultivation of cells

For the derivation of the model system, it is assumed that there are two cell types:

- (1.) A reference, here termed wild-type (WT) phenotype, described by a growth rate μ_{WT} and a biomass (number) yield Y_{WT} and
- (2.) a mutant strain with, in comparison to WT, lower growth rate μ_M and higher biomass yield Y_M .

The overall number of cells in the system x^0 at the beginning of an incubation step is composed of the two strains and represented by the variables x_{WT}^0 and x_M^0 .

$$x^0 = x_{WT}^0 + x_M^0 \quad (3.3)$$

At the start of the enrichment experiment, the enrichment e_M^0 of the mutant strain is defined as:

$$e_M^0 = \frac{x_M^0}{x_{WT}^0 + x_M^0}, \quad e_{WT}^0 = 1 - e_M^0 \quad (3.4)$$

The cells are homogenously distributed within the accessible culture volume V . This volume (and consequently the contained substrate) is then partitioned between the two populations M and WT, where the volumes partitioned to the respective populations are defined as V_{WT} and V_M .

$$V = V_M + V_{WT} \quad (3.5)$$

In order to express which amount of the accessible volume is partitioned to every population, the volume fraction k is defined.

$$k = \frac{V_M}{V} \quad (3.6)$$

Under ideal conditions, every cell population will obtain a volume proportional to its cell number and k will be equal to the initial enrichment e_M^0 .

$$k = \frac{x_M^0}{x^0} = e_M^0 \quad (3.7)$$

It should be noted that this will only hold true under ideal conditions, in contrast to the non-ideal cases that will be addressed later.

As a next step, the partitioned system with the initial number of cells is incubated until all accessible substrate is consumed by the cells, leading to an increase in cell numbers. The increase for the respective populations will depend on the available volume, the substrate concentration c_s^0 and their apparent biomass (number) yield Y .

$$\begin{aligned} x_M^1 &= x_M^0 + \Delta x_M \quad \text{with} \quad \Delta x_M = Y_M c_s^0 k V \\ x_{WT}^1 &= x_{WT}^0 + \Delta x_{WT} \quad \text{with} \quad \Delta x_{WT} = Y_{WT} c_s^0 (1-k) V \\ \Delta x &= \Delta x_M + \Delta x_{WT} \end{aligned} \quad (3.8)$$

The number of offspring only depends on the yield and is independent of the growth rate of the respective cell population assuming that the ATP expenditure for maintenance is negligible compared to growth¹⁷. After the incubation, the reached enrichment e^1 is calculated from the population cell numbers at the end of the cultivation (indicated by superscript 1),

$$e_M^1 = \frac{x_M^1}{x_M^1 + x_{WT}^1} \quad (3.9)$$

and with the initial enrichment, the successive increase in enrichment for the incubation:

$$\Delta e_M = e_M^1 - e_M^0 \quad (3.10)$$

Substituting equations (3.3), (3.8) and (3.9) in (3.10) and leads to an equation for Δe_M as a function of the initial enrichment, which can be rewritten in following to facilitate the interpretation:

$$\Delta e_M = \frac{x_M^0 + \Delta x_M}{x^0 + \Delta x} - e_M^0 = \frac{\Delta x_M - e_M^0 \Delta x}{1 + \frac{x^0}{\Delta x}} \quad (3.11)$$

The nominator expresses the capability of the system to enrich the culture; this will only be the case if the mutant is able to produce more offspring compared to the wild type. On the other hand, the enrichment in the system will only increase beyond

e_M^0 when the ratio of mutant offspring to all offspring is higher compared the starting enrichment. The physical meaning of the denominator is to correct for the initial cell population in the system i.e. mixing the offspring cells with the initial cell population; this will lower the achieved final enrichment systematically (under the premise that the nominator is positive).

In order to characterize the enrichment characteristics of a partitioned system, the point(s) where the system cannot be enriched from its initial enrichment are of particular interest, they will be further referred to as critical point(s).

$$\Delta e_M = 0 \quad (3.12)$$

This will always be the case when the nominator of eq. (3.11) is zero. This nominator is dimensionless and called enrichment efficiency (EE) from here on. The value of the enrichment efficiency is always between -1 and 1, where a positive value indicates that the mutant strain can be further enriched from the initial enrichment during incubation, whereas a negative value is obtained when the enrichment will decrease during incubation. If the number of initial cells is very small compared to the produced number of offspring, EE will converge to Δe .

$$EE = \frac{\Delta x_M}{\Delta x} - e_M^0 \quad (3.13)$$

In order to derive an expression for EE as a function of the model parameters Y_M and Y_{WT} the equation is extended using eq. (3.8).

$$EE = \frac{Y_M \cdot k}{Y_M \cdot k + Y_{WT} \cdot (1 - k)} - e_M^0 \quad (3.14)$$

As the enrichment efficiency of the system does not depend on the absolute values of the yield but only on the relative yield difference between the different cell types, the equation can be simplified further by substituting:

$$\gamma = \frac{Y_M}{Y_{WT}} - 1 \quad (3.15)$$

Leading to final expression for EE after expansion:

$$EE = -\frac{e_M^0 - k - \gamma k + e_M^0 \gamma k}{\gamma k + 1} \quad (3.16)$$

It can be seen that the enrichment efficiency is only a function of e_M^0 and γ , and therewith independent of the design variables x_0 and c_s^0 . This makes it suitable to compare the performance of the system independent of the design variables. From the roots of eq. (3.16) the critical initial enrichment can be derived as a function of k . This value denotes a zero in the EE at the initial enrichment e_M^0 showing that the system cannot be further enriched at this point (e.g. like an azeotrope in distillation).

$$e_{M,crit}^0 = \frac{k + \gamma k}{1 + \gamma k} \quad (3.17)$$

Vice versa also the critical k as a function of e_M^0 can be derived.

$$k_{crit} = \frac{e_M^0}{1 + \gamma - e_M^0 \gamma} \quad (3.18)$$

Lastly, also the equation for Δe as a function of EE can be derived, where the increase in enrichment for a number of transfers can be calculated recursively for a number of incubations.

$$\Delta e = \frac{EE}{1 + \frac{x_0}{c_s^0 V (Y_{WT} k + Y_M (1 - k))}} \quad (3.19)$$

Extension to non-ideal partitioned systems

The model so far did not include any provisions on how the partitioning of the system is achieved in practice, instead it was assumed that the system is ideally partitioned (see eq. (3.7)).

In a real system, this assumption does not necessarily hold true as droplets are used as a means to partition the cell populations into volumes. In droplets, ideal

partitioning can only be achieved under two premises; (1) the droplets are only occupied with one cell or empty and (2) the probability of a cell to be partitioned into a particular droplet is independent of the population it belongs to. Whereas the latter requirement is admissible for cells with similar morphology, the first one is not. To reflect this behavior, the model is extended in order to describe two effects leading to droplets with multiple occupancy; clustering of cells and multiple partitioning (See Figure 3.2).

These extensions can be introduced conveniently into the framework and will not require changes in the equations as derived, but instead only require a different definition of k .

Modelling the influence of cell clusters

The formation of cell clusters caused by the missing separation of cells after division has been observed in directed evolution experiments of *S. cerevisiae*¹⁸, where its occurrence obstructs the enrichment of the desired phenotype. Especially when the environment favours cell clustering e.g. in a sequential batch reactor cell clusters are retained in the bioreactor thus having a selective advantage.

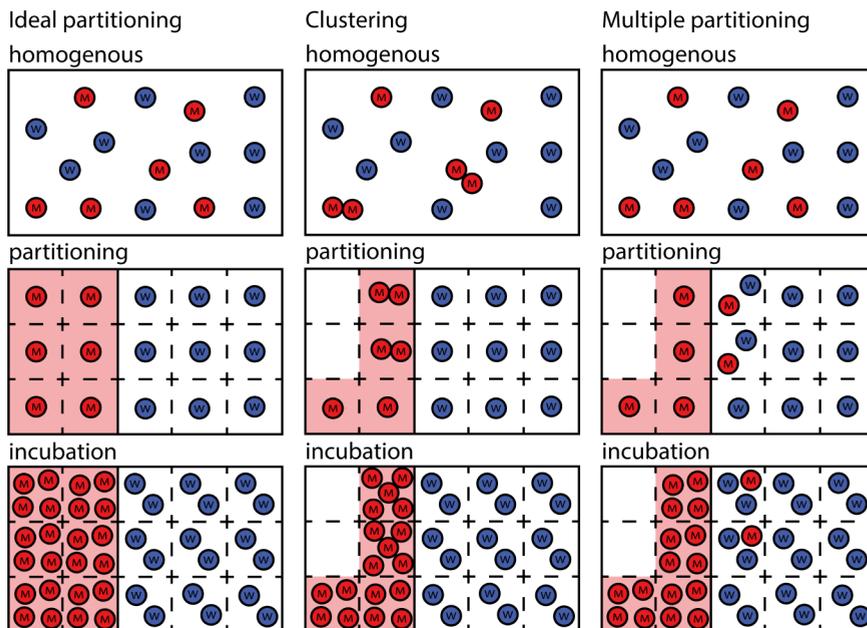


Figure 3.2. Schematic comparison of ideal and non-ideal partitioning

When cells form agglomerates, a discrepancy between the actual and the apparent number of cells in a population will arise. Recalling eq. (3.7), in the ideal case the

volume partitioned to the mutant population is proportional to the number of the initial mutant population. This means that cell clustering will lead to a systematically lower k compared to e_M^0 .

To reflect this behaviour the single cell ratio κ is introduced, describing the deviation from the ideal system by taking into account the deviation between the actual x^0 and the apparent cell number \tilde{x}_M^0 . In case of an optimal partitioned system, κ will be 1, otherwise smaller.

$$\kappa = \frac{\tilde{x}_M^0}{x_M^0} \quad (3.20)$$

This leads to the following equation for k .

$$k = \kappa e_M^0 \quad (3.21)$$

It is assumed that the wild-type population forms an insignificant amount of clusters. Moreover, it is assumed that cell clusters do not interfere with the droplet formation and the partitioning of the cell clusters into the droplets is independent of the cluster size.

Modelling the influence of multiple partitioned droplets

The generation of droplets leads to Poisson distributed droplet occupancies, where the mean is the average occupancy for a droplet. Additionally, the created droplet volume distribution is not uniform; this means also the average occupancy is a function of the droplet volume. Also, in contrast to clustering, this is a purely statistical process, i.e. cells belonging to different populations can be partitioned into one droplet, which in turn requires an analytic description for the direct competition for substrate in a droplet.

k can be predicted for a droplet volume distribution knowing the initial cell number per volume and the initial enrichment of the system. The droplet volume distribution consists of a number of bins characterized by the mean droplet volume in the bin, $V_{d,bin}$ and the number of droplets in the bin, n_{bin} . The sum of all bins resembles the total droplet volume, V_d . It is important to note that V_d is typically not equal to the accessible volume V , as V_d contains empty droplets. Empty droplet volume is not accessible to any cell population and therefore excluded from V_d . In contrast to the ideal model, it is no longer assumed that all substrate will be consumed at the

end of the incubation, which can also lead to a systematically lower V in comparison to V_d .

The volume fraction of each bin f_{bin} is determined by:

$$f_{bin} = \frac{n_{d,bin} V_{d,bin}}{V_d} \quad (3.22)$$

$$\text{with } V_d = \sum_{bin} n_{d,bin} V_{d,bin}$$

For a large number of droplets, the cells will be partitioned into the droplets following a Poisson distribution. The number of cells in a droplet is described by the variable occ where occ_{bin}^* is the average occupancy of a particular bin.

$$f_{occ,bin} = \frac{occ_{bin}^*}{occ!} e^{-occ_{bin}^*} f_{bin} \quad \text{for } occ = 0, 1, 2, 3, \dots \quad (3.23)$$

$$\text{with } occ_{bin}^* = \frac{x^0 V_{d,bin}}{V_d} = c_x^0 \cdot V_{d,bin}$$

The volume fraction for all occupancies is calculated by summation of all bins of the volume distribution.

$$f_{occ} = \sum_{bin} f_{occ,bin} \quad (3.24)$$

In order to calculate the split ratio k the volume fractions accessible for every population have to be determined. Here it has to be noted that the distribution is not selective for a specific cell (mutant or wild-type), but a random process. For example a double occupancy of a droplet can result in two wild type cells, two mutant cells or one each.

To describe the distribution of the volume for all different possibilities, the fractions \mathcal{E} are calculated from a binomial distribution.

$$\varepsilon_{M,occ,occ_M} = \binom{occ}{occ_M} \left(e_M^0 \right)^{occ_M} \left(e_{WT}^0 \right)^{occ-occ_M} \quad (3.25)$$

for $occ_M = 0, 1, 2, 3, \dots, occ_{max}$

In order to derive k it needs to be taken into account that the volumes of the multiple occupied droplets are shared by the respective cells in such droplets i.e. for each population M and WT.

Therefore, the fractions $\omega_{M,bin,occ_M,occ}$ and $\omega_{W,occ_M,occ,bin}$ are introduced and the split ratio k can be described by:

$$k = \kappa \cdot \frac{\sum_{bin} \sum_{occ} \sum_{occ_M} f_{occ,occ_M,bin} \cdot \varepsilon_{M,occ,occ_M} \cdot \omega_{M,occ,occ_M,bin}}{\sum_{bin} \sum_{occ} \sum_{occ_M} f_{occ,occ_M,bin} \cdot \varepsilon_{M,occ,occ_M} \cdot \omega_{M,occ,occ_M,bin} + \sum_{bin} \sum_{occ} \sum_{occ_M} f_{occ,occ_M,bin} \cdot (1 - \varepsilon_{M,occ,occ_M}) \cdot \omega_{M,occ,occ_M,bin}} \quad (3.26)$$

Also the accessible volume from the total droplet volume can be determined now.

$$V = \left(\sum_{bin} \sum_{occ} \sum_{occ_M} f_{occ,occ_M,bin} \cdot \varepsilon_{M,occ,occ_M} \cdot \omega_{M,occ,occ_M,bin} + \sum_{bin} \sum_{occ} \sum_{occ_M} f_{occ,occ_M,bin} \cdot (1 - \varepsilon_{M,occ,occ_M}) \cdot \omega_{M,occ,occ_M,bin} \right) \cdot V_d \quad (3.27)$$

In order to derive ω , a model that is able to describe the substrate competition for the two populations in a droplet is needed. Here, simple Monod kinetics are used to simulate every bin of the droplet volume distribution for all possible initial occupancies;

$$\begin{aligned} \frac{dx_{M,bin}^d}{dt} &= \mu_M x_{M,bin}^d & \mu_M &= \frac{\mu_{max,M} c_{S,bin}^d}{K_{S,M} + c_{S,bin}} \\ \frac{dx_{WT,bin}^d}{dt} &= \mu_{WT} x_{WT,bin}^d & \mu_{WT} &= \frac{\mu_{max,WT} c_{S,bin}^d}{K_{S,WT} + c_{S,bin}} \\ \frac{dc_{S,bin}^d}{dt} &= \left(-\mu_M \cdot \frac{x_{M,bin}^d}{Y_M} - \mu_{WT} \cdot \frac{x_{WT,bin}^d}{Y_{WT}} \right) \end{aligned} \quad (3.28)$$

With the respective initial conditions:

$$\begin{aligned} x_{M,bin}^{d,0} &= occ_M \\ x_{WT,bin}^{d,0} &= occ - occ_M \\ c_{S,bin}^d &= c_S^0 \end{aligned} \quad (3.29)$$

Where the ω calculate to:

$$\omega_{M,occ,occ_M,bin} = \frac{\int_0^{t_1} \frac{\mu_M \cdot x_{M,bin}^d}{Y_M} dt}{c_S^0} \quad (3.30)$$

$$\omega_{W,occ,occ_M,bin} = \frac{\int_0^{t_1} \frac{\mu_W \cdot x_{WT,bin}^d}{Y_{WT}} dt}{c_S^0}$$

Compared to the ideal model, this model does not make the presumption that all substrate is depleted at the end of the incubation. The effect of not completely consumed substrate, e.g. if the incubation time in large droplets is not sufficient to consume all the substrate is taken into account. In case all substrate in a mixed occupied droplet has been depleted, the following simpler relation will hold.

$$\omega_{WT,occ,occ_M,bin} = 1 - \omega_{M,occ,occ_M,bin} \quad (3.31)$$

Experimental methods

Measurement of the droplet size distribution

The droplet size distribution was photographed using a digital camera with a light microscope at 400x magnification. Subsequently the ImageJ image processing software was used, where the threshold color of the picture was adjusted to increase the visibility of the droplets. The droplet areas were then estimated using the inbuilt 'Analyze Particles' function. The conversion from pixels of the picture to length was calibrated using a stage micrometer. Assuming a spherical droplet shape, the determined areas were converted to volumes and binned into a droplet volume distribution using Microsoft Excel.

Strains

Two strains were used in this work, the reference strain CEN.PK113-7D^{19,20} (WT) and a strain with increased free energy conservation derived from CEN.PK113-7D, IMX771 (M). IMX771 expresses a heterologous maltose phosphorylase gene from *Lactobacillus sanfranciscensis*, the native *MAL11* maltose-transporter gene as the sole gene for maltose transport over the cell membrane and the beta-phosphoglucomutase of *Lactococcus lactis*. Strain IMX771 was derived from

Saccharomyces cerevisiae strain IMZ226⁵ that was transformed²¹ using a CAN1::KanMX deletion cassette²² in order to create a G418 resistant strain. Mutants were selected on solid YP medium (demineralized water, 10 g/L Bacto yeast extract, 20 g/L Bacto peptone, 2% (w/v) agar), supplemented with 200 mg/l G418 and the genotype was confirmed using PCR. IMX771 was kindly provided by the Industrial Microbiology group of Delft University of Technology. Both strains were stocked as working cell bank in 20% glycerol at -80°C.

Cultivation conditions

Pre-cultures were prepared in shake flasks using synthetic medium containing 17.3g/L maltose, 5g/L (NH₄)₂SO₄, 3g/L KH₂PO₄, 0.5g/L MgSO₄·7H₂O, 1g/L trace element solution²³ and 1g/L vitamin solution²⁴.

Enrichment in droplets were performed using synthetic medium as previously described^{25,26} with a maltose concentration of 34 g/L maltose, a TWEEN80 concentration of 0.78g/L and an ergosterol concentration of 0.015g/L.

YP maltose plates were used for determination of CFU and contained 10 g/l BD Bacto yeast extract, 20g/L Peptone, 15g/L Agar and 17.3g/L maltose. The required amounts of yeast extract, peptone and agar were dissolved in 90% of the total volume and autoclaved at 121°C. The required amount of maltose was dissolved in 10% of the final volume and added after filter sterilization. YP maltose G418 plates, used for selection of IMX771 were prepared the same way as YP maltose plates, but supplemented with 200 mg/L G418. All chemicals were purchased at Sigma-Aldrich.

Droplet cultivation system

Incubation of a mixed population containing CEN.PK113-7D and IMX771 in droplets was performed in 5 ml Eppendorf tubes (Eppendorf, Germany). A hole was drilled into the cap of 5ml Eppendorf tubes where a sterile filter (Whatman Puradisc FP 30 CA-S pore size 0.2 µm) was inserted in order to ensure sterile gas exchange with the (anaerobic) environment.

The tubes with filter units were autoclaved at 121°C for 20min before use. The Eppendorf tubes were then filled with 1300µl 3M Novec HFE-7500 including 0.2% Pico-Surf surfactant (Dolomite microfluidics, UK) and stored in an anaerobic chamber for at least 24h prior to use, in order to minimize the oxygen content of the oil. 700 µl of cell suspension was then added to the tube and vortexed for two minutes in order to create the emulsion. The Eppendorf tubes were incubated at 30°C in an anaerobic chamber for 72h.

To break the emulsion 600µl of PFOH (1H,1H,2H,2H-perfluorooctan-1-ol) was added. After vortexing for 2 minutes, the tube was set aside for at least 10 minutes to separate the phases. The required amount for analysis was withdrawn and the remaining volume was diluted back to the desired initial cell concentration and the process was repeated in a new tube.

Analytical methods

At the end of every incubation, a series of 10-fold dilutions in medium without trace elements, vitamins and maltose was prepared in duplicate to a total volume of 1 mL. 100 µL of the 10³ dilution were plated on YP maltose plates in duplicate and incubated for 3 days at 30C, whereupon the colonies were counted. Subsequently, the plates were replicated to YP maltose plates containing 0.2 mg/ml G418 using Velveteen squares and a Replica Plating Tool (both Bel-Art, US). The enrichment was calculated from the ratio between G418 resistant CFU and overall CFU.

To estimate the number of cells per gram dry weight, cells concentrations were determined using a Coulter counter (Z2 COULTER COUNTER, Beckman Coulter), where five independent measurements for a sample were performed and averaged. Dry weight concentration was determined by filtering a defined amount of cell suspension on a cellulose filter (pore size 0.22µm) and drying the filter at 105°C.

Optical cell density (OD) was determined using a spectrophotometer (Biochrom Libra S11) at a wavelength of 600 nm. Samples were diluted if needed such that the measurement was between OD of 0.05 and 0.6.

Results

Ideal performance and impact of suboptimal partitioning on the enrichment characteristics

The ideal system description allows to investigate the best possible performance of the approach by calculating the enrichment efficiency (EE) using eq. (3.16). The ideal system always shows a parabolic shape with roots only at 0 and 1 (see Figure 3.3, left panel), this means the ideally partitioned system will always allow for a complete enrichment regardless of the initial enrichment. For the relevant differences in biomass yield, the enrichment efficiency exhibits a maximum at an initial enrichment of around 0.5. Consequently, the enrichment is less efficient at low and high enrichments resulting in a sigmoidal shape of the enrichment as a function of the number of transfers (see Figure 3.6). Moreover, the enrichment

efficiency also decreases globally with decreasing difference in yield; this means that more transfers will be necessary in order to achieve the same increase in enrichment.

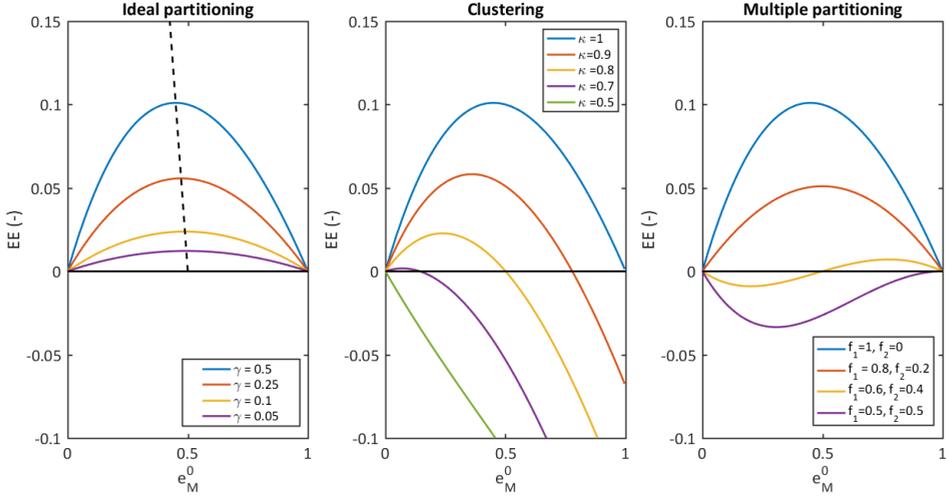


Figure 3.3. Enrichment efficiency (EE) as a function of the initial enrichment (e_M^0); Left: for the ideal partitioned system for different yield differences (γ). The dashed line indicates the trajectory for the maximal EE (for derivation see Appendix). Middle: Effect of cell clustering for different single cell ratios (κ) for $\gamma=0.5$. Right: Effect of different volume fractions of single and double occupied droplets (f_1 and f_2) on the enrichment efficiency for $\gamma=0.5$.

However, the partitioning may not be optimal and, in the case of cell clustering (see Figure 3.2), it leads to a global decrease in EE . Moreover, it can be observed that a mutant population exhibiting clustering cannot reach a full enrichment anymore. This effect increases with increased clustering, respectively lower single cell ratios.

By substituting eq. (27) in (23), the critical clustering factor for certain target enrichment e_M^0 can be derived. A function of the dimensionless yield difference is obtained:

$$\kappa_{crit} = \frac{1}{-e_M^0 \gamma + \gamma + 1} \quad (3.32)$$

For $e_M^0 = 0$, the single cell ratio below which no enrichment (κ_{min}) will be achieved can be found.

$$\kappa_{\min} = \frac{1}{\gamma + 1} \quad (3.33)$$

This demonstrates that the method is robust against clustering, which is a useful property, i.e. the method will not enrich for phenotypes having a stronger tendency to cluster than that of the reference population.

To qualitatively investigate the effect of multiple partitioning (Figure 3.3, right), simulations were performed under the following assumptions/simplifications:

- (1.) All droplet volume (V_d) is accessible and either single or double occupied, neglecting any higher occupancies;
- (2.) A droplet with double occupancy is assumed to be fully dominated by a present WT cell. This assumption reflects the worst-case enrichment caused by the occupancy effect and reduces the calculation complexity by setting all ω to either 0 or 1.
- (3.) There is no clustering of cells.

With these assumptions, k simplifies to:

$$k = f_1 \cdot e_M^0 + f_2 (e_M^0 \cdot e_M^0) \quad (3.34)$$

In contrast to the clustering effect, it is observed that multiple occupied droplets have a more critical effect on the performance of the system. Multiple occupancies reduce the enrichment efficiency primarily at low enrichments, where it can also introduce additional roots to EE . This behavior is a result of the small fraction of the double occupied volumes with mutant cells at low enrichment. This fraction increases (in this case quadratic) with increasing enrichment and hinders the capability of enriching a mutant population from a low initial enrichment, which is the main objective of the method.

Model driven design of an enrichment experiment

The modelling approach allows predicting, understanding and designing the trade-offs faced in partitioned cultivation systems. The enrichment characteristics of the system depends on six variables, (1) the yield difference γ , (2) the droplet volume distribution, (3) the enrichment at the beginning e^{start} and the target enrichment to be achieved in the end of the experiment e^{end} . Moreover, (4) the initial cell

concentration in the system $c_x^0 = \frac{x^0}{V_d}$ (see eq.(3.23)), (5) the initial substrate

concentration in the droplets c_s^0 , and (6) the incubation time per transfer t^1 .

The yield difference is determined by the individual cell populations in the system. The droplet volume distribution is determined by the experimental procedure and it can only be influenced to a minor extend the approach used here. The enrichment of the initial population can usually not be chosen. On the other hand, the target enrichment can be set and should be high enough to allow for extraction of cells with the desired phenotype. This means only the last three parameters can be chosen in an experiment, where the incubation time is not independent with respect to the chosen substrate concentration. This leads to two remaining independent design variables, c_s^0 and c_x^0 (see also Figure 3.4).

For practical reasons the transfer the culture is set to $t^1 = 72h$, in order to achieve the maximal enrichment Δe for each transfer transfers (see eq. (3.19)) a maximal c_s^0 should be chosen to minimize the number of transfers.

$$c_{s,opt}^0 = \arg \max c_s^0$$

such that (3.35)

$$\begin{cases} \omega_M + \omega_{WT} > 0.95 \\ c_s^0 < c_{s,max}^0 \end{cases}$$

However, some constraints have to be taken into account: The wild type population grows faster compared to the mutant population, it will deplete their substrate faster and therewith temporarily produce more offspring. If the experiment is stopped before all substrate is depleted, this will influence the EE negatively. Therefore the first constraint will enforce that at least 95% of the accessible substrate in the droplets will be depleted at the chosen incubation time. The second constraint ensures that the concentrations in the droplet are within physiological limits, in the case of ethanol buildup as in our anaerobic system this will eventually lead to growth inhibition²⁷ and is taken into account using the respective substrate concentration $c_{s,max}^0$. Other effects could be e.g. substrate repression at high substrate concentrations. A feasible solution was found for an initial maltose concentration of $c_s^0 = 34g/l$.

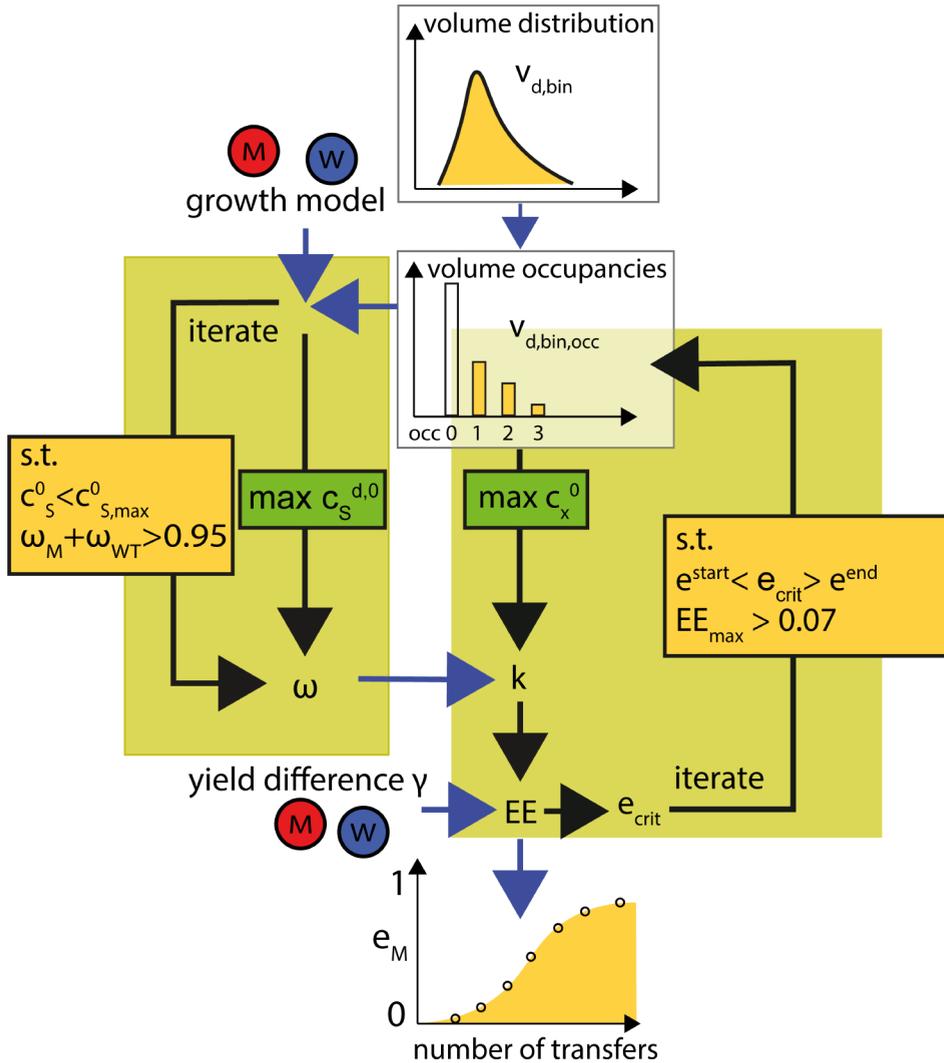


Figure 3.4. Schematic representation of the experimental design workflow

Moreover, the optimization yields a full set of ω which also determine k and therewith the enrichment efficiency. The values for droplet occupancies up to three cells are shown for the droplet volumes of 12 pL, the median droplet diameter of the distribution 70.9 pL and 330 pL (see Table 3.1). The values show that in multiple occupied droplets the mutant is only able to gather a minor share of the substrate in direct competition with the WT, therewith decreasing the enrichment efficiency. The effect is more pronounced in larger droplets, due to the exponential growth of the cells which leverages the advantage for the faster growing cell(s). This means

that the assumptions made in the previous qualitative prediction, i.e. as soon as there is a WT cell in a droplet it owns the complete volume, are admissible if the growth rate of the mutant strain is not known. In any case this is a worst-case assumption that will always underestimate the performance of the system where the selection of $c_S^{d,0}$ with respect to t^1 creates a cutoff for the growth rate of a mutant phenotype and it has to be selected accordingly.

Table 3.1. $\omega_{M,occ_M,occ}$ for the droplet volumes 12pL/70.9pL/330pL

Droplet occupancy		
1	2	3
$\omega_{M,0,0} = 0/0/0$	$\omega_{M,0,2} = 0/0/0$	$\omega_{M,0,3} = 0/0/0$
$\omega_{M,1,1} = 1/1/1$	$\omega_{M,1,2} =$ 0.1573/0.1066/0.0622	$\omega_{M,1,3} =$ 0.0900/0.0677/0.0428
	$\omega_{M,2,2} = 1/1/1$	$\omega_{M,2,3} =$ 0.2751/0.1983/0.1202
		$\omega_{M,3,3} = 1/1/1$

After determining the substrate concentration, the initial cell number c_x^0 can be estimated which will determine primarily the volume occupancies.

As already mentioned, the droplets occupancies will be Poisson distributed with their average occupancy occ_{bin}^* . Looking at equation (3.23), the average occupancy depends on the droplet volume, which leads to the effect that larger droplets have a higher average occupancy and are therewith more likely to be multiple occupied compared to smaller droplets (see also Figure 3.5). Keeping in mind that the volume increases with the order of three as a function of the droplet diameter, this means that even a low number of large droplets will contain significant fraction of the total droplet volume.

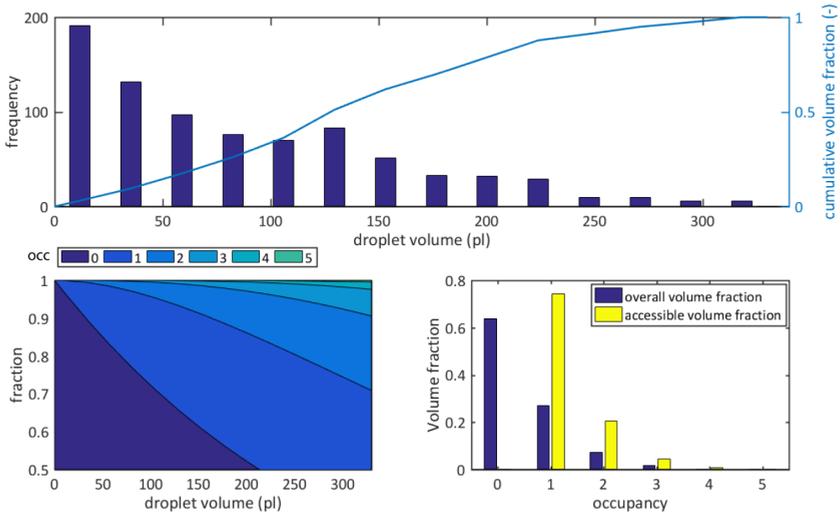


Figure 3.5. Top: Measured droplet volume distribution; Left: Predicted droplet occupancies as a function of droplet volume for $c_{x,opt}^0$; Right: Predicted overall and accessible volume fractions for the different occupancies over the measured droplet volume distribution for $c_{x,opt}^0$.

Large droplets will limit the initial cell concentration in order to keep the multiple occupied volume fractions small. This means on the other hand that that small droplets will be mostly non-occupied, limiting the volumetric mutant throughput of the method compared to a more uniform distribution. However, the major limiting effect for the throughput remains the Poisson distribution. Even in uniform droplets the average occupancy has to be chosen low to obtain mostly single occupied droplets²⁶.

With known volume fractions, the split ratio k can be calculated and consequently the enrichment efficiency.

Comparable to the previous optimization for the substrate concentration, another optimization is performed, to maximize the volumetric throughput of the experiment.

$$\begin{aligned}
 c_{x,opt}^0 &= \arg \max(c_x^0) \\
 &\text{such that} \\
 e^{start} &> e_{crit} < e^{end} \\
 EE_{max} &\geq 0.07
 \end{aligned}
 \tag{3.36}$$

A trade-off between throughput and performance has to be made, while ensuring feasibility of the method in the chosen range of enrichment. Here, the first constraint makes sure that there are no critical points e_{crit} in the chosen interval (i.e. roots in EE). Additionally, it has to be taken into account that this measure is not sufficient to ensure a desired performance of the system. To address this, a second constraint is introduced, limiting the admissible decrease in EE . Here also other criteria may be appropriate, but as the first constraint ensures a concave shape of the EE , the criterion can be expected to be sufficient in most cases. With the model an initial cell concentration c_x^0 of $3.25 \cdot 10^6$ cells/mL was estimated, corresponding to an initial OD of about 0.6.

Table 3.2. Parameters for the droplet enrichment simulation

Parameter	Value	Source (if applicable)
$\mu_{max,WT}$	0.28 h^{-1}	5
$\mu_{max,M}$	0.1 h^{-1}	5
K_S	4 mmol/L	28
$Y_{X/S,M}$	0.088 g _{DW} /g _{Maltose}	5
$Y_{X/S,WT}$	0.066 g _{DW} /g _{Maltose}	5
Y_M	$2.478 \cdot 10^9$ 1/g _{Maltose}	This work
Y_{WT}	$1.485 \cdot 10^9$ 1/g _{Maltose}	This work
γ	0.6687	calculated from eq. (3.15)
$c_{x,opt}^0$	$3.25 \cdot 10^6$ cells/ml; OD=0.6	from optimization
$c_{s,opt}^0$	34g/L	from optimization
$c_s^{0,max}$	240g/L	27
e^{start}	0.1	This work
e^{end}	0.95	This work

Experimental validation of the model

In order to validate the model, the predictions in enrichment for the designed experiment are compared to measurements from an experiment with the *S. cerevisiae* strains IMX771 and CEN.PK113-7D. The model simulations describe the found pattern within the experimental error (see Figure 3.6). Also, the loss in performance

due to multiple occupied droplets can be clearly seen when comparing to the predictions of the ideal system.

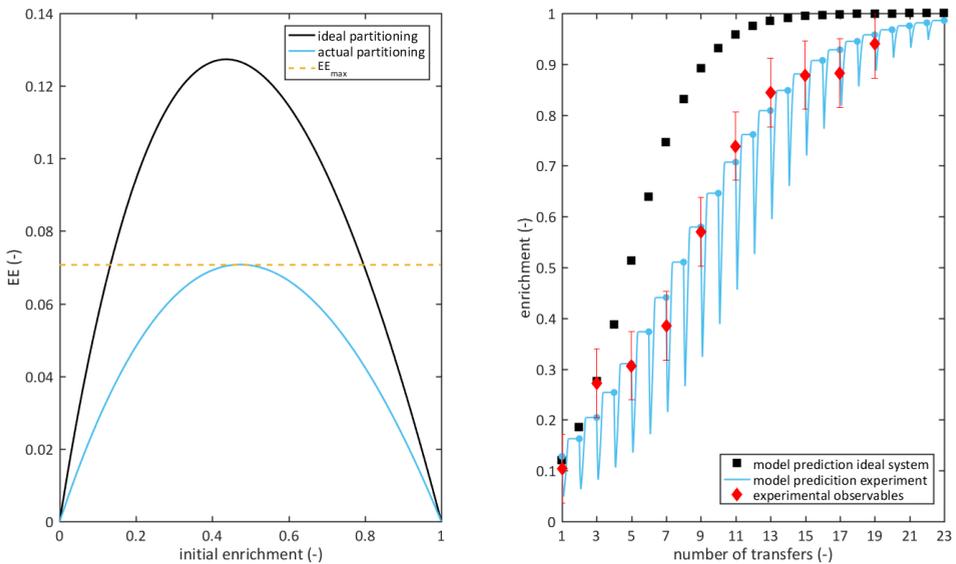


Figure 3.6. Left: Prediction of the enrichment efficiency for the ideal (blue) and the experimental droplet occupancies (red). Right: Simulation of the enrichment trajectory for the ideal system (black squares), Simulations (blue) for the experimental conditions with the transfers every 72h (blue dots) and experimental observations (red). The initial decrease in enrichment after a transfer is caused by the faster growth rate of the WT population.

Prediction of the enrichment from low mutation frequencies

In order to apply the approach for extraction of mutants with increased free energy conservation, the method must enrich from a potentially very low initial enrichment. The mutation frequencies depend on the objective, e.g. for auxotrophic *S. cerevisiae* they are found in the range of $5 \cdot 10^{-4}$ ²⁹, similar values have been observed for Streptomycin resistance³⁰.

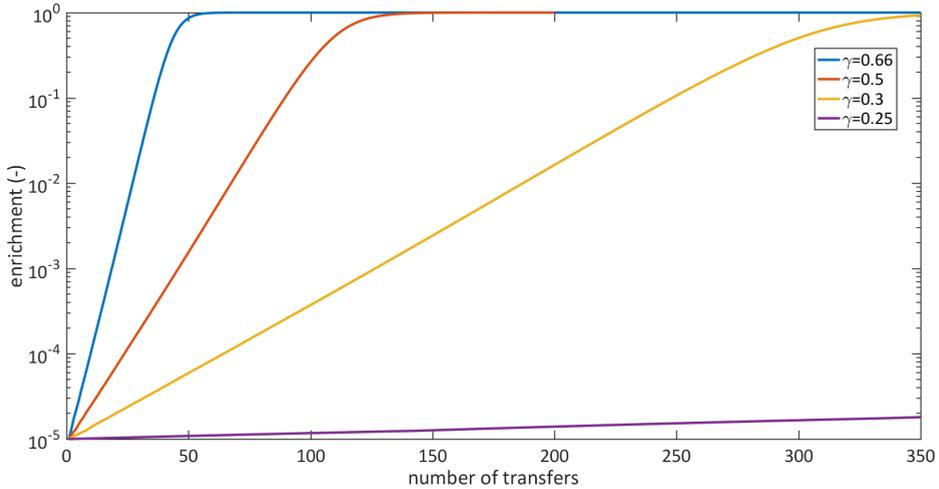


Figure 3.7. Model predictions for the enrichment characteristics starting from low initial enrichments ($e^{start} = 10^{-5}$) for different yield differences γ .

The calculations suggest that the current system requires approximately 10 transfers (720h) (see Figure 3.7) to increase the enrichment by one order of magnitude. For lower difference in biomass yield the approach the required number of transfers already reaches timescales of years for $\gamma=0.25$ (192 transfers per order of magnitude corresponding to 576 days).

Conclusions & Discussion

The aim of the current study was to assess the performance of enrichment for phenotypes with increased biomass yield in droplets.

A model with ideal partitioning characteristics was used to predict and assess the maximum performance of the approach. Also non-ideal partitioning i.e. existence of multiple occupied droplets in the system was taken into account. Furthermore, a non-uniform droplet volume distribution was integrated. Two effects were studied; (1) clustering of cells and (2) multiple partitioning of cells into one droplet, where both effects affect the performance of the system negatively. Here, the effect of multiple partitioning can be identified as more critical, as it decreases the performance of the system predominantly in the low enrichment range and can lead to the situation that the system cannot enrich anymore from low initial enrichments. On the other hand, the analysis of cell clustering showed that the droplet system is robust against phenotypes that exhibit cell clustering; this is an advantage

compared to non-partitioned systems for directed evolution, where this effect is an undesired outcome of an evolution experiment. This means partitioned systems could also be taken into consideration as a tool to select for phenotypes with reduced clustering.

A system containing defined and characterized *S. cerevisiae* strains with a known difference in ATP efficiency and thus biomass yield was researched. The model was used to design an experiment for our model system, requiring two constrained optimizations. First the maximization of the substrate concentration and second the trade-off between the volumetric throughput, i.e. the population size that can be screened and the performance of the system. The performance depends largely on the difference in biomass yield between the populations and the average droplet occupancies within the observed distribution in droplet volumes.

With the resulting operating conditions, the model was validated, where the model predictions could be reproduced within the experimental error. The model was also extrapolated to very low enrichments in the range of observed mutation frequencies for *S. cerevisiae*, where it would take about 10 transfers in order to increase the enrichment by one order of magnitude.

This demonstrates that the presented methodology is suited for directed evolution, however has limitations for lower differences in biomass yield between the populations where the number of transfers will rapidly increase with decreasing difference in biomass yield.

The performance in those ranges can only be increased by increasing k beyond the initial enrichment; this is not possible in the presented approach and would require an approach sorting the droplets for the contained number of offspring discarding the droplets with a lower number of offspring. However, such approach is technically significantly more complex and may not be competitive with respect to the mutant throughput.

Alternatively, one could think about methods penalizing faster growing cells e.g. lead to rapid loss in viability once the substrate is depleted. However, this also poses the risk of selecting false positive growth deficient phenotypes.

The droplet size distribution in the used experimental approach can hardly be influenced. However, also for approaches creating uniform droplets where the droplet occupancies are Poisson distributed, the average occupancy has to be chosen in a way that around 90% of the created droplets are empty. This means that the non-uniform distribution of droplet volumes is not the major limitation for throughput of the method. This could be improved using special microfluidic chips²⁶,

however it has to be noted that empty droplets do not have negative effects on the enrichment and therefore scaling the volume to increase the throughput is possible. An inherent characteristic of the partitioned cultivation is that the partitioning occurs at the cell level, leading to the enrichment of phenotypes with increased biomass number yield. It was found that the difference in number yield is higher compared to the biomass yield expressed per gram dry weight of cells (see Table 3.2), i.e. the system favors mutants having the same gravimetric biomass yield but exhibiting a higher number of cells per gram biomass dry weight. This effect leverages the enrichment and may be caused by the lower Y_{ATP} of slower growing cells due to a lower protein content¹⁷ (see also Chapter 5). However with the available data a correlation between energy efficient phenotypes and cell number yield cannot be generalized and should be further characterized in more detail. Currently the partitioned cultivation is labor intensive as the transfers are done by hand. To generate high throughput and allow parallelization automatization of the method would be desirable. With automatization this method has the potential to become a useful tool in selecting phenotypes with increased free energy conservation.

Symbols and Abbreviations

c	Concentration
e	Dimensionless enrichment
f	Volume fraction
EE	Enrichment efficiency
q_s	Specific substrate uptake rate
K	Affinity constant
k	Volume fraction accessible for mutant cells
n	Number of cells
n_{ATP}	ATP stoichiometry
occ	Occupancy of a droplet
RYT	Rate-yield trade-off hypothesis
V	Volume
x	Number of cells
\tilde{x}	Apparent number of cells
Y	Yield
ε	Fraction from probability distribution
μ	Growth rate

γ	relative yield difference
κ	Single cell ratio
ω	Volume fraction inside droplet

Subscript/Superscript

<i>bin</i>	Bin of a droplet volume distribution
<i>crit</i>	Critical
<i>d</i>	Droplet
<i>end</i>	End of experiment
<i>M</i>	Mutant (more ATP efficient)
<i>max</i>	maximal
<i>occ</i>	Droplet occupancy
<i>opt</i>	Optimized
<i>P</i>	Product
<i>S</i>	Substrate
<i>start</i>	Start of enrichment experiment
<i>WT</i>	Wild-type
<i>x</i>	Biomass
0	Start of incubation
1	End of incubation
*	Average

References

- 1 Straathof, A. J. J. in *Comprehensive Biotechnology (Second Edition)* 811-814 (Academic Press, 2011).
- 2 Straathof, A. J. J. Transformation of Biomass into Commodity Chemicals Using Enzymes or Cells. *Chemical Reviews* 114, 1871-1908 (2014).
- 3 Cueto-Rojas, H. F., van Maris, A. J., Wahl, S. A. & Heijnen, J. J. Thermodynamics-based design of microbial cell factories for anaerobic product formation. *Trends Biotechnol* 33, 534-546 (2015).
- 4 Ataman, M. & Hatzimanikatis, V. Heading in the right direction: thermodynamics-based network analysis and pathway engineering. *Current Opinion in Biotechnology* 36, 176-182 (2015).
- 5 de Kok, S. *et al.* Increasing free-energy (ATP) conservation in maltose-grown *Saccharomyces cerevisiae* by expression of a heterologous maltose phosphorylase. *Metab Eng* 13, 518-526 (2011).
- 6 de Kok, S., Kozak, B. U., Pronk, J. T. & van Maris, A. J. A. Energy coupling in *Saccharomyces cerevisiae*: selected opportunities for metabolic engineering. *FEMS Yeast Research* 12, 387-397 (2012).
- 7 Taymaz-Nikerel, E. e. a. in *Thermodynamics in Biochemical Engineering* 547-579 (EFPL Press, 2013).
- 8 Sauer, U. in *Metabolic Engineering* (eds Jens Nielsen *et al.*) 129-169 (Springer Berlin Heidelberg, 2001).
- 9 Wisselink, H. W., Toirkens, M. J., Wu, Q., Pronk, J. T. & van Maris, A. J. Novel evolutionary engineering approach for accelerated utilization of glucose, xylose, and arabinose mixtures by engineered *Saccharomyces cerevisiae* strains. *Applied and environmental microbiology* 75, 907-914 (2009).
- 10 Brennan, T. C. *et al.* Evolutionary engineering improves tolerance towards jet fuel replacements in yeast. *Applied and Environmental Microbiology*, AEM. 04144-04114 (2015).
- 11 Pfeiffer, T. & Morley, A. An evolutionary perspective on the Crabtree effect. *Frontiers in Molecular Biosciences* 1, 17 (2014).
- 12 Pfeiffer, T., Schuster, S. & Bonhoeffer, S. Cooperation and competition in the evolution of ATP-producing pathways. *Science* 292, 504-507 (2001).
- 13 Crabtree, H. G. The carbohydrate metabolism of certain pathological overgrowths. *Biochemical Journal* 22, 1289-1298 (1928).
- 14 Maja Novak, Thomas Pfeiffer, Richard E. Lenski, Uwe Sauer & Sebastian Bonhoeffer. Experimental Tests for an Evolutionary Trade-Off between Growth Rate and Yield in *E. coli*. *The American Naturalist* 168, 242-251 (2006).
- 15 Bachmann, H. *et al.* Availability of public goods shapes the evolution of competing metabolic strategies. *Proceedings of the National Academy of Sciences of the United States of America* 110, 14302-14307 (2013).

- 16 Bachmann, H. *et al.* Availability of public goods shapes the evolution of competing metabolic strategies. *Proceedings of the National Academy of Sciences* 110, 14302-14307 (2013).
- 17 Verduyn, C., Postma, E., Scheffers, W. A. & van Dijken, J. P. Energetics of *Saccharomyces cerevisiae* in anaerobic glucose-limited chemostat cultures. *J Gen Microbiol* 136, 405-412 (1990).
- 18 Oud, B. *et al.* Genome duplication and mutations in ACE2 cause multicellular, fast-sedimenting phenotypes in evolved *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* 110, E4223-4231 (2013).
- 19 Entian, K.-D. & Kötter, P. in *Methods in Microbiology* Vol. Volume 36 (eds Stansfield Ian & J. R. Stark Michael) 629-666 (Academic Press, 2007).
- 20 van Dijken, J. P. *et al.* An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. *Enzyme Microb Technol* 26, 706-714 (2000).
- 21 Daniel Gietz, R. & Woods, R. A. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods in Enzymology* 350, 87-96 (2002).
- 22 Guldener, U., Heck, S., Fielder, T., Beinhauer, J. & Hegemann, J. H. A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res* 24, 2519-2524 (1996).
- 23 Postma, E., Alexander Scheffers, W. & Van Dijken, J. P. Kinetics of growth and glucose transport in glucose-limited chemostat cultures of *Saccharomyces cerevisiae* CBS 8066. *Yeast* 5, 159-165 (1989).
- 24 Bruinenberg, P. M., de Bot, P. H. M., van Dijken, J. P. & Scheffers, W. A. The role of redox balances in the anaerobic fermentation of xylose by yeasts. *European journal of applied microbiology and biotechnology* 18, 287-292 (1983).
- 25 Canelas, A. B. *et al.* Quantitative Evaluation of Intracellular Metabolite Extraction Techniques for Yeast Metabolomics. *Analytical Chemistry* 81, 7379-7389 (2009).
- 26 Abate, A. R., Chen, C. H., Agresti, J. J. & Weitz, D. A. Beating Poisson encapsulation statistics using close-packed ordering. *Lab on a chip* 9, 2628-2631 (2009).
- 27 Casey, G. P. & Ingledew, W. M. M. Ethanol Tolerance in Yeasts. *CRC Critical Reviews in Microbiology* 13, 219-280 (1986).
- 28 Weusthuis, R. A., Adams, H., Scheffers, W. A. & van Dijken, J. P. Energetics and kinetics of maltose transport in *Saccharomyces cerevisiae*: a continuous culture study. *Appl Environ Microbiol* 59, 3102-3109 (1993).
- 29 Hashimoto, S. *et al.* Isolation of Auxotrophic Mutants of Diploid Industrial Yeast Strains after UV Mutagenesis. *Applied and Environmental Microbiology* 71, 312-319 (2005).

- 30 Kai, G. & Weber, A. S. Streptomycin-resistant mutant production in a continuous-flow UV mutation device. *Journal of industrial microbiology* 8, 107-112 (1991).

Supporting Appendix

The maximum of EE in the ideal system

Derivation of the maximum for the EE (eq. (3.16)) in the ideal system, i.e. $e_M^0 = k$ gives.

$$\frac{dEE}{de_M^0} = 1 - \frac{\gamma + 1}{(\gamma e_M^0 + 1)^2}$$

This means in the interval $0 < e_M^0 < 1$ there is only one maximum at:

$$e_M^0 = \frac{\sqrt{\gamma + 1} - 1}{\gamma}$$

This means for γ approaching zero the maximum is close to an e_M^0 of 0.5, whereas it asymptotically approaches 0 for $\lim_{\gamma \rightarrow \infty}$.

This behavior can be best understood looking at the generic definition of EE (see eq. (3.13)) where with increasing yield difference the first term dominates e_M^0 .

Chapter 4:

Exploring the links between energy metabolism, cellular physiology and protein allocation in the evolution of *S. cerevisiae* using structured mFBA modelling

in collaboration with S.A. Wahl and J.J. Heijnen

"The flow of energy through a system acts to organize that system."

Harold J. Morowitz

Abstract

Unravelling the mechanisms of evolution is crucial for the understanding of microbial ecology, but also in metabolic engineering where adaptive laboratory evolution is frequently used as a versatile tool for strain improvement.

Phenotypical evolutionary fitness, e.g. the growth rate, is the result of complex metabolic trade-offs, making it difficult to understand the relation between the environmental selective pressure and the resulting phenotype rationally. In these cases, a mechanistic model would allow understanding and predicting the outcome of laboratory evolution experiments.

A model for yeast combining kinetics, stoichiometric modelling and the allocation of cellular protein to describe the cellular physiology subject to morphological constraints, i.e. the limited availability of cell membrane surface, inner mitochondrial membrane surface and total protein content is derived and tested.

The model explicitly links the allocated cellular protein and the catalyzed metabolic rates, leading to a description of metabolic trade-offs subject to a metabolic objective e.g. growth rate and environmental variables e.g. substrate limitation. This makes the approach useful in predicting evolutionary trajectories and to infer purposeful adaptive laboratory evolution strategies for strain improvement.

The applicability of the approach is demonstrated by predicting optimal metabolic strategies in yeast under batch as well as in glucose and oxygen limited conditions. It was found that the model predicts the Crabtree effect in *S. cerevisiae*, where it is shown that the model exhibits a minimum threshold in mitochondrial protein content from whereupon ethanol fermentation can fully restore the maximal growth rate. It is hypothesized that this can explain the observed variety in the Crabtree effect among different yeast. Moreover the model shows that the occurrence of a Crabtree effect is strictly linked to the glucose import mechanism with the only change requiring the mutation from a glucose-proton symport to a glucose uniporter. The dramatic loss in glucose affinity associated with this change is considered and it is shown that the separation of a Crabtree positive phenotype from a Crabtree negative progenitor will require elevated glucose concentrations. Moreover the model shows that this step does not require oxygen limited or anaerobic conditions, as hypothesized previously. Furthermore the effect of increased total protein content compared to *S. cerevisiae* as observed in many Crabtree negative yeast on the evolutionary outcome is explored; here it is shown that with evolving a higher protein content such yeast can outcompete *S. cerevisiae* under all glucose and almost all oxygen limited conditions, including the niche it emerged from stressing the importance of ethanol tolerance. The model prediction is also applied to a practical

engineering target, homolactic acid fermentation, also known as Warburg effect in other organisms. The difference between the Crabtree effect and the Warburg effect (aerobic lactate production) is discussed under the premise of a limited availability of cell membrane surface. The fermentative pathways show the same ATP yield on glucose but have the distinction that ethanol can diffuse out of the cell, whereas lactate needs a transport protein. It is shown that the trade-off between the two transporters eliminates the solution space observed with the Crabtree effect i.e. a decrease in mitochondrial protein content can never restore the maximal growth rate anymore. Still, the model indicates that the directed evolution of a homolactic fermenting yeast, where rational knowledge is missing, should still be evolutionary favorable from a pyruvate decarboxylase (PDC) negative strain under oxygen limited conditions.

Introduction

Growth rate as selective pressure in evolution

According to life-history theory, growth rates are subject to strong directional selection due to reproductive and survival advantages^{1,2}. This holds in particular for exponentially growing populations of unicellular microorganisms, where faster growth allows assimilating more of the public substrate in the environment compared to slower growing competitors³.

Selection pressure is also frequently applied in metabolic engineering, where desirable phenotypes can often be linked to increased growth rate. This approach is commonly called adaptive laboratory evolution (ALE) or directed evolution⁴. The challenge is to design selective conditions that favor exclusively the growth of the desired phenotype⁵ to avoid undesired or false-positive outcomes. Such predictions are very challenging, as the interrelation between the environmental variables and the resulting growth rate is not known *a priori*. This is especially the case when the organism faces metabolic trade-offs, where the exact shape of the trade-off is not known (as discussed in Chapter 3).

A mechanistic model will allow predicting the outcomes of evolution and can provide a rational basis for the intrinsic understanding of the connection between the imposed environmental variables and the resulting growth rate. To do so the model has to reflect relevant mechanism determining the growth rate.

The link between energy metabolism, protein allocation and growth

Growth of organisms requires large amounts of energy to synthesize cellular components. This energy has to be gained from the Gibbs free energy released in the

catabolism of substrate. This cellular energy is exchanged between catabolism and anabolism using the high energy phosphoanhydride bonds of nucleoside triphosphates, mainly adenosine triphosphate (ATP). In the case of the yeast *S. cerevisiae* about one mole of ATP is needed to form 16 grams of cell dry matter from glucose^{6,7}. This yield will be further referred to as Y_{ATP} . The ATP demand for growth can be mainly attributed to protein synthesis⁸, therewith creating a close interrelation between the protein content and the Y_{ATP} , as well as the maximal biomass yield on substrate⁹. On the other hand, the amount of anabolic and catabolic protein to catalyze the biochemical reactions determines the achievable metabolic rates. Thus, there is also a direct interdependency between proteome and the fluxome, e.g. the ATP production rate or growth rate of an organism.

Overflow metabolism as a prominent example for a metabolic trade-off

An interesting metabolic phenomenon with respect to the ATP yield in catabolism is the occurrence of overflow metabolism, which describes a behavior where cells show respiro-fermentative activity in the presence of the electron acceptor oxygen.

Most prominent examples are the production of acetate in the prokaryote *E. coli*, lactate production in mammalian cancer cells termed Warburg effect¹⁰ and aerobic ethanol formation in the eukaryote *S. cerevisiae* known as Crabtree effect¹¹. The difference in ATP yield between the respiratory and the fermentative pathway is striking, i.e. in *S. cerevisiae* the complete aerobic catabolism of 1 mol glucose yields about 16 moles of ATP, whereas the fermentation to ethanol only yield 2 moles of ATP. The same yield of 2 ATP/glucose can be obtained with homolactic fermentation, under the premise that the formed lactate and proton are exported without expense of ATP. In order to acquire a higher energy production rate needed for faster growth, organisms must maximize the product of substrate uptake rate and the catabolic ATP stoichiometry¹².

Under the prerequisite that the availability of substrate is usually finite in natural environments¹³ the occurrence of 'wasteful' overflow metabolism seems counterintuitive at first glance. One would expect that evolution favors efficiency under such conditions, as it would lead to a larger ATP production rate and therewith a higher growth rate at the same substrate uptake rate.

However, this line of thought assumes that the rate and the ATP yield of metabolic pathways are independent of each other. In metabolic pathways, there seems to be a metabolic trade-off between the maximal metabolic rate and the yield of ATP production (further referred to as RYT hypothesis)¹⁴, leading to a pareto-optimal solution maximizing the growth rate¹⁵. In 2001 the publication of Pfeiffer et al.¹⁶

received wide attention, applying the RYT concept to explain overflow metabolism. The authors hypothesize that such trade-off could lead to a growth rate advantage of phenotypes exhibiting a Crabtree effect^{17,18}. This hypothesis can explain the usage of pathways with a lower ATP yield, as it would lead to a competitive advantage and has also been discussed extensively as an example for the tragedy of the commons¹⁹ leading to an energetically suboptimal use of substrate^{3,13,20,21}.

Alternative hypothesis explaining the emergence of overflow metabolism such as the Crabtree effect

The RYT is not the only hypothesis trying to explain the occurrence of the Crabtree effect in *S. cerevisiae*, another popular hypothesis is the Make-Accumulate-Consume hypothesis (MAC)^{22,23}, stating that ethanol production confers an evolutionary advantage over its competitors because ethanol is toxic to most other microbes, where at the same time the increased glucose uptake rate starves-out competitors. The ethanol can then later be converted back to acetaldehyde and consumed. There is also no consensus about the ecological niche where the Crabtree effect evolved, which could require anaerobic conditions²⁴, semi-anaerobic conditions²⁵ to no niche at all²⁶.

Modelling the link between metabolic rates, proteome and cellular physiology

With the wider availability of proteome data, the RYT hypothesis could also be discussed in the context of proteome efficiency (also often referred to as catalytic efficiency) of pathways²⁷ (i.e. the amount of enzyme needed to catalyze a certain flux). This allows determining the cellular proteome allocation, which can be compared to optimal expression levels of enzymes inferred from a metabolic model subject to a metabolic target²⁸. This also led to efforts to incorporate constraints reflecting the cellular morphology. Most prominent for this approach are the works in prokaryotes of Molenaar and Zhuang^{2,29} introducing a constraint for the availability of membrane surface area and Basan et al.³⁰ introducing a constraint for the total protein pool of a cell. Although those models are comparably abstract with respect to cellular metabolism and morphology, they already demonstrate a capability for the mathematical description of overflow metabolism.

More rigorous and mechanistic models have been developed based on classical structured or stoichiometric models together with the optimality model flux balance analysis (FBA)³¹ and used to predict evolutionary outcomes^{32,33} or overflow metabolism in *E. coli*³⁴. A key part of FBA is the ability to add constraints on the

fluxes; such constraints are useful in FBA, as they lead to a reduction in the solution space and also serve as an upper boundary for the optimized set of fluxes without setting explicit bound constraints³⁵. Numerous constraints and optimization targets have been proposed for FBA and shall not be reviewed in detail here, as this contribution focusses exclusively on mass constraints for cellular proteins³⁶ (thus further referred to as mFBA).

In mFBA, the cost function associates an amount of protein to each flux in the metabolic network, i.e. the flux divided by the respective specific activity of the protein catalyzing the reaction. The fluxes are then estimated subject to a constrained mass of protein. Therefore, this approach introduces an explicit link between the proteome and the fluxome for steady-state conditions. This approach has recently been applied to the eukaryote *S. cerevisiae*³⁷, using a single protein mass constraint on intermediary metabolism and has been shown to predict the Crabtree effect.

Aim of the work

Current models are limited to specific conditions e.g. carbon limitation and have a coarse description of yeast morphology and physiology. Here an extension of the mFBA approach is presented that will allow to predict the outcomes of evolutionary experiments under various conditions by introducing limiting nutrient kinetics that allow predicting the cellular rates and protein composition under different environmental conditions.

The model framework is extended to multiple protein constraints, reflecting the morphology of eukaryotic cells more closely, i.e. multiple protein constraints posed by limited availability of cellular and inner mitochondrial membrane are implemented (Figure 4.1). Additionally, special attention is set in order to model the physiology of yeast quantitatively, in particular with respect to the metabolic stoichiometry and the interrelation between the total protein content and cellular energetics (Y_{ATP}).

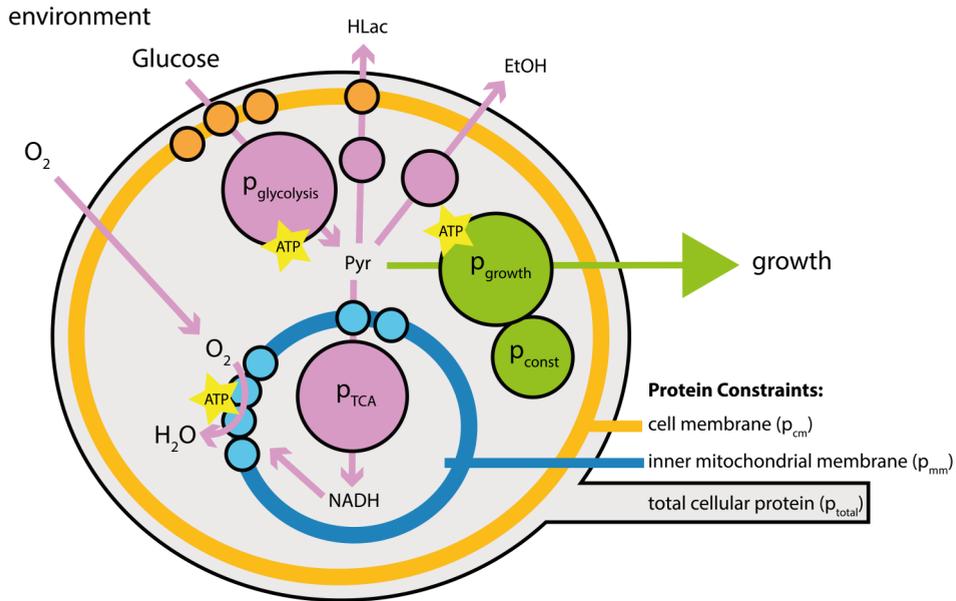


Figure 4.1: Schematic representation of the model containing lumped metabolic fluxes (arrows) and the associated protein p (circles), where purple indicates catabolic and green anabolic reactions. Moreover showing the three incorporated protein mass constraints for cell membrane p_{cm} , inner mitochondrial membrane p_{mm} and total cellular protein p_{total} .

With the model, different optimal metabolic strategies can be predicted with respect to the achieved growth rate and -yield, where the results are compared to the outcomes of published experimental data and interpreted with respect to the resulting evolutionary fitness under different conditions. The potential of the approach is demonstrated by addressing a number of long standing questions about yeast physiology and evolution in a comprehensive way. These questions are:

- (1) Can the concept of proteome efficiency explain why under batch conditions there is such a variance in the ratio between fermented and respired glucose among different Crabtree positive yeast species and strains?³⁸
- (2) How could evolution divide yeast into Crabtree positive and Crabtree negative species and in which environmental conditions are required for this divergence?
- (3) Why can the expression of a glucose proton glucose symport in the cell membrane be associated with the Crabtree negative phenotype, whereas the expression of a glucose uniporter can be associated with a Crabtree positive phenotype and how does the dramatic loss in glucose affinity affect the evolutionary fitness?³⁹

- (4) If the Crabtree effect causes a growth advantage, why are there Crabtree negative yeasts, e.g. *C. utilis* exhibiting a higher maximal growth rate compared to *S. cerevisiae*? What role does the overall protein content play?
- (5) Can the evolutionary fitness of *S. cerevisiae* be understood better in terms of the RYT hypothesis⁴⁰ or the “make-accumulate-consume” hypothesis (MAC)²²?
- (6) How does the evolutionary fitness of the Warburg effect compare to the Crabtree effect and under which conditions could directed evolution lead to a homolactic fermenting *S. cerevisiae*?

Methods

In this contribution a combination of stoichiometric modelling, kinetic modelling of the limiting nutrients and mFBA with multiple protein pool constraints is used. The model allows simulating of different environmental conditions and is used to predict optimal metabolic strategies in batch and substrate-limited culture for the eukaryote *S. cerevisiae*. This work builds upon a recently published model³⁷ and introduces several changes and extensions.

Model Reactions

In order to keep the model comprehensive the linear pathways were lumped and the respective specific activities of the single enzymes were summed up in the pathway as described in the Supporting material (see Table 4.S1).

The model is structured, meaning it possesses different compartments. Whereas all proteins contribute to the total amount of protein, membrane proteins are also assigned either to the cell membrane or the inner mitochondrial membrane and will therefore also contribute to the respective protein masses.

The stoichiometry of respiration corresponds to a P/O ratio for NADH and FADH of 1.2⁴¹ or 16.4 moles of ATP per glucose (including 2 ATP from glycolysis). The conserved moieties ADP and NAD⁺, Q and CoA have been excluded from the stoichiometry matrix as they are redundant in the stoichiometry matrix. GTP is balanced as ATP, as both are assumed interconvertible.

Table 4.1. Lumped reactions with designated location according to the 3 protein mass constraints in the cell membrane (p_{cm}), inner mitochondrial membrane (p_{mm}) and the overall cellular protein (p_{total})

Name	Reaction	p_{cm}	p_{mm}	p_{total}
HXT (Glucose import)	$Glc_{ec} \rightarrow Glc$ (uniporter) or $Glc_{ec} + ATP \rightarrow Glc$ (glucose-proton symport)	X		X
JEN1 (Lactic acid export)	$HLac \rightarrow HLac_{ec}$	X		X
JEN1 (Lactic acid import)	$HLac_{ec} \rightarrow HLac$	X		X
Glycolysis	$Glc \rightarrow 2Pyr + 2ATP + 2NADH$			X
PDC + ADH	$Pyr + NADH \rightarrow EtOH + CO_2$			X
LDH	$Pyr + NADH \rightarrow HLac$			X
MPC _{fwd}	$Pyr \rightarrow Pyr_{mit}$		X	X
MPC _{bwd}	$Pyr_{mit} \rightarrow Pyr$		X	X
TCA cycle	$Pyr_{mit} \rightarrow 3CO_2 + 4NADH + FADH_2 + ATP$			X
NDI, NDE	$NADH \rightarrow QH_2$		X	X
SDH	$FADH_2 \rightarrow QH_2$		X	X
2RIP1 + COX	$O_2 + 2QH_2 + 9H^+_{mit} \rightarrow 9H^+$		X	X
ATP	$3.75H^+ \rightarrow ATP + 3.75H^+_{mit}$		X	X
DW _{raw}	$0.413Pyr + 0.005NH_3 + 0.65ATP + 0.15H_2O \rightarrow$ $1C_1H_{1.94}O_{0.91}N_{0.005} + 0.014NADH + 0.24CO_2$			
Protein	$0.413Pyr + 0.275NH_3 + 4.59ATP \rightarrow$ $1C_1H_{1.581}O_{0.318}N_{0.275} + 0.0065NADH + 0.24CO_2 +$ $0.442H_2O$			X
Glycerol production	$0.5Glc + NADH + ATP \rightarrow Glycerol$			X
CYB2 + 0.5 COX	$HLac + 0.5 O_2 + 3H^+_{mit} \rightarrow Pyr + 3H^+$		X	X
ADH+ALD	$EtOH + 2ATP \rightarrow AcCoA + 2NADH$			X
Glyoxyate + SFC + PEPCK +PK	$2AcCoA \rightarrow Pyr + CO_2 + 2NADH + 1FADH_2$		X (only SFC)	X
TCA cycle AcCoA + CAT2 + CRC1	$AcCoA \rightarrow 2 CO_2 + 3NADH + FADH + ATP$		X (only CRC1)	X

The biomass equation, protein allocation and cellular maintenance for growth

Biomass (DW) is assumed to be composed of two parts: (1) DW_{raw} containing the cell constituents carbohydrates, lipids, DNA and RNA⁴² and (2) a protein fraction p_{total} containing all cellular protein. Due to a lack of complete data, the composition of DW_{raw} is considered to independent of the growth rate. This simplification can be

justified, because the named constituents make up a smaller amount of the biomass compared to proteins⁴³ and more importantly have a very small influence on the bioenergetics (Y_{ATP}) of the cell compared to protein synthesis^{44,45}. Additionally, this has the advantage that all protein fractions p can be expressed over DW_{raw} , leading to the growth reaction with the following overall biomass constitution.

$$DW = DW_{raw}(1 + p_{total}) \quad (4.1)$$

The reconciled elemental biomass composition (DW), as well as the protein composition (p_{total}) have been taken from literature⁴⁶. To obtain DW_{raw} , the protein content (47% (w/w) at $D=0.1$ 1/h⁴⁷) was subtracted from the biomass composition. As the biomass and the protein reactions have 5 unknowns but only 4 moieties (C,H,O,N), the molar biomass yield on substrate was taken from literature⁶ to balance the equations, where the substrate needed for the formation of NADPH via the Pentose phosphate pathway has been accounted for and is therewith not explicitly considered in the model. To find a compromise for the predictions with fermentative and non-fermentative carbon sources, intracellular pyruvate was chosen as the precursor for the biomass and the protein synthesis reaction.

In contrast to the original model³⁷, the overall protein content (p_{total}) is implemented directly into the model. Therefore p_{total} is divided into two fractions, a constitutive amount of protein p_{const} which is independent of the fluxes and a flux-associated protein fraction p_{flux} ³⁰ which is a function of the necessary protein needed to catalyze the metabolic fluxes in the model (see also eq.(4.3)).

$$P_{total} = P_{flux} + P_{const} \quad (4.2)$$

The amount of growth independent protein content p_{const} has been parametrized to 36.5% (w/w) (corresponding to 0.56 gProtein/g DW_{raw}) at zero growth from experimental observations in glucose limited chemostat cultures using linear extrapolation^{9,48}.

The ATP cost for DW_{raw} has been taken from calculations based on the ATP requirements for the synthesis of the respective cell constituents⁹, whereas the ATP cost for protein synthesis has been estimated using experimental Y_{ATP} data⁷, so that the overall growth model describes the interrelation between observed Y_{ATP} and protein content as observed experimentally in *S. cerevisiae* (see Figure 4.S1 in the Supporting material).

As in all models linking cellular energetics and growth there is a systematic discrepancy between the apparent and the theoretical $Y_{\text{ATP, max}}$ ^{47,49-51}. This is usually accounted for with a maintenance coefficient linked to the growth rate and a growth independent maintenance coefficient⁵². As the model predicts rates as a function of protein masses, this ATP gap is included into the protein reaction i.e. the ATP cost is a linear function of the protein content (also see Table 4.1). This explains the about 2.5 times higher ATP stoichiometry for cellular protein synthesis compared to the usually considered ATP demand^{6,8,50}, which may also be too low^{53,54}.

Specific protein activities

The specific protein activities have been used as published before³⁷. Additionally, the *in vivo* specific enzyme activity for the Hexose transporter has been taken from literature⁵⁵. Although being aware that there are many more glucose transporters^{56,57} in *S. cerevisiae* than assumed in our model abstraction, due to a lack of data the high affinity glucose uniporter HXT2 was used as a proxy. The same situation holds for the different cytochrome c oxidases⁵⁸. It is assumed that the specific activity of the glucose-proton symport is equal to the glucose uniporter as also found experimentally⁵⁹. This assumption was also tested in the model, where the model turned out to be robust with respect to this value in the case of a glucose-proton symport. By altering the value by plus and minus 50%, only a marginal effect the model prediction could be observed (data not shown).

Fortunately, the lactate-proton symport JEN1 was characterized using the same method as the Hexose transporter⁶⁰. The specific activity of JEN1 was also used for the mitochondrial pyruvate channel (MPC). No specific *in vitro* protein activities could be found for the mitochondrial AcCoA transport, which takes place via multiple pathways in *S. cerevisiae*⁶¹. All Acetyl-CoA transport was assumed to be carnitine-dependent and the specific activity of the carnitine transporter CRC1 was parametrized in the order of magnitude of the JEN1 transporter. The same holds for the mitochondrial succinate-fumarate carrier (SFC)⁶² which is needed for growth on C-2 compounds to link the glyoxylate pathway to gluconeogenesis (succinate dehydrogenase is only expressed in the mitochondrial matrix but not in the cytosol). The pathways involving those reactions are only used in growth on non-fermentable carbon sources, so the parametrization does not have influence on predictions for growth on glucose and a very low sensitivity has been found for the growth predictions for the non-fermentable substrates (results not shown). In contrast to lactate, ethanol can pass the membrane without a transporter by diffusion. This means the ethanol flux is not bounded by the protein constraint for the cell

membrane in the model and could theoretically exceed physiological limits. All predicted rates are well below the estimated maximal ethanol production rate of about 31 mmol/gDW/l⁷ in anaerobic conditions, thus it was assumed that diffusion resistance is not the rate-limiting step under the imposed conditions. Due to the very high membrane diffusivity, diffusion limitation is also excluded for the oxygen uptake. The specific protein activity for the growth reaction to catalyze the synthesis of biomass e.g. ribosomes (p_{growth} in Figure 4.1) was calculated based on the linear increase in experimentally observed protein content as found in aerobic^{9,48} and anaerobic⁶³ glucose limited chemostat cultivations (Figure 4.3, right panel). A detailed table with the used specific protein activities can be found in the Supporting Materials (refer to Table 4.S1 in the Supporting material).

Flux prediction

Flux balance analysis (FBA) is used to optimize a linear combination of fluxes. Here the maximization of the growth rate is the objective function subject to a number of constraints. Steady-state conditions are enforced in the network and moreover all fluxes were constrained to be strictly positive e.g. a reversible reaction was introduced as 2 reactions with positive sign to circumvent the calculation of negative protein masses.

$$\begin{aligned}
 \mu_{\max} &= \arg \max(\mu) \\
 \text{such that :} \\
 N \cdot v &= 0 \\
 v_i &> 0 \\
 \sum \frac{v_i}{\sigma_i \cdot a_{i,cm}} &< p_{cm} \\
 \sum \frac{v_i}{\sigma_i \cdot a_{i,mm}} &< p_{mm} \\
 \sum \underbrace{\frac{v_i}{\sigma_i \cdot a_i}}_{P_{flux}} + p_{const} &< p_{total}
 \end{aligned} \tag{4.3}$$

The difference between FBA and mFBA are the inequality constraints in eq. (4.3), where the flux vector is divided by the specific protein activity a (with the units

catalyzed flux/mass of enzyme) and a dimensionless saturation factor σ_i (can have values between 0 and 1) and are set by default to 0.5 as in the original model³⁷.

$$\sigma_i = 0.5 \quad (4.4)$$

In contrast to the original model³⁷ three protein mass constraints are introduced, where p_{cm} constrains the protein mass in the cell membrane, p_{mm} the mass of protein in the mitochondrial membrane²⁹ and the overall protein of the cell p_{total} (eq.(4.2), see also Figure 4.1). A constraint for the mitochondrial lumen was not introduced as the mitochondrial morphology and therewith the surface to volume ratio of mitochondria was shown to be highly variable in *S. cerevisiae*⁶⁴.

Implementation of the interdependence between the predicted overall protein content and Y_{ATP}

However this is not sufficient to reflect the physiology of yeast with respect to the multiple protein constraints. As p_{total} is an inequality constraint, this means that it is not guaranteed that the model will converge towards the total protein content constraint. The model has to account for this behavior i.e. describe that the overall biomass composition DW (and therewith the Y_{ATP}) in the stoichiometry matrix N is dependent on p_{total} (eq.(4.1)), which is in turn is a result of the simulation itself (eq. (4.3)). This leads to an implicit problem, which is solved by an additional iterative optimization minimizing the difference between the estimated and the actual protein content in the model.

$$\left[\begin{array}{l} p_{total} - \hat{p}_{total} = 0 \\ \text{such that} \\ DW = DW_{raw}(1 + \hat{p}_{total}) \quad \text{eq. (1)} \\ \left[(p_{total}, \dots) = \min(-\mu) \right] \quad \text{eq. (3-5)} \\ \text{s. t. ...} \end{array} \right] \quad (4.5)$$

The inner optimization problem was solved with the MATLAB *linprog* solver using the dual simplex algorithm therein and the MATLAB *fzero* algorithm was used to solve the outer loop.

Simulation of substrate-limited conditions

For the simulation of oxygen and glucose limited conditions the saturation constant σ_i of the hexose transporter or the cytochrome c oxidase (COX) was set to values in the interval $[0, 1[$. The respective broth (glucose or oxygen) concentrations are derived from the Michaelis-Menten equation (see eq.(4.4)).

$$c = \frac{\sigma \cdot K_m}{1 - \sigma} \quad (4.6)$$

The apparent affinity constants K_m have been taken from literature as 1 mmol glucose/l for the glucose uniport and 0.025 mmol glucose/l for the glucose-proton symport mechanism⁶⁵ and 0.75×10^{-3} mmol oxygen/l for the cytochrome c oxidase^{66,67} corresponding to 0.32 % air saturation at standard pressure and a temperature of 30 °C (Oxygen conversion table, PreSens, Germany).

Results and Discussion

Base case model parametrization for *S. cerevisiae* and *C. utilis*

There is surprisingly little data about the quantitative physiology for different strains of *S. cerevisiae* with the additional problem that most of all available data has been obtained with 'domesticated' laboratory or industrial strains, which may have altered properties due to selection for certain properties^{68,69}. For example the strain DS28911, where the model of Nilsson and Nielsen³⁷ is based on is an aneuploid industrial strain used for baking. Such strains are selected for a high rate of CO₂ production under anoxic conditions and thus high alcoholic fermentation rate⁴⁸.

Similar the popular laboratory strain CEN.PK 113-7D has been optimized for laboratory (chemostat) fermentations and genetic accessibility⁷⁰. It is known that this strain has numerous mutations in the cAMP signaling pathway⁷¹ that affect the regulation of glucose repression⁷². It cannot be excluded that the latter or other mutations alter the respiro-fermentative characteristics^{73,74}, in particular as CEN.PK 113-7D shows a very similar behavior compared to DS28911⁷⁵.

The strain that has been thoroughly characterized with respect to quantitative physiology and may resemble wild-type characteristics better is *S. cerevisiae* CBS 8066, which is a hybrid of the strain Y55 originally isolated from wine grapes in France during the 1930s⁷⁶ and a not further described 'wild-type' (CBS strain database). Therefore the model will be based in following primarily on data obtained with CBS 8066. For Crabtree-negative yeast even less quantitative physiological

data is available, here the well characterized *C. utilis* CBS 621 is used as a proxy to parametrize the model.

With the known stoichiometry and the specific enzyme activities, the model has only 3 free variables to parametrize, the mass constraints for the (1) protein mass in the cell membrane p_{cm} , (2) the inner mitochondrial membrane p_{mm} and (3) the total protein mass p_{total} . As the biomass composition changes with the predicted overall protein content, the constraints are expressed relative to the not changing raw biomass amount in $gProtein/gDW_{raw}$.

Table 4.2. Parameter definition and model parametrization for *S. cerevisiae* and *C. utilis*

Term	Value	Definition
p_{cm}	0.012 $gProtein/gDW_{raw}$	Protein constraint for Glucose and lactic acid transporters in the cell membrane, parametrized to reflect a maximal glucose uptake rate of 16 mmol/gDW/h.
p_{total}	<i>S. cerevisiae</i> (all): 1.37 $gProtein/gDW_{raw}$ <i>C. utilis</i> CBS 621: 2.00 $gProtein/gDW_{raw}$	Constraint for the overall cellular protein, parametrized to reflect the maximal growth rate
p_{mm}	<i>S. cerevisiae</i> CBS 8066: $p_{mm} = p_{mm,min}$ <i>S. cerevisiae</i> DS 28911: 0.16 $gProtein/gDW_{raw}$ <i>C. utilis</i> CBS 621: unbounded	Protein constraint for the inner mitochondrial membrane
$p_{mm,min}$	<i>S. cerevisiae</i> (all): 0.23 $gProtein/gDW_{raw}$ <i>C. utilis</i> : n.a.	Protein constraint for the inner mitochondrial membrane, subscript min refers to the minimal protein mass allowing to reach the maximal growth rate
$p_{mm,max}$	0.45 $gProtein/gDW_{raw}$	Protein constraint for the inner mitochondrial membrane, subscript <i>max</i> refers to the protein mass allowing to reach the maximal oxygen uptake rate of 20 mmol/gDW/h in <i>S. cerevisiae</i> ⁸¹

First p_{cm} was parametrized to a maximal glucose uptake rate of 16 mmol/gDW/h⁷⁷, while leaving the other two constraints unbounded. As a second step the value of p_{total} was estimated according to the observed maximal aerobic growth rate for the *S. cerevisiae* strain CBS8066 on glucose of 0.49 1/h in minimal medium with NH_4^+ as a nitrogen source⁷⁸, where p_{mm} was left unbounded. For *C. utilis* p_{total} is parametrized in the same way to a maximal growth rate of 0.59 1/h⁷⁹. This parametrization also leads to a prediction of p_{mm} as the model minimizes the protein cost for all constraints simultaneously. This is the minimal amount required to reach the maximal growth rate on glucose and is therefore further referred to as $p_{mm,min}$. This means a further increase in mitochondrial membrane protein beyond $p_{mm,min}$ cannot increase the maximal growth rate anymore. Also $p_{mm,max}$ was estimated, which is the amount of mitochondrial enzyme allowing for the maximal oxygen uptake rate as observed experimentally under uncoupling conditions with benzoic acid⁸⁰.

Mitochondrial protein constraint p_{mm} with different carbon sources under batch conditions

In the initial parametrization $p_{mm,max}$ is about two times the value of $p_{mm,min}$ showing a significant variability. This raises the question how the mitochondrial constraint p_{mm} affects the growth on the non-fermentable carbon sources ethanol and lactic acid which consume more oxygen per produced biomass and protein. This leads to a different situation, i.e. here an in comparison higher mitochondrial protein mass could further increase the maximal growth rate (see Figure 4.2, green line). The results indicate that the mitochondrial protein content is only partly derepressed during growth on ethanol with a p_{mm} about 25% larger compared to $p_{mm,min}$ and significantly lower compared to uncoupling conditions ($p_{mm,max}$) which correspond to the maximal oxygen uptake rate of 20 mmol/gDW/h. The ethanol uptake rate at the maximal growth rate is predicted to 6.5 mmol/gDW/h, if compared to the ethanol production rate under anaerobic conditions this makes a diffusion limitation of ethanol in the cell membrane unlikely. It remains unclear if the necessary additional mitochondrial surface causes the observed change in mitochondrial morphology between growth on glucose and ethanol or under uncoupling conditions^{64,81}. In the case of growth on lactic acid one would expect a behavior similar to ethanol; however the observed growth rate corresponds to a p_{mm} below the $p_{mm,min}$ on glucose. Therefore, it is unlikely that the growth rate on lactate is limited by the mitochondrial membrane capacity or it could be that the experimental growth rate on lactate is biased due to uncoupling⁸². The predicted p_{mm} for the maximal growth

rate for a fully respiratory strategy, which was achieved by reducing the glycolytic flux using chimeric glucose transporters⁸³ is predicted close to $p_{mm,min}$.

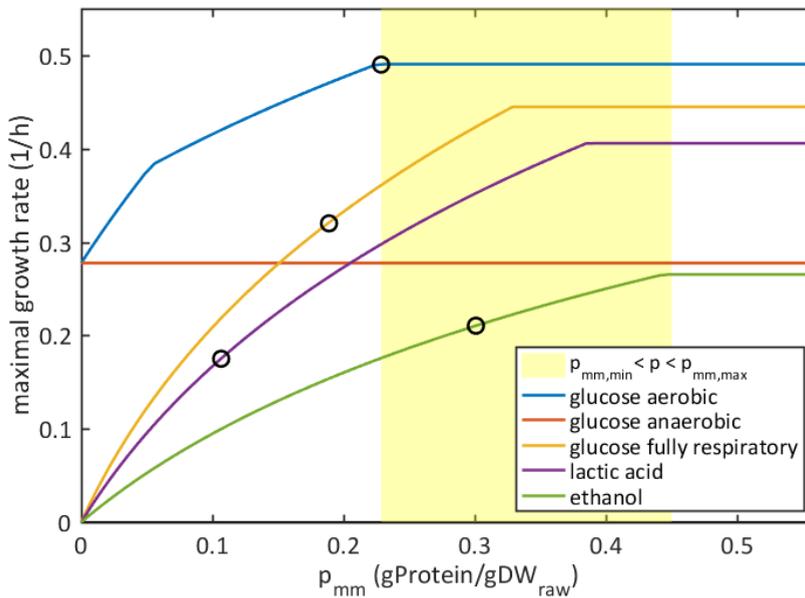


Figure 4.2. Model prediction for the maximal growth rate of *S. cerevisiae* as a function of the mitochondrial protein constraint $p_{mm,min}$ on different substrates. The predictions are compared to the empirically observed maximal growth rates in *S. cerevisiae* (black circles). Glucose aerobic 0.49 1/h⁷⁸, Glucose anaerobic 0.31 1/h⁷, glucose fully respiratory (34% reduction from wild-type⁸³) 0.32 1/h, lactic acid 0.18 1/h⁸⁴ and growth on ethanol 0.21 1/h⁷⁷.

Simulations for *S. cerevisiae* under glucose limited conditions

In order to investigate the effect of p_{mm} on the cellular physiology further, the estimated parameters were used to predict the growth rate of *S. cerevisiae* under glucose limited conditions, where p_{mm} was either set to $p_{mm,min}$ or left unconstrained.

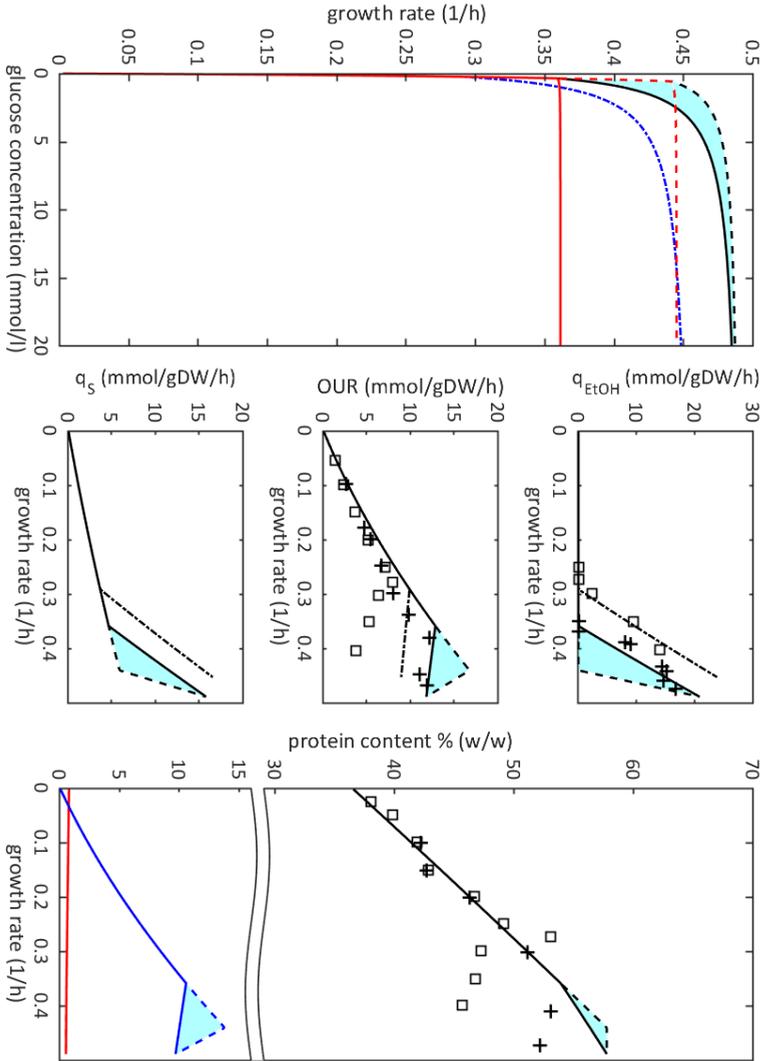


Figure 4.3. Left: growth rate as a function of broth glucose concentration for a Crabtree positive phenotype (black lines) with $\rho_{mm} = \rho_{mm,min}$ (-) and unbounded ρ_{mm} (--). The cyan area in between shows the solution space leading to the maximal growth rate of 0.49 1/h. Red lines show a purely respiratory phenotype and the blue line a purely fermenting phenotype. Middle: Biomass specific rates for a phenotype with $\rho_{mm} = \rho_{mm,min}$ (-), for unbounded ρ_{mm} (--), and for $\rho_{mm} < \rho_{mm,min}$ (.-), the cyan area shows the solution space leading to the maximal growth rate of 0.49 1/h, together with experimental observables in aerobic glucose limited chemostat cultures for the *S. cerevisiae* strains CBS 8066 (+) and DS 28911 (\square). Right: Protein content as a function of growth for a phenotype with $\rho_{mm} = \rho_{mm,min}$ (solid line) and for unbounded ρ_{mm} (dashed line) for ρ_{total} (black), ρ_{mm} (blue) and ρ_{cm} (red) together with experimental observables for ρ_{total} in aerobic glucose limited chemostat cultures for the *S. cerevisiae* strains CBS 8066 (+) and DS 28911 (\square).

The model predicts ethanol production also for the unconstrained p_{mm} , however this happens only once p_{total} converges to the boundary, forcing the cell to replace the oxidative phosphorylation by the more protein efficient fermentative pathway. This transition starts at a growth rate of 0.44 1/h (see Figure 4.3), which is also the maximal growth rate predicted for a purely respiratory phenotype. The protein rearrangement towards a Crabtree effect is abrupt in the $q_i-\mu$ plots, also because the protein cost for growth rises about the same as the necessary protein cost for the catabolic pathways as a function of the growth rate (See Figure 4.S2). It can be seen that the higher protein efficiency of the fermentative pathway allows for a 10% increase in maximal growth rate compared to a non-fermenting phenotype under a limitation in the overall protein content.

Otherwise it is under glucose limited conditions always beneficial in terms of growth rate to invest protein into the oxidative pathway. The reason for this behavior is that the glucose influx is limited under the applied glucose limited conditions by the mass of glucose transporters p_{cm} that can be expressed in the cell membrane, therewith favoring the pathway with the higher ATP yield on substrate.

Interestingly, the Crabtree effect can always restore the maximal growth rate when p_{mm} is above $p_{mm,min}$, this means that this is a free variable in cellular decision making, as above this threshold ethanol fermentation with the higher proteome efficiency can compensate a reduction in p_{mm} .

It is found that *S. cerevisiae* CBS 8066 operates at the lower limit of this solution space in p_{mm} , corresponding to an onset of the Crabtree effect at a dilution rate of ~ 0.37 1/h, therewith finding an optimal compromise between maximizing the substrate uptake rate and the ethanol production rate without compromising the maximal growth rate. Still it has to be noted that the Crabtree effect comes at the cost of a clear disadvantage in growth rate at the same residual glucose concentration, despite the higher protein efficiency of the fermentative pathway and thus a fully respiratory strategy would outperform this Crabtree positive strategy at glucose concentrations below ~ 2 mmol/l glucose in terms of growth rate. The situation looks different for the strain DS 28911 which operates outside the predicted solution space with an onset of the Crabtree effect at a dilution rate of ~ 0.28 1/h. According to the model this must lead to a reduced maximal growth rate predicted slightly higher compared to the experimentally observed growth rate (0.45 vs. 0.42 1/h⁷⁵). The reason for this over prediction may be explained by the strong decay in the overall protein content and OUR which the model cannot reproduce with the chosen objective function of maximizing growth. Also CBS 8066 shows this behavior at high growth rates, although much less pronounced, whereas it

has not been observed in anaerobic conditions⁶³. The reason for this behavior cannot be elucidated with the available data, but based on the model simulations the possibility that a drop in protein content is caused by the enzymatic rearrangement from respiration to fermentation can be excluded. The effect could be caused by a changed cellular morphology at high growth rates leading to a reduction of the growth independent protein content p_{const} or lead to a systematic measurement error in the biomass concentration⁸⁵. The model cannot describe such behavior on a mechanistic level; however it is clear that a reduction of the growth independent protein content would increase the Y_{ATP} and therefore is positively correlated with the growth rate. The benefit of this effect is independent of the metabolic strategy and will therefore not influence the general outcomes of this comparative study.

Exploring the link between the glucose uptake mechanism and the Crabtree effect in yeast

So far only yeast phenotype expressing a glucose uniporter as glucose uptake mechanism have been considered, also referred to as facilitated diffusion. It is assumed that the progenitor of the modern yeast showed a respiratory behavior²⁴, therefore in order to understand the relevant evolutionary mechanism the physiology of Crabtree positive and Crabtree negative yeast has to be compared (see Figure 4.4). Interestingly, the occurrence of a glucose proton symport has been strictly associated with a Crabtree negative phenotype³⁹. This has severe impact on the ATP yield as the import of one proton per glucose molecule will cost ATP as the proton has to be exported again by the proton ATPase. This will reduce the ATP yield of the catabolic pathways, whereas the protein cost for the pathway enzymes stays constant. With a proton ATPase stoichiometry of 1proton/ATP⁸⁶ this leads to a 6% lower ATP yield per glucose in the respiratory pathway and a 50% decrease in the fermentative pathway. On the other hand, the import of glucose by facilitated diffusion makes this reaction reversible and leads to a dramatic loss of glucose affinity with an about 40 fold increase in the apparent affinity constant from 0.025mmol/l to 1mmol/l³⁹.

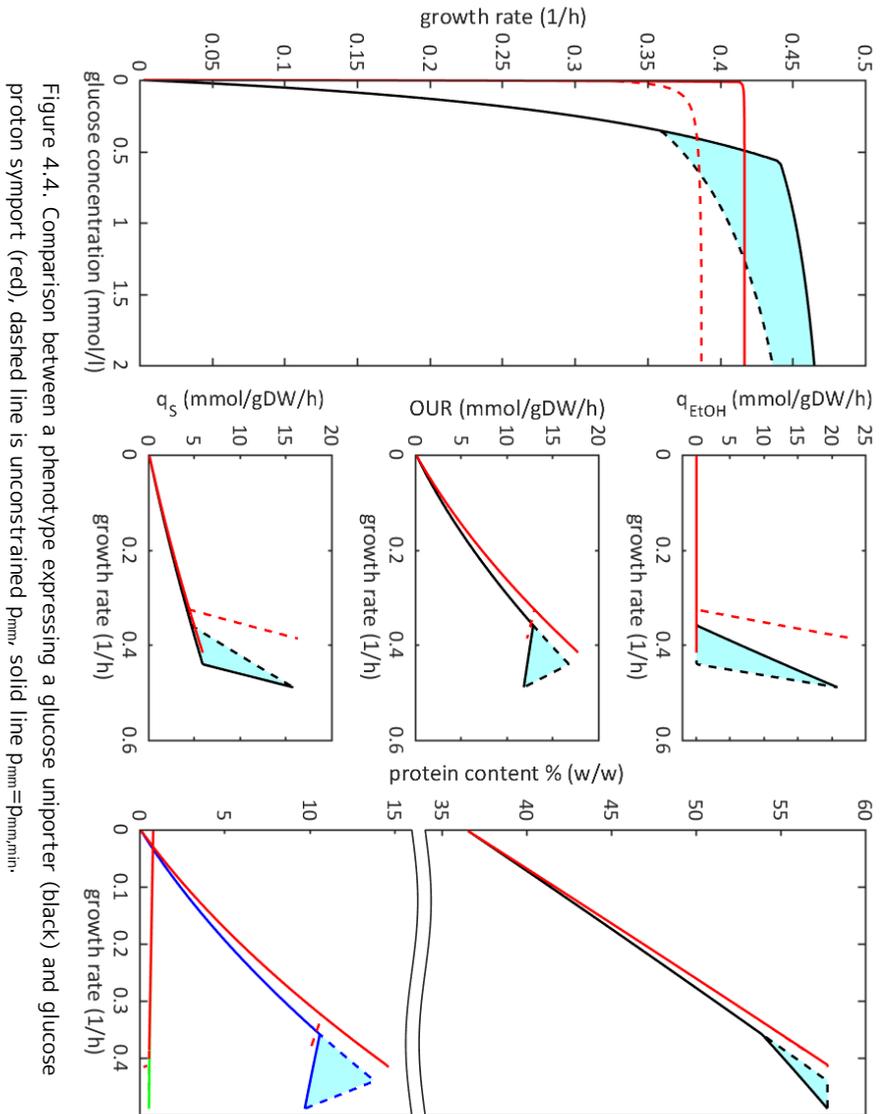


Figure 4.4. Comparison between a phenotype expressing a glucose uniporter (black) and glucose proton symport (red), dashed line is unconstrained p_{mm} , solid line $p_{mm}=p_{mm,init}$.

The model predicts the occurrence of a Crabtree negative phenotype with the only change being the introduction of a glucose proton symporter instead of a glucose uniporter. Reason is that with the 50% decreased ATP yield of the fermentative pathway the protein efficiency of the fermentative pathway is no longer higher than the respiratory pathway and therewith the occurrence of ethanol fermentation leads to a reduction in the maximal growth rate. With the applied protein requirements for the critical ATP stoichiometry for the import of glucose leading to the occurrence of a Crabtree effect can be calculated from the model to 0.93 mol ATP/mol glucose

imported (see Figure 4.S3) and is therewith only slightly below the ATP stoichiometry of 1mol ATP/mol glucose when expressing a glucose proton symport. Even below this threshold $p_{mm,min}$ is positively correlated with the stoichiometry, thus reducing the solution space for the Crabtree effect considerably (see also Figure 4.S3).

Due to the lower overall ATP yield phenotypes showing ATP consuming glucose transport always exhibits a lower maximal growth rate compared to the phenotype expressing the glucose uniporter at the same overall protein content. On the other hand, the advantage of the strategy is, due to the increased glucose affinity, an increased glucose uptake rate and therewith growth rate at low glucose concentrations despite lower ATP yield of the catabolic pathway.

The link between total protein content and Crabtree phenotype

In practice Crabtree negative yeasts often exhibit higher maximal growth rates compared to *S. cerevisiae*, these can only be obtained in the model by an increased total protein pool.

To investigate this behavior simulations were performed varying the overall protein constraint for phenotypes with glucose uniporter and bottlenecked mitochondria and compare it to the Crabtree positive phenotype with glucose proton symport and constrained p_{mm} .

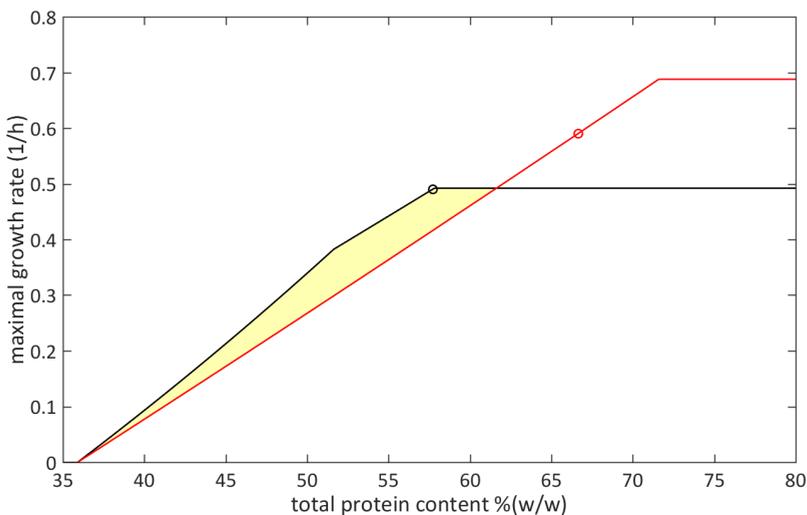


Figure 4.5. Predictions for the maximal growth rate as a function of overall protein content for the Crabtree positive (black) (glucose uniporter and $p_{mm}=p_{mm,min}$) and Crabtree negative (red) (glucose proton symport) strategy. Markers indicate observed maximal growth rates for *S. cerevisiae* CBS 8066 (0.49 1/h⁷⁸) and *C. utilis* CBS 621 (0.59 1/h⁷⁹). The yellow area indicates the protein content where the Crabtree positive strategy is favorable under glucose excess.

The results show that the constraint in the mitochondrial protein content p_{mm} of the Crabtree positive phenotype limits the possible maximal growth rate (see Figure 4.5), which can also not be further increased with increased total protein content. Exactly opposite is the situation for the phenotype with glucose proton symport. Here an increase in overall protein content is directly related to an increase in growth rate, also beyond the maximal growth rate of *S. cerevisiae*. This is in agreement with experimental observations which show that Crabtree negative yeasts show higher protein content compared to *S. cerevisiae* e.g. for *C. utilis* CBS 621 an about 5% higher protein content has been determined (62.9%(w/w)) and for *Kluyveromyces marxianus ATCC 26548* a protein content of up to 71.9% has been observed^{6,87}. In general this means that a Crabtree positive strategy i.e. constraining p_{mm} leads to a higher growth rate at equivalent protein content however at the same time this strategy also imposes a limit to the achievable growth rate achievable with an increase in protein content; in contrast to Crabtree negative phenotypes.

Prediction of competition between *S. cerevisiae* and *C. utilis* under glucose and oxygen limitation

In order to investigate this further, the competition between *S. cerevisiae* CBS 8066 and the Crabtree negative yeast *C. utilis* CBS 621 with a maximal growth rate of 0.59 1/h was predicted under substrate limitation. Although a higher P/O ratio has been hypothesized for *C. utilis*⁶ due to a proton translocating NADH dehydrogenase, newer research suggests that this is only the case for the external NADH dehydrogenase⁸⁸, whereas the internal NADH dehydrogenase does not export protons (as in *S. cerevisiae*⁸⁹). The loss of the respiratory complex I took place earlier in the evolution of yeast²⁴ and seems therefore not to be directly linked to the development of the Crabtree effect.

The simulations show that *C. utilis* CBS 261 has a globally higher growth rate at the same glucose concentration and also at the same oxygen partial pressure (see Figure 4.6). The only environment where *S. cerevisiae* can prevail is under severely oxygen limited conditions below a dissolved oxygen concentration of ~0.02 % air saturation. These predictions are in agreement with experimental observations from competition experiments in chemostat culture⁷⁹.

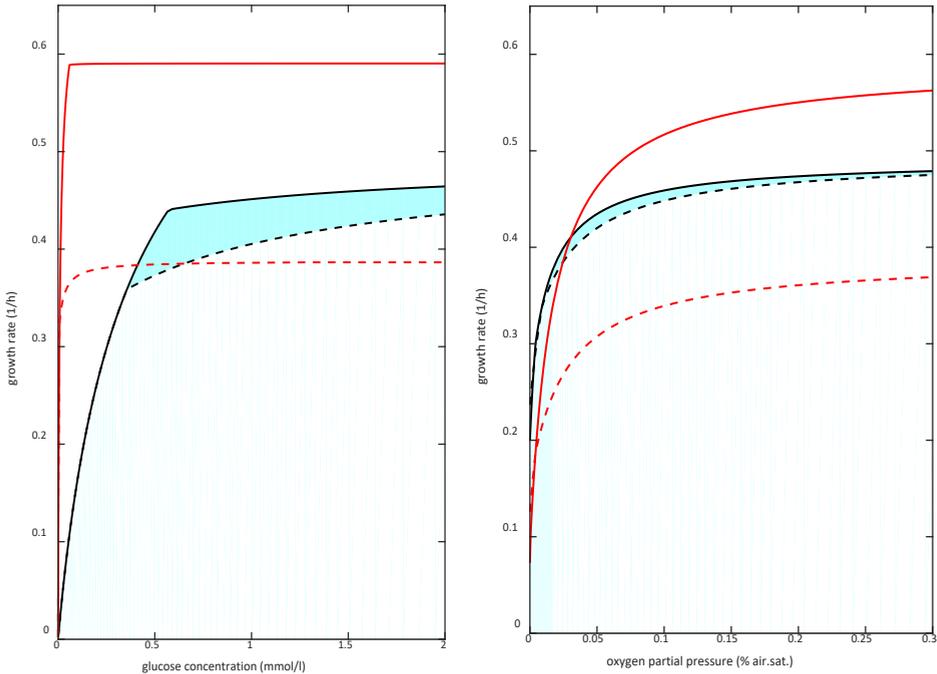


Figure 4.6. Predicted growth rates of *C. utilis* CBS 261 (Crabtree negative, red) and *S. cerevisiae* CBS8066 (Crabtree positive, black) yeast under glucose (left) and oxygen limitation (right). Dashed lines is unconstrained p_{mm} , solid lines $p_{mm}=p_{mm,min}$.

This is interesting from an evolutionary point of view as *S. cerevisiae* cannot compete against the Crabtree negative phenotype in terms of glucose affinity nor in maximal growth rate nor in growth on non-fermentable carbon sources (data not shown) nor in utmost aerobic conditions.

Comparison of the Crabtree and the Warburg effect

Frequently an analogy between the Crabtree and the Warburg effect is made, as the lactate fermentation leads to the same ATP yield on glucose compared to ethanol fermentation. This raises the question whether the properties observed for the Crabtree effect also hold for the Warburg effect.

While ATP yields are equal, there is one significant difference between the two pathways, which is that lactate is dissociated at near neutral cytosolic pH and can only leave the cell by a transporter whereas ethanol can diffuse through the cell membrane without a transporter.

Under the constraint of cell membrane protein (p_{cm}) this leads to a metabolic trade-off between expression of the glucose uniporter (HXT) and the lactate-proton symporter (JEN1) in the cell membrane. This means that the lactate export rate can only be increased on cost of a decrease in the specific glucose uptake rate, therewith creating to a trade-off between the two rates.

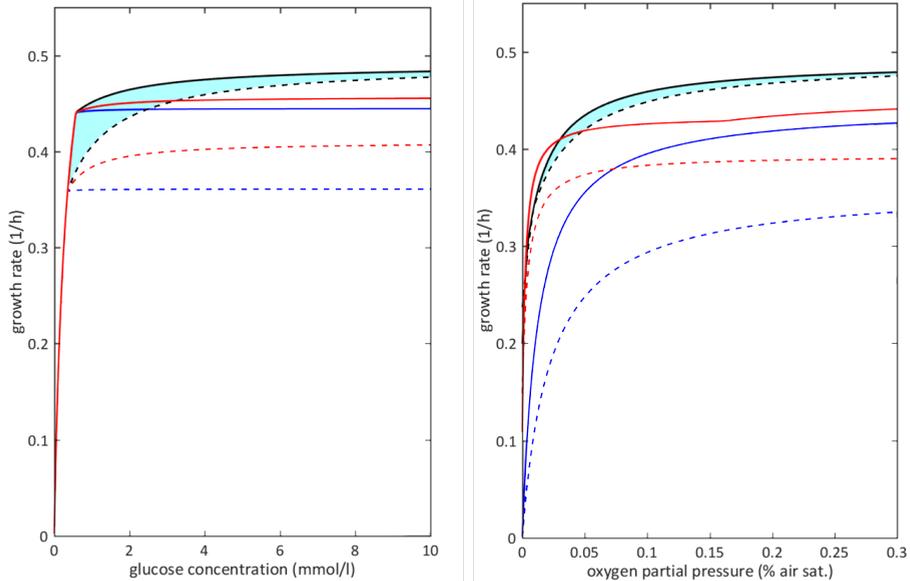


Figure 4.7. Comparison of fully respiratory (blue), Crabtree (black) and Warburg (red) under glucose (left) and oxygen limitation (right). Dashed line is $p_{mm} = p_{mm, min}$, solid line p_{mm} unbounded.

Comparing the predictions to the ethanol producing stain, it can be seen that this leads to the effect that a mitochondrial repression can no longer restore the maximal growth rate comparable to the findings in the Crabtree negative yeast (see Figure 4.7 left). This situation changes slightly in oxygen limited conditions (Figure 4.7 right), however also there the lactate producing phenotype with repressed mitochondria can only prevail in severe oxygen limitation below 0.07 % air saturation compared to the fully respiratory strategy. Also the model predicts that an anaerobic growing lactate producing cell is possible, where the maximal growth rate is estimated to ~ 0.2 1/h, about one third less than for anaerobic growth on ethanol.

Discussion

The Crabtree effect leads to a relaxation of mitochondrial function at increased glucose concentration, which can explain the observed variance in Crabtree effect in nature

It has been shown that the model exhibits a minimum threshold in p_{mm} from whereon the Crabtree effect can always restore the maximal growth rate (see cyan colored area in Figure 4.3) and the p_{mm} also directly determines the growth rate where fermentation starts. This has two major implications; (1) the fermentative pathway can substitute the mitochondrial function with respect to the maximal growth rate and (2) means that the growth rate from where the ethanol production starts is a free variable for cellular decision making. Therewith the model can reflect the relaxation of mitochondrial function as shown in genetic studies compared to Crabtree negative yeast⁹⁰ whereas the occurrence of the solution space can explain the observed variance in the occurrence of the Crabtree effect^{25,38} between different yeast strains and species.

However a relaxation of mitochondrial function i.e. a lower p_{mm} is shown to be strictly associated to a growth disadvantage under glucose limiting conditions. This growth disadvantage decreases with increasing residual concentration e.g. a phenotype with $p_{mm}=p_{mm,min}$ exhibits 95% of the maximal growth rate at a glucose concentration of about 5 mmol/l. From an evolutionary standpoint that means that phenotypes with a lower p_{mm} can only be competitive at elevated glucose concentrations whereas low glucose concentrations will pose an evolutionary pressure towards a higher p_{mm} .

High glucose concentrations and a glucose uniporter are necessary conditions for the emergence of the Crabtree effect, whereas anoxic conditions are not

There is limited information available to deduce how the environment looked like 100 million years ago in the Cretaceous age when the Crabtree effect arose²⁶. Although it has been hypothesized that the emergence of Crabtree positive yeast is linked⁹¹ to the diversification of flowering plants (angiosperms) from gymnosperms^{92,93}, which has taken place around the same time it is hard to infer the necessary environmental conditions for this evolutionary step making the presented model approach particularly useful.

The model simulation show that aerobic fermenting phenotypes can emerge from a non-fermentative progenitor in terms of growth rate when only two premises have

to be met (1) the mutation from a glucose proton symport to a glucose uniporter and (2) elevated glucose concentrations above ~1mmol/l in order to compensate for the low glucose affinity of the uniporter (see Figure 4.4). This transition may require as little as a single mutation in the H⁺ binding site, as shown in the glucose transporter from *Staphylococcus epidermidis* which shows high sequence homology with eukaryotic hexose transporters⁹⁴.

It was found that low oxygen partial pressures mostly favor Crabtree negative phenotypes as they exhibit a larger p_{mm} and therefore can take up more oxygen at the same oxygen partial pressure. The Crabtree positive phenotype can only compensate this disadvantage by fermentation at very low oxygen partial pressures (see Figure 4.6). Under this consideration the evolution of Crabtree effect in anaerobic conditions is conceivable but not likely also considering the high diffusivity of oxygen from the environment. The observed inability of anaerobic growth observed in some yeast is not caused by an energetic limitation, but by the inability of sterol import⁹⁵, because the synthesis of sterols is an oxygen requiring process⁹⁶. Research indicates that the ability for sterol import is not strictly associated with the occurrence of a Crabtree positive phenotype⁹⁷ and could have evolved independently.

Another prerequisite for the Crabtree effect is the expression of the enzymes pyruvate decarboxylase (PDC) and alcohol dehydrogenase (ADH) converting pyruvate to ethanol and CO₂. Also Crabtree negative yeast are able to ferment glucose to ethanol under oxygen limitation⁹⁸ and exhibit a phenomenon called short term Crabtree effect⁹⁹ during transition from glucose limitation to glucose excess. In oxygen limited chemostat cultures on glucose *C. utilis* CBS 621 exhibits specific ethanol production rates up to 15 mmol/gDW/h⁹⁸ and therewith in the range as also observed for *S. cerevisiae*. This is in agreement with the findings that the protein cost for PDC and ADH are negligible compared to the protein cost of glycolysis (see also Figure 4.S2) and indicates that the metabolic capacity for glucose uptake, glycolysis and the fermentative node may not have been the limiting steps for the emergence of the Crabtree effect.

The role of whole genome duplication (WGD) in the emergence of the Crabtree effect

The model shows that besides the glucose uniporter all essential proteins for the emergence of the Crabtree effect are included, raising the question about the role of the whole genome duplications (WGD) which has been linked to the emergence of the Crabtree effect^{18,24}.

Clearly WGD is a key process in the evolutionary history and duplicated genes can be found in almost all eukaryotic genomes. WGD is hypothesized to have two major functions in evolution (1) it creates a reproductive barrier, where the mechanism has been hypothesized to be hybridization followed by a doubling of the genome to restore fertility¹⁰⁰ and (2) it allows the divergence of gene functions. Modelling shows that genome duplication is negatively associated with evolutionary fitness and typically only few genes are retained⁹¹ favoring the reproductive barrier hypothesis. On the other hand research shows a correlation between WGD and the evolution of novel regulatory functions^{22,101-103} which could mean that the WGD played an important role in rewiring the metabolism towards glucose repression¹⁰⁴.

The rate-yield trade-off hypothesis (RYT) can explain the emergence of Crabtree positive yeast but not the evolutionary persistence of *S. cerevisiae* in its niche

The simulations show that the evolution of the Crabtree effect from a progenitor is possible, however at the same time it is shown that the Crabtree effect also impedes the evolution towards higher protein content, leading to the effect that a Crabtree negative phenotype with high protein content can gradually outcompete the Crabtree positive phenotype, also in the niche it emerged from (see Figure 4.5).

This finding strongly indicates that the trade-off between growth rate and growth yield as stated in the RYT is not the only relevant mechanism in explaining the evolutionary persistence of *S. cerevisiae* in high sugar environments, bringing the trait of ethanol tolerance into focus as emphasized by the make-accumulate-consume (MAC) hypothesis. Ethanol tolerance is often identified as relevant in batch competition experiments, where *S. cerevisiae* exhibits an extraordinary high ethanol tolerance compared to other yeast^{105,106}. The used competition experiment in chemostat culture leads to significantly lower ethanol concentrations compared to aerobic batch conditions due to the washout of ethanol. Experimentally the MIC (minimum inhibitory concentration) of *C. utilis* was determined to 6.3 % (v/v) ethanol and the K_i (50% growth inhibition concentration) of 3.4%(v/v) whereas *S. cerevisiae* shows about double the ethanol tolerance as well as double the K_i ^{107,108}.

Ethanol toxicity cannot explain the lower protein content of *S. cerevisiae* compared to most Crabtree positive yeast *per se*, as a higher overall protein content in the model is strictly linked to a higher maximal growth rate. This contradiction between the model prediction and the experimental observations in batch culture can only be resolved by a synergistic interrelation between ethanol tolerance and the protein

content i.e. lower total protein or mitochondrial protein content mitigates ethanol toxicity.

The molecular mechanisms of ethanol tolerance in yeast are very diverse as reviewed^{109,110} and not fully understood. To our knowledge no model exists to describe the effects of ethanol stress on yeast mechanistically and although ethanol tolerance has been linked to cellular energetics also no quantitative physiological data on the energetic cost of ethanol tolerance could be found, e.g. the Y_{ATP} . The model cannot definitely answer the question about the underlying fundamental mechanism and specific and systematic research on the quantitative physiology of yeast under ethanol stress would be required to answer the question.

Is the directed evolution towards a homolactic fermenting yeast possible from a PDC negative phenotype?

The model allows predicting the growth rate of a lactate producing strain by implementing a heterologous lactate dehydrogenase. It is further elaborated on the physiology of a homolactic strain, i.e. it does not produce ethanol and acetate, which is usually achieved by removing all PDC genes¹¹¹. Such strain has been shown not to grow anaerobically^{112,113}. The difference between Crabtree and the Warburg effect is that although both pathways lead to 2 ATP per glucose fermented the weak acid lactic acid which is dissociated at physiological pH requires a transporter to leave the cell. This means that the hexose and the lactate transporter both compete for limited membrane space and consequently leading to a trade-off between glucose uptake and lactate export. This trade-off eliminates the solution space as observed for the Crabtree effect i.e. no $p_{mm,min}$ exists; this means that such phenotype can never compensate the loss in growth rate upon a lowering in p_{mm} by lactate fermentation.

This also means that an analogy between the Crabtree effect and Warburg effect should be made cautiously. In spite of this trade-off, it cannot explain the lack of an anaerobic growing lactate fermenting strain, where the model predicts a maximal growth rate of about 0.2 1/h. However the trait only leads to a clear growth advantage independent of the p_{mm} under severely oxygen limited conditions (see Figure 4.7). However the original function of JEN1 is lactate uptake, therefore the transporter is subject to a phenomenon called glucose repression (or more general carbon catabolite repression). This means that *S. cerevisiae* prefers the uptake of glucose even when lactate is present in the environment. Therefore, the activity of the promoter of JEN1 was shown to be inversely related to the sugar concentration^{114,115} a behavior which can be also predicted by the model (Figure

4.S4). This repression is more complicated and takes also place on the post-transcriptional¹¹⁶ and post-translational level¹¹⁷. However it seems to be possible to select mutants not showing glucose repression of lactate uptake¹¹⁸.

Additionally, the transporter thermodynamics have to be considered; clearly the ATP yield of 2 ATP per glucose can only be conserved when the lactate is exported with a lactate-proton symport. Such transporter has a maximal out/in ratio of 1 when the cytosolic pH is equal to the extracellular pH and decreasing with lower pH¹¹⁹. This means a homolactic fermenting strain will likely only be able to grow at near neutral pH. Secondly, this means that the JEN1 transporter is ineffective in mitigating weak acid stress making it likely that there are additional active exporters that will have to be removed. To our best knowledge these transporters have not been identified therewith making a rational engineering of a homolactic engineering strain tedious, whereas directed evolution could lead to the desired outcome.

General Conclusions

The remarkable capacity of a multiple constrained structured mFBA model for the description of eukaryotic physiology was demonstrated. It was shown that the approach is useful in understand evolution, also with respect to laboratory evolution and allows to deduce consistent trajectories of yeast evolution towards the Crabtree effect. The particular strength of the approach is the capability in predicting metabolic trade-offs with respect to protein allocation e.g. in catabolism or carbon catabolite repression.

The presented model should be extended, here the major bottleneck are *in vivo* specific protein activities in particular for isoenzymes and transporters. It is believed that the approach can be useful in exploring many metabolic phenomena, especially metabolic trade-offs, on a more mechanistic level and will become a standard tool in future.

Symbols and Abbreviations

a	Specific protein activity
DW	Dry weight
K_m	Affinity constant
N	Stoichiometry matrix
p	Protein content
v	flux
Y	Yield
σ	Enzyme saturation factor
μ	Growth rate
PDC	Pyruvate decarboxylase

Subscript/Superscript

ATP	Adenosine triphosphate
$const$	Constitutive/growth independent
cm	Cell membrane
$flux$	Flux associated
i	Generic variable for fluxes
max	maximal
min	minimal
mm	Mitochondrial membrane
raw	Cellular dry weight without protein
$total$	Total cellular protein

References

- 1 Dmitriew, C. M. The evolution of growth trajectories: what limits growth rate? *Biological reviews of the Cambridge Philosophical Society* 86, 97-116 (2011).
- 2 Molenaar, D., Van Berlo, R., De Ridder, D. & Teusink, B. Shifts in growth strategies reflect tradeoffs in cellular economics. *Molecular systems biology* 5, 323 (2009).
- 3 Bachmann, H. *et al.* Availability of public goods shapes the evolution of competing metabolic strategies. *Proceedings of the National Academy of Sciences of the United States of America* 110, 14302-14307 (2013).
- 4 Sauer, U. in *Metabolic Engineering* (eds Jens Nielsen *et al.*) 129-169 (Springer Berlin Heidelberg, 2001).
- 5 Dykhuizen, D. E. & Hartl, D. L. Selection in chemostats. *Microbiological Reviews* 47, 150-168 (1983).
- 6 Verduyn, C., Stouthamer, A. H., Scheffers, W. A. & Dijken, J. P. A theoretical evaluation of growth yields of yeasts. *Antonie Van Leeuwenhoek* 59 (1991).
- 7 Verduyn, C., Postma, E., Scheffers, W. A. & van Dijken, J. P. Physiology of *Saccharomyces cerevisiae* in anaerobic glucose-limited chemostat cultures. *J Gen Microbiol* 136, 395-403 (1990).
- 8 Stouthamer, A. H. A theoretical study on the amount of ATP required for synthesis of microbial cell material. *Antonie van Leeuwenhoek* 39, 545-565 (1973).
- 9 Verduyn, C. in *Quantitative Aspects of Growth and Metabolism of Microorganisms* (ed A. H. Stouthamer) 325-353 (Springer Netherlands, 1992).
- 10 Warburg, O., Wind, F. & Negelein, E. THE METABOLISM OF TUMORS IN THE BODY. *The Journal of general physiology* 8, 519-530 (1927).
- 11 Crabtree, H. G. The carbohydrate metabolism of certain pathological overgrowths. *Biochemical Journal* 22, 1289-1298 (1928).
- 12 Waddell, T. G., Repovic, P., Melendez-Hevia, E., Heinrich, R. & Montero, F. Optimization of glycolysis: A new look at the efficiency of energy coupling. *Biochemical Education* 25, 204-205 (1997).
- 13 Bachmann, H., Bruggeman, F. J., Molenaar, D., Branco dos Santos, F. & Teusink, B. Public goods and metabolic strategies. *Current Opinion in Microbiology* 31, 109-115 (2016).
- 14 Heinrich, R., Montero, F., Klipp, E., Waddell, T. G. & Melendez-Hevia, E. Theoretical approaches to the evolutionary optimization of glycolysis: thermodynamic and kinetic constraints. *European journal of biochemistry / FEBS* 243, 191-201 (1997).
- 15 Westerhoff, H. V., Hellingwerf, K. J. & Van Dam, K. Thermodynamic efficiency of microbial growth is low but optimal for maximal growth rate. *Proceedings*

- of the National Academy of Sciences of the United States of America 80, 305-309 (1983).
- 16 Pfeiffer, T., Schuster, S. & Bonhoeffer, S. Cooperation and Competition in the Evolution of ATP-Producing Pathways. *Science* 292, 504-507 (2001).
- 17 Pfeiffer, T. & Bonhoeffer, S. in *Zeitschrift für Physikalische Chemie* Vol. 216 51 (2002).
- 18 Pfeiffer, T. & Morley, A. An evolutionary perspective on the Crabtree effect. *Frontiers in Molecular Biosciences* 1, 17 (2014).
- 19 Hardin, G. The Tragedy of the Commons. *Science* 162, 1243-1248 (1968).
- 20 MacLean, R. C. The tragedy of the commons in microbial populations: insights from theoretical, comparative and experimental studies. *Heredity* 100, 233-239 (2008).
- 21 MacLean, R. C. & Gudelj, I. Resource competition and social conflict in experimental populations of yeast. *Nature* 441, 498-501 (2006).
- 22 Piskur, J., Rozpedowska, E., Polakova, S., Merico, A. & Compagno, C. How did *Saccharomyces* evolve to become a good brewer? *Trends in genetics : TIG* 22, 183-186 (2006).
- 23 Hagman, A., Sall, T., Compagno, C. & Piskur, J. Yeast "make-accumulate-consume" life strategy evolved as a multi-step process that predates the whole genome duplication. *PLoS one* 8, e68734 (2013).
- 24 Dashko, S., Zhou, N., Compagno, C. & Piškur, J. Why, when, and how did yeast evolve alcoholic fermentation? *Fems Yeast Research* 14, 826-832 (2014).
- 25 Hagman, A. & Piškur, J. A Study on the Fundamental Mechanism and the Evolutionary Driving Forces behind Aerobic Fermentation in Yeast. *PLoS one* 10, e0116942 (2015).
- 26 Goddard, M. R. & Greig, D. *Saccharomyces cerevisiae*: a nomadic yeast with no niche? *FEMS Yeast Res* 15 (2015).
- 27 Flamholz, A., Noor, E., Bar-Even, A., Liebermeister, W. & Milo, R. Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proceedings of the National Academy of Sciences of the United States of America* 110, 10039-10044 (2013).
- 28 Beg, Q. K. *et al.* Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity. *Proceedings of the National Academy of Sciences of the United States of America* 104, 12663-12668 (2007).
- 29 Zhuang, K., Vemuri, G. N. & Mahadevan, R. Economics of membrane occupancy and respiro-fermentation. *Molecular Systems Biology* 7, 500-500 (2011).
- 30 Basan, M. *et al.* Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. *Nature* 528, 99-104 (2015).
- 31 Orth, J. D., Thiele, I. & Palsson, B. O. What is flux balance analysis? *Nat Biotech* 28, 245-248 (2010).

- 32 Harcombe, W. R., Delaney, N. F., Leiby, N., Klitgord, N. & Marx, C. J. The ability of flux balance analysis to predict evolution of central metabolism scales with the initial distance to the optimum. *PLoS computational biology* 9, e1003091 (2013).
- 33 Großkopf, T. *et al.* Metabolic modelling in a dynamic evolutionary framework predicts adaptive diversification of bacteria in a long-term evolution experiment. *BMC Evolutionary Biology* 16, 163 (2016).
- 34 Varma, A. & Palsson, B. O. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl Environ Microbiol* 60, 3724-3731 (1994).
- 35 Berkhout, J. *et al.* How biochemical constraints of cellular growth shape evolutionary adaptations in metabolism. *Genetics* 194, 505-512 (2013).
- 36 Adadi, R., Volkmer, B., Milo, R., Heinemann, M. & Shlomi, T. Prediction of microbial growth rate versus biomass yield by a metabolic network with kinetic parameters. *PLoS computational biology* 8, e1002575 (2012).
- 37 Nilsson, A. & Nielsen, J. Metabolic Trade-offs in Yeast are Caused by F1FO-ATP synthase. *Scientific reports* 6, 22264 (2016).
- 38 De Deken, R. H. The Crabtree effect: a regulatory system in yeast. *J Gen Microbiol* 44, 149-156 (1966).
- 39 van Urk, H., Postma, E., Scheffers, W. A. & van Dijken, J. P. Glucose transport in crabtree-positive and crabtree-negative yeasts. *J Gen Microbiol* 135, 2399-2406 (1989).
- 40 Pfeiffer, T., Schuster, S. & Bonhoeffer, S. Cooperation and competition in the evolution of ATP-producing pathways. *Science* 292, 504-507 (2001).
- 41 Gulik, W. M. v. & Heijnen, J. J. A metabolic network stoichiometry analysis of microbial growth and product formation. *Biotechnology and Bioengineering* 48, 681-698 (1995).
- 42 Stückerath, I. *et al.* Characterization of null mutants of the glyoxylate cycle and gluconeogenic enzymes in *S. cerevisiae* through metabolic network modeling verified by chemostat cultivation. *Biotechnology and Bioengineering* 77, 61-72 (2002).
- 43 Verduyn, C. Physiology of yeasts in relation to biomass yields. *Antonie van Leeuwenhoek* 60, 325-353 (1991).
- 44 Verduyn, C., Stouthamer, A. H., Scheffers, W. A. & van Dijken, J. P. A theoretical evaluation of growth yields of yeasts. *Antonie van Leeuwenhoek* 59, 49-63 (1991).
- 45 Babel, W. & Müller, R. H. Correlation between cell composition and carbon conversion efficiency in microbial growth: a theoretical study. *Applied Microbiology and Biotechnology* 22, 201-207 (1985).
- 46 Lange, H. C. & Heijnen, J. J. Statistical reconciliation of the elemental and molecular biomass composition of *Saccharomyces cerevisiae*. *Biotechnology and Bioengineering* 75, 334-344 (2001).

- 47 Verduyn, C., Postma, E., Scheffers, W. A. & van Dijken, J. P. Energetics of
Saccharomyces cerevisiae in anaerobic glucose-limited chemostat cultures. *J*
Gen Microbiol 136, 405-412 (1990).
- 48 Hoek, P., Dijken, J. P. & Pronk, J. T. Effect of specific growth rate on
fermentative capacity of baker's yeast. *Appl Environ Microbiol* 64 (1998).
- 49 Stouthamer, A. H. Theoretical study on amount of ATP required for
synthesis of microbial cell material. *Antonie Van Leeuwenhoek* 39 (1973).
- 50 Lynch, M. & Marinov, G. K. The bioenergetic costs of a gene. *Proceedings of*
the National Academy of Sciences 112, 15690-15695 (2015).
- 51 Wagner, A. Energy Constraints on the Evolution of Gene Expression.
Molecular Biology and Evolution 22, 1365-1374 (2005).
- 52 Förster, J., Famili, I., Fu, P., Palsson, B. Ø. & Nielsen, J. Genome-Scale
Reconstruction of the Saccharomyces cerevisiae Metabolic Network. *Genome*
research 13, 244-253 (2003).
- 53 Weijland, A. & Parmeggiani, A. Toward a model for the interaction between
elongation factor Tu and the ribosome. *Science* 259, 1311-1314 (1993).
- 54 Schimmel, P. GTP hydrolysis in protein synthesis: two for Tu? *Science* 259,
1264-1265 (1993).
- 55 Kruckeberg, A. L., Ye, L., Berden, J. A. & van Dam, K. Functional expression,
quantification and cellular localization of the Hxt2 hexose transporter of
Saccharomyces cerevisiae tagged with the green fluorescent protein.
Biochem J 339 (Pt 2), 299-307 (1999).
- 56 Boles, E. & Hollenberg, C. P. The molecular genetics of hexose transport in
yeasts. *FEMS Microbiol Rev* 21, 85-111 (1997).
- 57 Reifemberger, E., Boles, E. & Ciriacy, M. Kinetic characterization of individual
hexose transporters of Saccharomyces cerevisiae and their relation to the
triggering mechanisms of glucose repression. *European Journal of*
Biochemistry 245, 324-333 (1997).
- 58 Burke, P. V., Raitt, D. C., Allen, L. A., Kellogg, E. A. & Poyton, R. O. Effects of
oxygen concentration on the expression of cytochrome c and cytochrome c
oxidase genes in yeast. *J Biol Chem* 272, 14705-14712 (1997).
- 59 Lichtenberg-Frate, H., Naeschen, T., Heiland, S. & Hofer, M. Properties and
Heterologous Expression of the Glucose Transporter GHT1 from
Schizosaccharomyces pombe. *Yeast* 13, 215-224 (1997).
- 60 Paiva, S., Kruckeberg, A. L. & Casal, M. Utilization of green fluorescent
protein as a marker for studying the expression and turnover of the
monocarboxylate permease Jen1p of Saccharomyces cerevisiae. *Biochem J*
363, 737-744 (2002).
- 61 Strijbis, K. & Distel, B. Intracellular acetyl unit transport in fungal carbon
metabolism. *Eukaryot Cell* 9, 1809-1815 (2010).

- 62 Palmieri, L. *et al.* The mitochondrial dicarboxylate carrier is essential for the growth of *Saccharomyces cerevisiae* on ethanol or acetate as the sole carbon source. *Mol Microbiol* 31, 569-577 (1999).
- 63 Verduyn, C., Postma, E., Scheffers, W. A. & Dijken, J. P. Physiology of *Saccharomyces cerevisiae* in anaerobic glucose-limited chemostat cultures. *J Gen Microbiol* 136 (1990).
- 64 Visser, W. *et al.* Effects of growth conditions on mitochondrial morphology in *Saccharomyces cerevisiae*. *Antonie van Leeuwenhoek* 67, 243-253 (1995).
- 65 Van Dijken, J. P., Weusthuis, R. A. & Pronk, J. T. Kinetics of growth and sugar consumption in yeasts. *Antonie van Leeuwenhoek* 63, 343-352 (1993).
- 66 Rosenfeld, E., Beauvoit, B., Rigoulet, M. & Salmon, J. M. Non-respiratory oxygen consumption pathways in anaerobically-grown *Saccharomyces cerevisiae*: evidence and partial characterization. *Yeast* 19, 1299-1321 (2002).
- 67 Dodia, R., Meunier, B., Kay, Christopher W. M. & Rich, Peter R. Comparisons of subunit 5A and 5B isoenzymes of yeast cytochrome c oxidase. *Biochemical Journal* 464, 335-342 (2014).
- 68 Spor, A. *et al.* Niche-driven evolution of metabolic and life-history strategies in natural and domesticated populations of *Saccharomyces cerevisiae*. *BMC Evolutionary Biology* 9, 296 (2009).
- 69 Boynton, P. J. & Greig, D. The ecology and evolution of non-domesticated *Saccharomyces* species. *Yeast* 31, 449-462 (2014).
- 70 Entian, K.-D. & Kötter, P. in *Methods in Microbiology* Vol. Volume 36 (eds Stansfield Ian & J. R. Stark Michael) 629-666 (Academic Press, 2007).
- 71 Nijkamp, J. F. *et al.* De novo sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN.PK113-7D, a model for modern industrial biotechnology. *Microb Cell Fact* 11 (2012).
- 72 Vanhalewyn, M. *et al.* A mutation in *Saccharomyces cerevisiae* adenylate cyclase, Cyr1K1876M, specifically affects glucose- and acidification-induced cAMP signalling and not the basal cAMP level. *Molecular Microbiology* 33, 363-376 (1999).
- 73 Fang, M. & Butow, R. A. Nucleotide reversal of mitochondrial repression in *saccharomyces cerevisiae*. *Biochemical and Biophysical Research Communications* 41, 1579-1583 (1970).
- 74 Diaz-Ruiz, R., Rigoulet, M. & Devin, A. The Warburg and Crabtree effects: On the origin of cancer cell energy metabolism and of yeast glucose repression. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1807, 568-576 (2011).
- 75 Hoek, P., Dijken, J. P. & Pronk, J. T. Regulation of fermentative capacity and levels of glycolytic enzymes in chemostat cultures of *Saccharomyces cerevisiae*. *Enzyme Microb Tech* 26 (2000).
- 76 Sniegowski, P. D., Dombrowski, P. G. & Fingerman, E. *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* coexist in a natural woodland site in

- North America and display different levels of reproductive isolation from European conspecifics. *FEMS Yeast Research* 1, 299-306 (2002).
- 77 Postma, E., Alexander Scheffers, W. & Van Dijken, J. P. Kinetics of growth and glucose transport in glucose-limited chemostat cultures of *Saccharomyces cerevisiae* CBS 8066. *Yeast* 5, 159-165 (1989).
- 78 van Dijken, J. P. *et al.* An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. *Enzyme Microb Technol* 26, 706-714 (2000).
- 79 Postma, E., Kuiper, A., Tomasouw, W. F., Scheffers, W. A. & van Dijken, J. P. Competition for glucose between the yeasts *Saccharomyces cerevisiae* and *Candida utilis*. *Appl Environ Microbiol* 55, 3214-3220 (1989).
- 80 Verduyn, C., Postma, E., Scheffers, W. A. & Dijken, J. P. Effect of benzoic acid on metabolic fluxes in yeasts—a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast* 8 (1992).
- 81 Verduyn, C., Postma, E., Scheffers, W. A. & Van Dijken, J. P. Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast* 8, 501-517 (1992).
- 82 Narendranath, N. V., Thomas, K. C. & Ingledew, W. M. Effects of acetic acid and lactic acid on the growth of *Saccharomyces cerevisiae* in a minimal medium. *Journal of Industrial Microbiology and Biotechnology* 26, 171-177 (2001).
- 83 Otterstedt, K. *et al.* Switching the mode of metabolism in the yeast *Saccharomyces cerevisiae*. *EMBO Reports* 5, 532-537 (2004).
- 84 Dejean, L., Beauvoit, B., Guerin, B. & Rigoulet, M. Growth of the yeast *Saccharomyces cerevisiae* on a non-fermentable substrate: control of energetic yield by the amount of mitochondria. *Biochimica et biophysica acta* 1457, 45-56 (2000).
- 85 Noorman, H. J., Baksteen, J., Heijnen, J. J. & Luyben, C. A. M. K. The bioreactor overflow device: an undesired selective separator in continuous cultures? *Microbiology* 137, 2171-2177 (1991).
- 86 Ambesi, A., Miranda, M., Petrov, V. V. & Slayman, C. W. Biogenesis and function of the yeast plasma-membrane H (+)-ATPase. *Journal of Experimental Biology* 203, 155-160 (2000).
- 87 Fonseca, G. G., Gombert, A. K., Heinzle, E. & Wittmann, C. Physiology of the yeast *Kluyveromyces marxianus* during batch and chemostat cultures with glucose as the sole carbon source. *FEMS Yeast Res* 7, 422-435 (2007).
- 88 Averet, N., Jobin, M. L., Devin, A. & Rigoulet, M. Proton pumping complex I increases growth yield in *Candida utilis*. *Biochimica et biophysica acta* 1847, 1320-1326 (2015).

- 89 Bakker, B. M. *et al.* Stoichiometry and compartmentation of NADH metabolism in *Saccharomyces cerevisiae*. *FEMS Microbiology Reviews* 25, 15-37 (2001).
- 90 Jiang, H., Guan, W., Pinney, D., Wang, W. & Gu, Z. Relaxation of yeast mitochondrial functions after whole-genome duplication. *Genome research* 18, 1466-1471 (2008).
- 91 van Hoek, M. J. & Hogeweg, P. Metabolic adaptation after whole genome duplication. *Mol Biol Evol* 26, 2441-2453 (2009).
- 92 De Bodt, S., Maere, S. & Van de Peer, Y. Genome duplication and the origin of angiosperms. *Trends in Ecology & Evolution* 20, 591-597 (2005).
- 93 Project, A. G. The Amborella genome and the evolution of flowering plants. *Science* 342, 1241089 (2013).
- 94 Iancu, C. V., Zmoon, J., Woo, S. B., Aleshin, A. & Choe, J. Y. Crystal structure of a glucose/H⁺ symporter and its mechanism of action. *Proceedings of the National Academy of Sciences of the United States of America* 110, 17862-17867 (2013).
- 95 Jacquier, N. & Schneider, R. Mechanisms of sterol uptake and transport in yeast. *The Journal of steroid biochemistry and molecular biology* 129, 70-78 (2012).
- 96 Hannich, J. T., Umebayashi, K. & Riezman, H. Distribution and Functions of Sterols and Sphingolipids. *Cold Spring Harbor Perspectives in Biology* 3, a004762 (2011).
- 97 Zavrel, M., Hoot, S. J. & White, T. C. Comparison of Sterol Import under Aerobic and Anaerobic Conditions in Three Fungal Species, *Candida albicans*, *Candida glabrata*, and *Saccharomyces cerevisiae*. *Eukaryotic Cell* 12, 725-738 (2013).
- 98 Weusthuis, R. A., Visser, W., Pronk, J. T., Scheffers, W. A. & van Dijken, J. P. Effects of oxygen limitation on sugar metabolism in yeasts: a continuous-culture study of the Kluyver effect. *Microbiology* 140 (Pt 4), 703-715 (1994).
- 99 Van Urk, H., Mark, P. R., Scheffers, W. A. & Van Dijken, J. P. Metabolic responses of *Saccharomyces cerevisiae* CBS 8066 and *Candida utilis* CBS 621 upon transition from glucose limitation to glucose excess. *Yeast* 4, 283-291 (1988).
- 100 Wolfe, K. H. Origin of the Yeast Whole-Genome Duplication. *PLoS Biology* 13, e1002221 (2015).
- 101 Goode, D. K., Callaway, H. A., Cerda, G. A., Lewis, K. E. & Elgar, G. Minor change, major difference: divergent functions of highly conserved cis-regulatory elements subsequent to whole genome duplication events. *Development (Cambridge, England)* 138, 879-884 (2011).
- 102 Ihmels, J. *et al.* Rewiring of the Yeast Transcriptional Network Through the Evolution of Motif Usage. *Science* 309, 938-940 (2005).

- 103 Arsovski, A. A., Pradinuk, J., Guo, X. Q., Wang, S. & Adams, K. L. Evolution of
Cis-Regulatory Elements and Regulatory Networks in Duplicated Genes of
Arabidopsis. *Plant Physiology* 169, 2982-2991 (2015).
- 104 Kayikci, Ö. & Nielsen, J. Glucose repression in *Saccharomyces cerevisiae*. *FEMS
Yeast Research* 15 (2015).
- 105 Williams, K. M., Liu, P. & Fay, J. C. Evolution of ecological dominance of yeast
species in high-sugar environments. *Evolution* 69, 2079-2093 (2015).
- 106 Zhou, N. *et al.* Coevolution with bacteria drives the evolution of aerobic
fermentation in *Lachancea kluyveri*. *PLoS one* 12, e0173318 (2017).
- 107 Antoce, O. A. Characterization of ethanol tolerance of yeasts using a
calorimetric technique. *VITIS* 35, 105-106 (1996).
- 108 Prior, B., Kilian, S. & Lategan, P. Growth of *Candida utilis* on ethanol and
isopropanol. *Archives of Microbiology* 125, 133-136 (1980).
- 109 Stanley, D., Bandara, A., Fraser, S., Chambers, P. J. & Stanley, G. A. The
ethanol stress response and ethanol tolerance of *Saccharomyces cerevisiae*.
Journal of Applied Microbiology 109, 13-24 (2010).
- 110 Ma, M. & Liu, Z. L. Mechanisms of ethanol tolerance in *Saccharomyces
cerevisiae*. *Applied Microbiology and Biotechnology* 87, 829-845 (2010).
- 111 van Maris, A. J. *et al.* Directed evolution of pyruvate decarboxylase-negative
Saccharomyces cerevisiae, yielding a C2-independent, glucose-tolerant, and
pyruvate-hyperproducing yeast. *Appl Environ Microbiol* 70, 159-166 (2004).
- 112 Abbott, D. A., van den Brink, J., Minneboo, I. M., Pronk, J. T. & van Maris, A.
J. Anaerobic homolactate fermentation with *Saccharomyces cerevisiae*
results in depletion of ATP and impaired metabolic activity. *FEMS Yeast Res*
9, 349-357 (2009).
- 113 van Maris, A. J. A., Winkler, A. A., Porro, D., van Dijken, J. P. & Pronk, J. T.
Homofermentative Lactate Production Cannot Sustain Anaerobic Growth of
Engineered *Saccharomyces cerevisiae*: Possible Consequence of Energy-
Dependent Lactate Export. *Applied and Environmental Microbiology* 70,
2898-2905 (2004).
- 114 Chambers, P., Issaka, A. & Palecek, S. P. *Saccharomyces cerevisiae* JEN1
Promoter Activity Is Inversely Related to Concentration of Repressing Sugar.
Applied and Environmental Microbiology 70, 8-17 (2004).
- 115 Lascaris, R. *et al.* Overexpression of HAP4 in glucose-derepressed yeast cells
reveals respiratory control of glucose-regulated genes. *Microbiology* 150,
929-934 (2004).
- 116 Andrade, R. P., Kotter, P., Entian, K. D. & Casal, M. Multiple transcripts
regulate glucose-triggered mRNA decay of the lactate transporter JEN1
from *Saccharomyces cerevisiae*. *Biochem Biophys Res Commun* 332, 254-262
(2005).

- 117 Paiva, S. *et al.* Glucose-induced ubiquitylation and endocytosis of the yeast Jen1 transporter: role of lysine 63-linked ubiquitin chains. *J Biol Chem* 284, 19228-19236 (2009).
- 118 Tsuboi, H. *et al.* Analysis of the pyruvate permease gene (JEN1) in glucose derepression yeast (*Saccharomyces cerevisiae*) Isolated from a 2-deoxyglucose-tolerant mutant, and its application to sake making. *Bioscience, biotechnology, and biochemistry* 67, 765-771 (2003).
- 119 van Maris, A. J., Konings, W. N., van Dijken, J. P. & Pronk, J. T. Microbial export of lactic and 3-hydroxypropanoic acid: implications for industrial fermentation processes. *Metab Eng* 6, 245-255 (2004).

Supporting Appendix

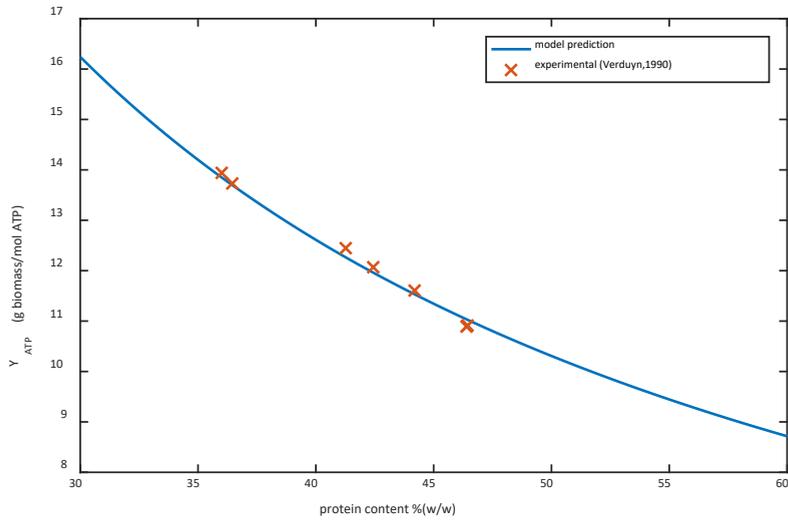


Figure 4.S1. Experimental observables and model fit for the biomass yield on ATP as a function of the protein content.

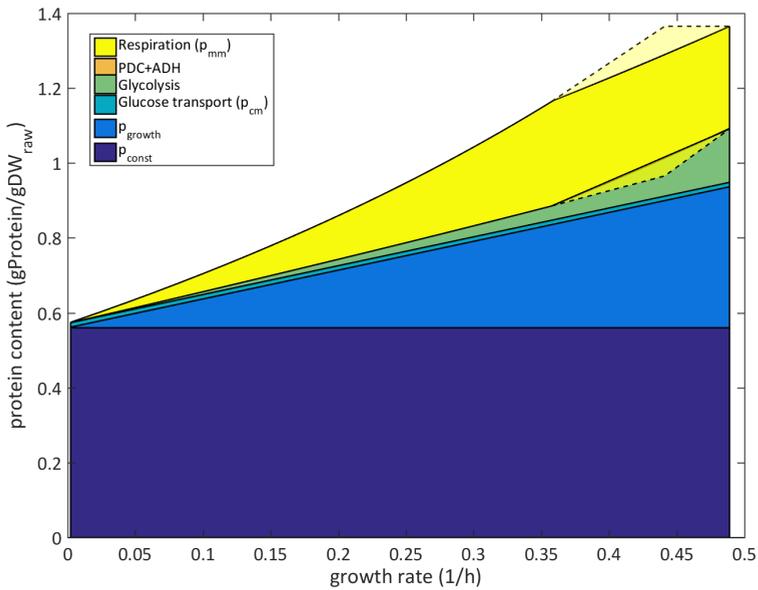


Figure 4.S2. Predicted protein fractions (by function) under glucose limited conditions, (--) $p_{mm}=p_{mm,min}$ and (-) p_{mm} unbounded. The enzyme amount for PDH and ADH is so small that it is not visible in the figure.

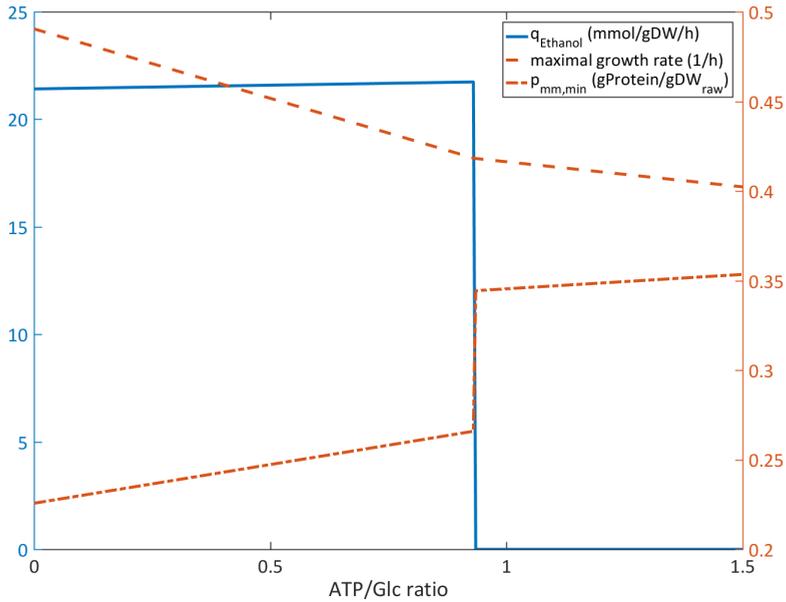


Figure 4.S3. Effect of the ATP expenditure in glucose import on the Crabtree effect. An abscissa value of 0 corresponds to the glucose uniporter and a value of 1 to the glucose-proton symport.

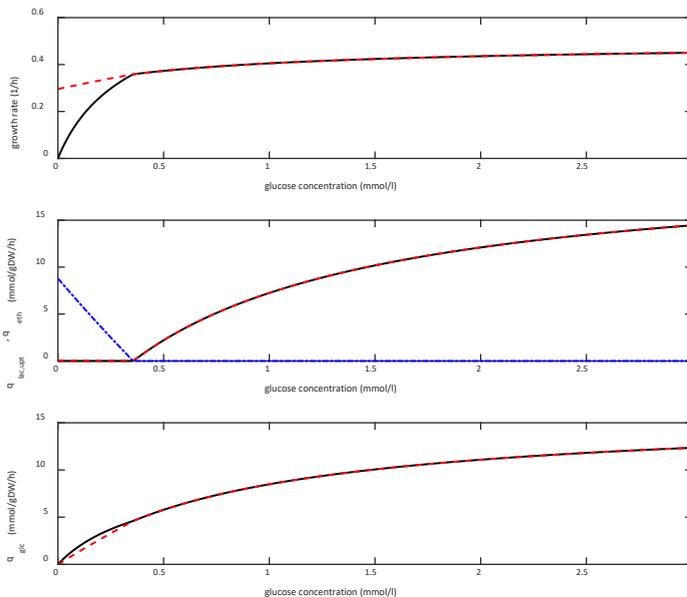


Figure 4.S4. Prediction of the physiology in presence of glucose and lactate excess (red and blue (q_{lac})) and glucose only (black) with respect to growth rate (top), lactate uptake and ethanol production rate (middle) and glucose uptake rate (bottom).

Table 4.S1. Specific protein activities for the network reactions

		Enzymes	a_{cm} (mmol/h/ mg)	a_{mm} (mmol/h/ mg)	a_{total} (mmol/h/ mg)
HXT (Glucose import)	$Glc_{cec} \rightarrow Glc$ (uniporter) or Glc_{cec} + 1ATP \rightarrow Glc (glucose-proton symport)		3.19		3.19
JEN1 (Lactic acid export)	$HLac \rightarrow HLa_{cec}$		6.38		6.38
JEN1 (Lactic acid import)	$Hla_{cec} \rightarrow HLac$		6.38		6.38
Glycolysis	$Glc \rightarrow 2Pyr + 2ATP$ + 2NADH	HXK + PGI + PFK + FBA + TPI + 2 GLD + 2 PGK + 2 GPM + 2 ENO + 2 CDC + 2 PK			0.52
PDC + ADH	$Pyr + NADH \rightarrow EtOH$ + CO_2				88.86
LDH	$Pyr + NADH \rightarrow HLac$				137.52
MPC_fwd	$Pyr \rightarrow Pyr_{mit}$			2.952	2.952
MPC_bwd	$Pyr_{mit} \rightarrow Pyr$			2.952	2.952
TCA cycle	$Pyr_{mit} \rightarrow 3CO_2 +$ $4NADH + FADH_2 + ATP$	PDH + CIT + ACO + IDH + KGDKG2 + SDH12 + MDH1 + FUM1			0.272
NDI, NDE	$NADH \rightarrow QH_2$			32.14	32.14
SDH34	$FADH_2 \rightarrow QH_2$			18.80	18.80
2RIP1 + COX	$O_2 + 2QH_2 + 9H^+_{mit}$ $\rightarrow 9H^+$			1.25	1.25
ATP1	$3.75H^+ \rightarrow ATP +$ $3.75 H^+_{mit}$			0.672	0.672
DW_{raw}	0.413 Pyr + 0.005NH ₃ + 0.65 ATP + 0.15 H ₂ O \rightarrow $1C_1H_{1.94}O_{0.91}N_{0.005} +$ 0.014 NADH + 0.24 CO ₂				
Protein	0.413 Pyr + 0.275NH ₃ + 4.59 ATP \rightarrow				0.091

	$1C_1H_{1.581}O_{0.318}N_{0.275}$ + 0.0065 NADH + 0.24 CO ₂ + 0.442 H ₂ O				
Glycerol production	0.5 Glc + NADH + ATP → Glycerol				0
CYB2 + 0.5 COX	HLac + 0.5 O ₂ + 3H ⁺ _{mit} → Pyr + 3H ⁺			2.86	2.86
ADH+ALD	EtOH + 2ATP → AcCoA + 2NADH				1.99
Glyoxylate cycle + SFC + FUM1 + PCK + PK	2 AcCoA → Pyr + CO ₂ + 2NADH + 1 FADH ₂	CIT + ACO + ICL1 +MLS1 + MDH1 + FUM1 + PCK + PK + SFC		2 (only SFC)	0.17
TCA cycle AcCoA + CAT2 + CRC1	AcCoA → 2 CO ₂ + 3NADH + FADH + ATP	CIT + ACO + IDH + KGDKGD2 + SDH12 + MDH1 + FUM1 + CAT2 + CRC1		2 (only CRC1)	0.33

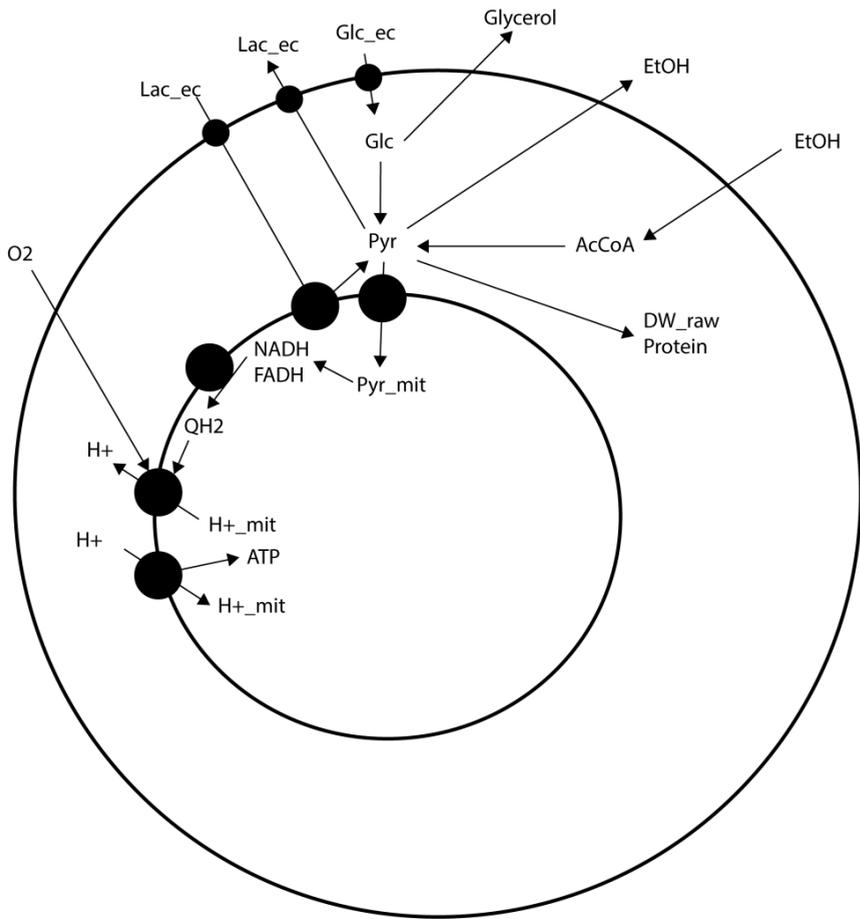


Figure 4.S5. Visualization of the model stoichiometry

Chapter 5:

Effective Estimation of Dynamic Metabolic Fluxes Using ^{13}C Labeling and Piecewise Affine Approximation: From Theory to Practical Applicability

in collaboration with S.A. Wahl

"All entities move and nothing remains still"

Heraclitus in Plato's Cratylus

Published as:

Schumacher, R.; Wahl, S.A. Effective Estimation of Dynamic Metabolic Fluxes Using ^{13}C Labeling and Piecewise Affine Approximation: From Theory to Practical Applicability. *Metabolites* 2015, 5, 697-719.

Abstract

The design of microbial production processes relies on rational choices for metabolic engineering of the production host and the process conditions. These require a systematic and quantitative understanding of cellular regulation. Therefore, a novel method for dynamic flux identification using quantitative metabolomics and ^{13}C labeling to identify piecewise-affine (PWA) flux functions has been described recently. Obtaining flux estimates nevertheless still required frequent manual reinitialization to obtain a good reproduction of the experimental data and, moreover, did not optimize on all observables simultaneously (metabolites and isotopomer concentrations). In our contribution we focus on measures to achieve faster and robust dynamic flux estimation which leads to a high dimensional parameter estimation problem. Specifically, we address the following challenges within the PWA problem formulation: (1) Fast selection of sufficient domains for the PWA flux functions, (2) Control of over-fitting in the concentration space using shape-prescriptive modeling and (3) robust and efficient implementation of the parameter estimation using the hybrid implicit filtering algorithm. With the improvements we significantly speed up the convergence by effectively exploiting that the optimization problem is partly linear and reducing the search space with the constraints. This allows application to larger-scale metabolic networks and demonstrates that the proposed approach is not purely theoretical, but also practically applicable.

Introduction

In the natural environment, microorganisms are exposed to transient environmental conditions that can trigger a wide range of genetic, regulatory and metabolic responses to adapt and survive ^{1,2}. These adaptation mechanisms are described to operate at very different timescales:

- (1) Genetic adaptation by mutagenesis and selection with a timescale of several generation times ³,
- (2) Adaptation of gene expression levels with a timescale of minutes ⁴,
- (3) Post-translational modifications with a time constant in the order of seconds and
- (4) Kinetic response which is considered an inherent property of the enzymes in the metabolic network and therefore persistent.

Kinetic metabolic regulation is a result of complex interactions between enzymes, substrates and allosteric effectors. Enzyme kinetics can be investigated *in vitro*, but the validity of estimated kinetic parameters has been shown to be limited for reconstruction of the observed *in vivo* metabolic network properties ⁵. *In vivo*, only the whole network response can be analyzed, making the identification of kinetics highly challenging. Pioneering *in vivo* research ⁶ used stimulus-response experiments, where typically a perturbation is introduced to the extracellular space and the (intracellular) response of the system is captured by rapid sampling and quantitative metabolomics. The observations are then interpreted by generation of kinetic model(s) and parameter fitting.

These models use kinetic formalisms to describe the flux as a function of enzyme activity (e), substrate and effector concentrations (c_i) and kinetic parameters (θ) and are typically non-linear. The latter can be either derived from mechanistic assumptions e.g., Michaelis-Menten kinetics, or be non-mechanistic e.g., power law leading to parameters with no physical meaning ⁷.

$$v = e \cdot f(c_i, \theta) \quad (5.1)$$

There are mainly two inherent conceptual drawbacks of kinetic modeling:

- (1) The model equations (metabolite balances) are strongly dependent on each other, which means the system can hardly be solved from a decomposition and commonly also leads to highly correlated parameters ^{8,9}.
- (2) The kinetic functions have to be chosen *a priori*, which means the mechanism of every enzyme and all interactions between metabolites and enzymes involved in the regarded metabolic network have to be selected before

parameter optimization. Especially, the kinetic formats of each reaction in larger metabolic networks are yet unknown, unclear, or shall be deduced from the captured observables.

One solution that has been used to address this issue was to define a family of different models, also referred to as ensemble modeling¹⁰⁻¹³, where it is good practise to find the simplest model that can describe the data (Occam's razor). Also, strategies to speed up parameter estimation in kinetic models have been proposed using the estimation of fluxes from the metabolic network stoichiometry as an intermediate step and scaffold^{14,15}. This approach allows exploiting the linear properties of the stoichiometry matrix, *i.e.*, the nullspace, however, the methodology requires selecting a kinetic model *a priori* and cannot directly identify dynamic metabolic fluxes. Regardless of those challenges, the number of uncertain reaction mechanisms in a kinetic model leads to a rapidly increasing number of models to be compared, together with the challenges of parameter estimation in non-linear systems, which result in putative non-convex optimization landscapes that quickly reach computationally infeasible scales.

To circumvent these challenges, the proposed hybrid modeling approach^{16,17} strives for direct identification of the intracellular fluxes without a priori assumptions of kinetics (see Figure 5.1). The approach builds on piecewise affine (PWA) flux functions, which approximate the real fluxes throughout the metabolic network. This means only (more readily available) information on the metabolic network structure, including the atom transitions of each reaction, together with the directionality of the fluxes, is required a priori.

The approach could be regarded as an extension of DMFA framework¹⁸, which also builds on PWA flux functions but is only suitable for determined or redundant metabolic networks and for underdetermined systems by incorporation of ¹³C labelling¹⁹. Therewith the resulting parameter estimation problem becomes much more challenging, as the balances for the tracer atoms introduce non-linearity into the so far linear system, which moreover has to be solved by numerical integration instead of analytically. The metabolic tracers add a second set of observables, leading to a multi-objective parameter estimation problem that requires definition of a trade-off between the different types of observables. Moreover parameter correlations, and especially high dimensionality, need to be tackled. Classical non-linear optimization approaches were not found practical for the arising ill-posed inverse problems at a larger scale as they usually lead to very slow convergence and are often not robust enough for practical application in larger-scale metabolic networks.

Experimental Requirements for *in vivo* Dynamic Flux Estimation

Crucial for every modeling approach is sufficient information for the identification of the model parameters from the observables²⁰. The identifiability of *in vivo* flux functions strongly depends on the specific metabolic network, the experimental design and the captured observables. This has been extensively analyzed for steady-state²¹ and the same methodologies can be used in dynamic conditions. Although the experimental design is not the focus of this contribution, the main challenges and requirements are shortly discussed:

- A stimulus-response experiment should lead to perturbation(s) strong enough to cover a significant (metabolite) concentration space for good identification of the kinetic parameters²².
- The metabolomics should preferably have a complete coverage of the regarded intracellular metabolic response, e.g., intracellular concentrations, which moreover have to be quantitative.
- Use of a ¹³C labeled substrate for improved flux tracing, as the concentration information alone is mostly not sufficient to identify all intracellular fluxes^{18,23}.

These experimental requirements are partly conflicting, especially when concentration measurements are performed using ¹³C labeled internal standards²⁴, as they require non-labeled intracellular metabolite pools for quantification. The putative experimental setups therefore either require repetitive (cyclic) conditions or multiple experimental runs to capture the metabolite and ¹³C enrichment information.

In this work, we build on one dynamic setup that has been demonstrated to be advantageous in facilitating concentration and labeling measurements—there cyclic and block-wise feeding are used leading to so-called feast-famine conditions^{25,26}. Compared to other experimental stimulus-response setups, e.g., pulse experiments or step changes, the feast-famine conditions show the following characteristics:

- (1) They generate repetitive concentration patterns in time, allowing for dense sampling from multiple cycles as well as application of ¹³C labeling from a single experiment.
- (2) The feast/famine perturbation includes both: the transient from limitation to excess as well as a return to limitation in a short timeframe of minutes.
- (3) In this setup, the starting metabolite concentrations of each cycle are the same as the endpoint, which means there is no net metabolite accumulation during one cycle (material is washed out with the biomass).

These properties are beneficial for flux identification, as the concentration and enrichment information can be captured over several cycles, i.e., improving data density and accuracy.

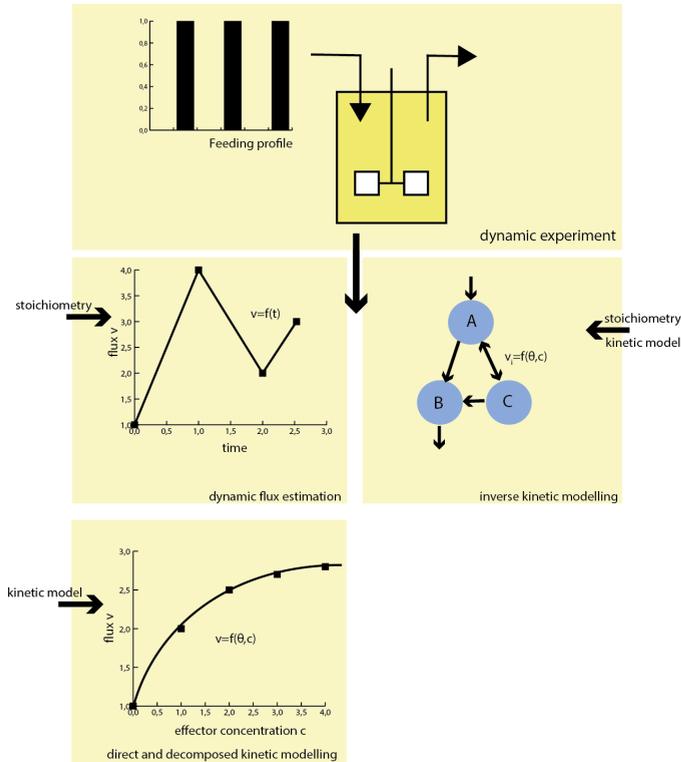


Figure 5.1. “Classical” kinetic modeling requires *a priori* defined kinetic mechanisms. Using the hybrid modeling approach, only the metabolic network structure (incl. atom transitions) is required *a priori*. As a result, the flux profile in time can be identified rather than kinetic parameters. With the flux functions on hand, the kinetics of the metabolic network can be investigated decoupled from the overall network.

Materials and Methods

The flux estimation follows a workflow, which will be described in detail in the following subchapters. First step is the domain selection procedure selecting the breakpoints for the flux functions. This is followed by definition and introduction of the non-negativity (prevents negative concentrations) and shape-prescriptive constraints. For use of the final step, first a feasible initial iterate has to be computed as implicit filtering requires this as an input. The final step performs the actual flux identification with incorporation of the enrichment data.

Used Models and Data

The *in silico* Spiral model was used as described previously¹⁶.

For the practical PenG example, concentration and labeling measurements as well as the metabolic network stoichiometry were obtained from a feast/famine cultivation of *P. chrysogenum* as previously reported¹⁷. In brief: The culture was supplied by a block-wise feeding regime (36 s feed, 324 s no feed) at an average dilution rate of $D = 0.05 \text{ h}^{-1}$. The feed contained minimal medium with a glucose concentration of 15 g/L. After several cycles, repetitive offgas and DO measurements were obtained, a biomass concentration of 5.7 g/L was obtained (average over the cycle). Samples for intracellular metabolites were withdrawn using a rapid-sampling device²⁷, quenched, extracted and analyzed using the ID-MS protocol^{17,28}. After sampling for intracellular concentrations, the feed was switched to a medium with the same composition, but containing fully ¹³C-labeled glucose. The enrichment was monitored during three consecutive cycles using rapid sampling and MS analysis. It has to be noted that the applicability of the approach is not limited to a particular organism but can be directly applied to any organism e.g. *S. cerevisiae*²⁹ under the provision that the necessary quantitative data can be captured and the pathway stoichiometry is known or can be inferred from this data.

Mathematical Modeling of Dynamic ¹³C Labeling Experiments Using PWA Flux Functions

The modeling approach is based on balances for metabolite concentrations and enrichments quantified by isotopomers (cumumers or other formalisms)^{23,30}. For isotopically non-stationary labeling the enrichment can also be simplified using C-molar enrichment (average enrichment of a metabolite over all C-atoms)¹⁷. The respective balances are generated based on the stoichiometry of each reaction and the respective C atom transitions. As the approach is well documented in literature (especially for steady-state), only an abstract description is given. The metabolite balance is based on the flux functions α and the metabolic network stoichiometry N .

$$\frac{dc}{dt} = f(N, \alpha) \quad (5.2)$$

Based on the atom transitions and the stoichiometry, a non-linear function g can be derived describing the balance of each single isotopomer concentration (or cumumer, EMU, *etc.*) x of the network:

$$\frac{d\mathbf{x}}{dt} = \mathbf{g}(\mathbf{c}, \boldsymbol{\alpha}, \mathbf{x}, \mathbf{x}^{inp}) \quad (5.3)$$

Integration of the isotopomer balances will yield the time-course of the isotopomer concentrations, which can be used to calculate the observables, i.e., mass-isotopomer fractions.

Flux Functions and Nomenclature in the PWA Flux Framework

Using piece-wise affine functions for the flux description changes some properties of the equation system. The piece-wise affine flux functions are defined in time by a set of breakpoints t_j valid for all fluxes in the considered metabolic network, partitioning the flux function v_i into number of breakpoints minus one domains in time:

$$v_{i,j}(t) = \begin{cases} v_{i,0} + \frac{v_{i,1} - v_{i,0}}{t_{i,1}} t & t_{j=0} \leq t \leq t_1 \\ v_{i,1} + \frac{v_{i,2} - v_{i,1}}{t_{i,2} - t_{i,1}} (t - t_1) & t_1 < t \leq t_2 \\ v_{i,2} + \frac{v_{i,3} - v_{i,2}}{t_{i,3} - t_{i,2}} (t - t_2) & t_2 < t \leq t_3 \end{cases} \quad (5.4)$$

The parameters of this function are the value of respective fluxes v_i at the breakpoint t_j and defined strictly positive (so a reversible reaction will have two fluxes). Flux values between breakpoints are calculated by linear interpolation on its two adjacent breakpoints. Therefore a metabolic network with i fluxes and j breakpoints has $i \times j$ parameters to be estimated. As will be discussed later, this number can decrease for specific experimental setups that introduce additional constraints on the concentrations or flux functions. This definition is also valid for higher order piecewise-defined flux functions as long as they are to be defined unique on the breakpoints, e.g., smooth quadratic splines, but will not be discussed in the scope of this paper.

Balancing of Metabolites and Solution of the Metabolite Mass Balances

With the previous definition of the flux functions in time, the metabolite mass balances can be formulated locally at every breakpoint j by linear combinations according to the stoichiometry matrix which can be further divided into known v_b and unknown v_n fluxes³¹.

$$\begin{aligned}
\mathbf{N} &= [\mathbf{N}_b \mathbf{N}_n] \\
\frac{d\mathbf{c}}{dt}(t_j) &= \mathbf{N}_b \mathbf{v}_b + \mathbf{N}_n \mathbf{v}_n \\
\mathbf{v}_n &= \mathbf{N}_n^{-1} \left(\frac{d\mathbf{c}}{dt}(t_j) - \mathbf{N}_b \mathbf{v}_b \right)
\end{aligned} \tag{5.5}$$

If \mathbf{N}_n is square and nondegenerate, the matrix can be inverted and a unique solution of unknown fluxes can be identified from the concentration transients and the known fluxes.

$$\mathbf{v}_n = \mathbf{N}_n^{-1} \left(\frac{d\mathbf{c}}{dt}(t_j) - \mathbf{N}_b \mathbf{v}_b \right) \tag{5.6}$$

In practice most metabolic networks are underdetermined with respect to the number of unknown fluxes f , i.e., they contain parallel pathways, cycles or reversible reactions.

$$\text{rank}(\mathbf{N}_n) < f \tag{5.7}$$

This means the metabolic network does not contain enough measurable (known) fluxes in order to identify all fluxes. In this case \mathbf{N}_n spans a null space (kernel), which contains all linear combinations of its basis vector(s) describing the infinite combinations of fluxes that lead to the same solution of $\frac{d\mathbf{c}}{dt}$.

$$\{ \mathbf{v}_n + \mathbf{w} \mid \mathbf{N}_n \mathbf{v}_n = \mathbf{const} \wedge \mathbf{w} \in \text{Null}(\mathbf{N}_n) \} \tag{5.8}$$

This is called practical non-identifiability and also directly explains why isotopic tracers are needed in order to identify all flux functions of such metabolic networks. The right hand side (rhs) is a linear combination of the flux functions \mathbf{v} , thus also results in a PWA function. Therewith the rhs can be described by a linear function u on the same breakpoints as \mathbf{v} :

$$\frac{d\mathbf{c}}{dt}(t) = u(t) \quad \text{with: } u(t) = \begin{cases} u_0 + \frac{u_1 - u_0}{t_1} t & t \leq t_1 \\ u_1 + \frac{u_2 - u_1}{t_2 - t_1} (t - t_1) & t_1 < t \leq t_2 \\ u_2 + \frac{u_3 - u_2}{t_3 - t_2} (t - t_2) & t_2 < t \leq t_3 \end{cases} \tag{5.9}$$

This function can easily be integrated (analytical), leading to a continuous, quadratic solution for the concentration in time with the initial concentrations \mathbf{c}_0 :

$$\mathbf{c}(\mathbf{u}, t) = \mathbf{c}_0 + \begin{cases} u_0 t + \frac{1}{2} \frac{u_1 - u_0}{t_1} t^2 & t \leq t_1 \\ u_0 t_1 + \frac{1}{2} (u_1 - u_0) t_1 + u_1 (t - t_1) + \frac{1}{2} \frac{u_2 - u_1}{t_2 - t_1} (t - t_1)^2 & t_1 < t \leq t_2 \\ u_0 t_1 + \frac{1}{2} (u_1 - u_0) t_1 + u_1 (t_2 - t_1) + \frac{1}{2} (u_2 - u_1) (t_2 - t_1) + (u_3 - u_2) (t - t_2) + \frac{1}{2} \frac{u_3 - u_2}{t_3 - t_2} (t - t_2)^2 & t_2 < t \leq t_3 \\ \vdots & \vdots \end{cases} \quad (5.10)$$

This function can easily be integrated (analytical), leading to a continuous, quadratic solution for the concentration in time with the initial concentrations \mathbf{c}_0 :

$$\mathbf{c}(\mathbf{u}, t) = \mathbf{c}_0 + \begin{cases} u_0 t + \frac{1}{2} \frac{u_1 - u_0}{t_1} t^2 & t \leq t_1 \\ u_0 t_1 + \frac{1}{2} (u_1 - u_0) t_1 + u_1 (t - t_1) + \frac{1}{2} \frac{u_2 - u_1}{t_2 - t_1} (t - t_1)^2 & t_1 < t \leq t_2 \\ u_0 t_1 + \frac{1}{2} (u_1 - u_0) t_1 + u_1 (t_2 - t_1) + \frac{1}{2} (u_2 - u_1) (t_2 - t_1) + (u_3 - u_2) (t - t_2) + \frac{1}{2} \frac{u_3 - u_2}{t_3 - t_2} (t - t_2)^2 & t_2 < t \leq t_3 \\ \vdots & \vdots \end{cases} \quad (5.11)$$

As can be seen, this solution is moreover linear with respect to $u_{i,j}$; with available concentration measurements, a linear regression can be performed to obtain the best estimate of \mathbf{u} . The measurements \mathbf{c}_M at timepoints \mathbf{t}_M can be obtained by generation of a matrix \mathbf{Y} :

$$\mathbf{c}_M = \underbrace{\begin{pmatrix} \mathbf{1}^T & \mathbf{Y}^* \end{pmatrix}}_{= \mathbf{Y}} \begin{pmatrix} \mathbf{c}_0 \\ \mathbf{u} \end{pmatrix} \quad (5.12)$$

The weighted linear regression problem (with weight matrix \mathbf{W}_M) then reads:

$$\begin{pmatrix} \mathbf{c}_0 \\ \mathbf{u} \end{pmatrix} = \left(\mathbf{Y}^T \mathbf{W}_M \mathbf{Y} \right)^{-1} \mathbf{Y}^T \mathbf{W}_M \mathbf{Y} \mathbf{c}_M \quad (5.13)$$

Solving the optimization problem

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha}} R_c \quad (5.14)$$

Sufficient Breakpoint Selection

For the flux functions the location of the breakpoints in time has to be selected as also derived previously¹⁸. Best practice would be to use all observables (concentration and enrichments) for this selection; however, considering the

complexity and dimensionality of the resulting parameter optimization problem, this is computationally not feasible on standard computer hardware. The optimization is implemented in a nested manner (also see Figure 5.2 top) with the inner loop solving equation (5.13) and the outer loop iterating on the breakpoints minimizing the weighted residual sum of squares on the concentration observables R_c using the PSwarm optimization algorithm³².

$$\hat{t}_j = \arg \min_{t_j} R_c \quad (5.15)$$

From R_c the adjusted R^2 as well as the AIC have been calculated^{33,34}.

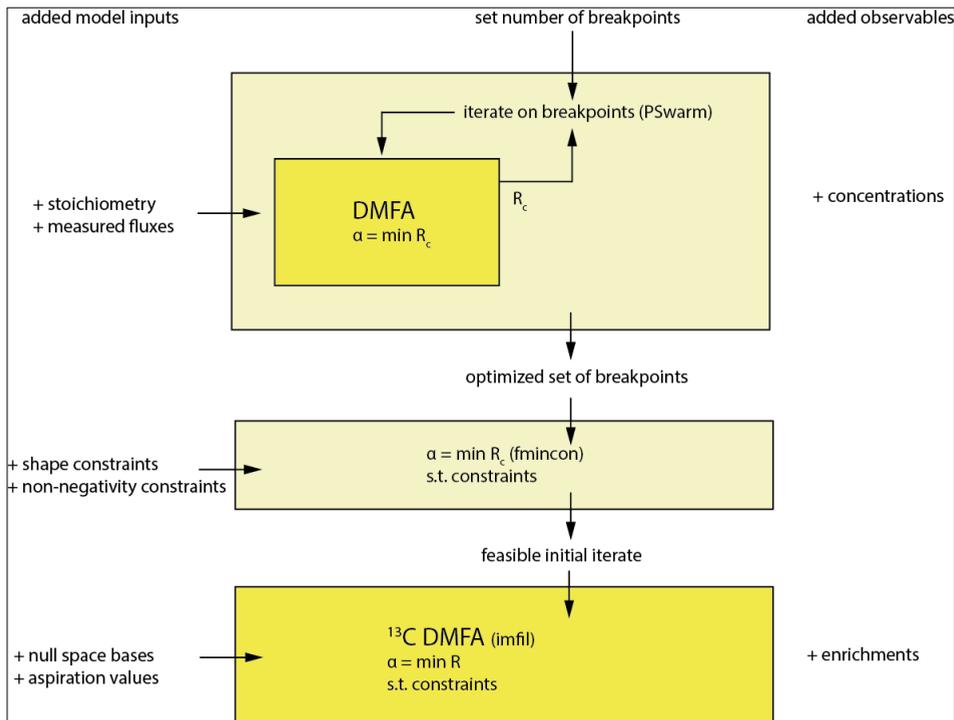


Figure 5.2. Overall workflow of the method with breakpoint selection as the first step followed by the main optimization and used optimizers in parentheses. The model inputs are successively expanded as described on the left columns. The first two steps only use the concentration observables whereas the final step additionally incorporates the enrichment observables from the ^{13}C tracer.

Introducing Constraints

In practical applications, measurement data with random errors (noise) can lead to issues during the optimization. Especially, cases that impair the stability of the numerical integration have to be prevented to ensure maximal robustness of the

high-dimensional parameter estimation: (1) Negative values in the approximation of the concentration profile; (2) Overfitting. Additionally, the experimental setup could enforce certain conditions that should be reflected in the parameter estimation, like in the case of the feast/famine setup, the first and last domain (beginning and end of the cycle) have to be equal.

Moreover the introduction of constraints can significantly reduce the search space and will therewith facilitate convergence of the optimization.

Specific Constraints for Feast-Famine Conditions

In the previous paragraph, the solution for the best estimate of the concentration profile with the given breakpoints has been demonstrated. In the special case of a feast/famine experiment the solution should additionally fulfill two (linear) constraints:

- (1) In a stable feast-famine regime the metabolite concentration at the end of one cycle has the same concentration as in the beginning (of the next cycle).
- (2) Similarly, the flux at the end of the feast famine cycle has to be the same as in the beginning. Otherwise no stable, repetitive cycles were obtained.

The concentration at the end of the cycle is the result of the integration using the parameters u (and c_0). A column vector C is generated, representing a column of matrix Y for the timepoint t_{end} :

$$C = Y_{t_{end}} \quad (5.16)$$

Integration of the constraint is obtained using Lagrange multipliers λ :

$$\begin{pmatrix} \hat{\mathbf{q}} \\ \lambda \end{pmatrix} = \mathbf{Z} \begin{pmatrix} \mathbf{Y}^T \mathbf{W}_M \mathbf{Y} \mathbf{c}_M - c_0 \\ \mathbf{b} \end{pmatrix} \quad \text{with } \mathbf{Z} = \begin{pmatrix} \mathbf{Y}^T \mathbf{W}_M \mathbf{Y} & \mathbf{C}^T \\ \mathbf{C} & \mathbf{0} \end{pmatrix}, \hat{\mathbf{q}} = \begin{pmatrix} \hat{\mathbf{u}} \\ \hat{c}_0 \end{pmatrix} \quad (5.17)$$

A comparable approach is taken to introduce the flux constraint $v_1 = v_j$. The vectors C and b are extended by an additional row:

$$C = (1 \quad 0 \quad \dots \quad -1), \quad \mathbf{b} = 0 \quad (5.18)$$

As can be seen this equality constraint reduces the number of parameters by one per balanced metabolite.

Non-Negativity Constraints

In order to ensure numerical stability of the integration for the isotopomer equations, it is crucial that computed negative approximated concentrations are detected and eliminated. In order to find a sufficient constraint the piecewise quadratic concentration pattern has to be evaluated in every domain.

In the convex case, the minimum of the vertex value has been computed, whereas in the concave case it has to be checked for roots in its interval. These computations are trivial but correspond to a nonlinear (quadratic) inequality constraint.

Rationale for Shape-Prescriptive Constraints

In practical applications overfitting has to be prevented. Reasons for overfitting can be (1) noisy data, (2) outlier or leverage points, (3) suboptimal choice of domains for flux functions, (4) suboptimal sample time points, (5) systematic error in observables or (6) missing data. Overfitting will lead to flux functions with potentially too high gradients. Additionally, because metabolite pools are often closely linked, the overfitting will propagate through the metabolic network, leading to an ill-posed optimization landscape hindering the parameter estimation progression and biased flux functions.

One obvious approach is to optimize the domain selection as discussed. Nevertheless, there are limits to the reduction of domains as all flux functions are defined on the same domains, and a trade-off between overfitting prevention and achievable convergence of the flux functions has to be found. It has to be noted that the proposed linear solution of the domain selection can only be applied to a subset of the observables (the concentration measurements), while non-linear differential equation systems are required to obtain the labeling profile. Consequently, the domains are chosen on a subset of the available data leading to a risk that the domain selection yields a good approximation but potentially not the global optimum for all data.

Shape constraints can be applied to “guide” the optimization procedure using shape primitives; together with the least squares objective function. This additional information on the shape of the (quadratic) concentration patterns has several benefits:

- (1) It can facilitate higher convergence orders of the flux functions;
- (2) The constraints restrict the search space and prevent that the optimization algorithm to converge in local optima in a noisy optimization landscape;
- (3) The chosen constraints do not introduce additional (unobserved) parameters;
- (4) The evaluation of the constraints is computationally very cheap even at large scale.

Shape constraints are (with exception of the non-negativity constraint) linear inequality equations, which can be implemented rapidly and intuitive in the PWA framework as linear inequality constraints on the parameter vector. This also means that feasible parameter sets can be computed almost instantly using linear programming. It has to be noted that shape constraints can also be introduced for metabolites where no observables are on hand (e.g., the metabolite is very unstable and cannot be quantified) which are exceptionally prone to overfitting as they do not contribute to the objective function. A strategy that has been used in the Penicillium model was to assume non-observed pools to follow the same dynamics as their surrounding metabolites.

Monotonicity and Convexity Constraints

When constraining the monotonicity in a domain four cases have to be considered for PWA flux functions: (1) Monotonous decreasing, (2) Monotonous increasing, (3) Switch from monotonous increasing to monotonous decreasing and (4) Switch from monotonous decreasing to monotonous increasing. In the case of a quadratic function this behavior can be enforced by constraining the right hand side (rhs) u at the breakpoints adjacent to the respective domain (see Figure 5.3. Inequality equations for the different shape constraints.).

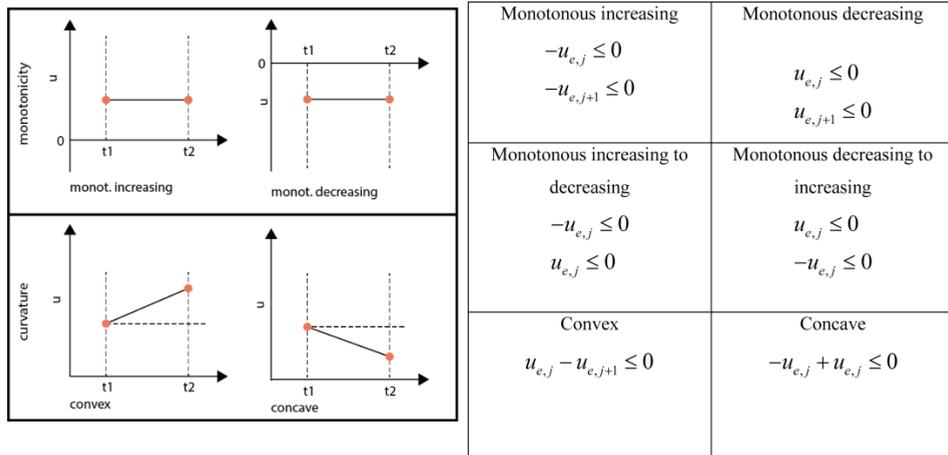


Figure 5.3. Inequality equations for the different shape constraints.

A direct switch from monotonous increasing to decreasing and the other way round is only possible if the rhs between those neighboring domains was zero, leading to a double-inequality constraint which is equivalent to an equality constraint, *i.e.*, forcing the inflection point to the respective middle knot. Generally, convexity and curvature can be constrained independently. Nevertheless,

in the case that a change of monotonicity is enforced, also the curvature is constrained. It is good practice to check the constraints for feasibility, e.g., by using a linear programming solver, e.g., linprog in MATLAB.

Equality Shape Constraints

In special cases a more stringent constraint than the inequality constraints could be needed. Here, the slope of the concentrations at the breakpoints can also directly be constrained using an equality constraint:

$$\mathbf{u}_{e,j} = \mathbf{A} \tag{5.19}$$

This constraint has been applied in the Penicillium model for metabolites with very noisy concentration measurements that did not allow for identifying a trend. The slopes have been set to zero, *i.e.*, the (net) influx equals the outflux of a pool at all times (details can be found in the Supplementary Material).

¹³C DMFA

The final flux (parameter) estimation problem has to minimize the error subject to two sets of objectives: (1) The concentration measurements and (2) The ¹³C enrichment observables. The weighted residual sum of squares (R_i) was chosen, whereas a constant α was introduced as scaling factor to weight the two R_i, which will lead to different points on the respective pareto frontier; it has been set to 1.317 to normalize for the different number of observations in either dataset. This leads to a multi-objective optimization problem and the L2 global criterion was used to define the overall objective function on previously computed aspiration values (best fit on either dataset) for the two residual sum of squares (R_c and R_x).

$$R_c = a \cdot \sum_m \left(\frac{y_{c,m} - y_{c,m}(\boldsymbol{\theta}, \mathbf{t})}{\boldsymbol{\sigma}_{c,m}} \right)^2, \quad R_x = \sum_m \left(\frac{y_{x,m} - y_{x,m}(\boldsymbol{\theta}, \mathbf{t})}{\boldsymbol{\sigma}_{x,m}} \right)^2$$

$$R = \left(\left(\frac{R_c - R_{c,asp}}{R_{c,asp}} \right)^2 + \left(\frac{R_x - R_{x,asp}}{R_{x,asp}} \right)^2 \right)$$

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} R$$

(5.20)

*s.t.**shape const.**nonneg. constr.*

Computing the aspiration values on first sight seems an unnecessary additional effort, but it has to be noted that the best fit of the concentration data is already available from the domain selection procedure and a good approximation of the enrichments can usually be computed fast, neglecting the concentration observables. In practice this has the advantage that the whole dataset can be inspected for outliers and leverage values and those observations can be removed or downweighted. This procedure will lead to a better posed optimization landscape and can speed up the overall parameter estimation process considerably. If no good estimate for the errors is available, those could also be estimated using, e.g., an iteratively reweighted least squares approach like robust regression. With the current scope to analyze the practical optimization convergence, a simplifying assumption of an error of 5% of the average value within each dataset was used.

The Implicit Filtering (imfil) Optimization Algorithm

The implicit filtering algorithm has been applied as provided (version for MATLAB)³⁵ with minor modifications and bug fixes. In brief, implicit filtering is a hybrid sampling algorithm that samples points on a so called *stencil* (with variable size in the parameter space called *scale* centered at the current best iterate) in the parameter space comparable to a pattern search or direct search algorithm. From the collected points in the iteration, projected gradients are computed, which are then used to perform a quasi-Newton iteration. In case no improvement was found using the quasi-Newton iteration, the solver will move the center of the stencil to the best sampled point. If no better point was sampled, the actual stencil scale is adjusted (decreased) or the optimization is terminated. In this work, the quasi-Newton iteration with the SR1 Hessian update was used.

Implementation of the Constraints in the Implicit Filtering Optimizer

imfil has to be initialized with a strictly feasible initial parameter set, which was computed using the MATLAB *fmincon* solver. In contrast to *fmincon*, using *imfil*, constraints have to be implemented as so-called hidden constraints, *i.e.*, a parameter set outside the feasible territory is rejected and a failure is returned to the algorithm (see Figure 5.4).

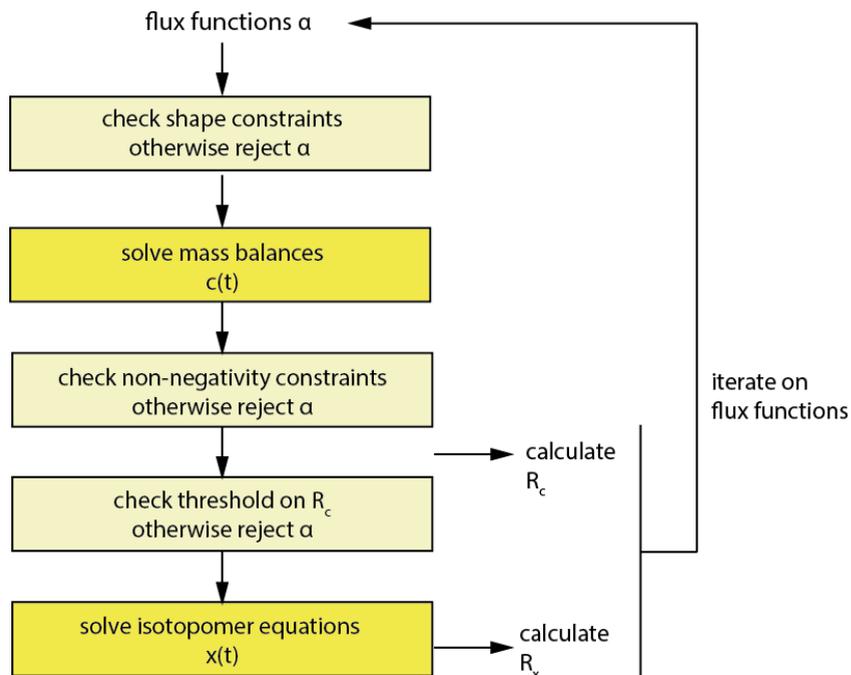


Figure 5.4. Workflow of the optimization in *imfil*.

This might at first seem overly simplistic, but considering that in the code 99.8% of the computational time in the forward simulation is being spent for the numerical integration of the isotopomer equations, rejecting parameter vectors without performing the numerical integration for non-feasible parameter sets can reduce the computational cost notably. Therefore all mentioned constraints have been introduced as hidden constraints when *imfil* was used. Beside the already discussed constraints, also a constraint on R_c has been implemented, and this was typically set to two times the initial R_c and prevents that a numerical integration of flux functions is performed for parameter vectors with a very bad objective function value on the concentrations. It is also useful in the initial phase of the optimization, when R_x dominates the objective function, in order to constrain the redistribution of error into R_c .

This might at first seem overly simplistic, but considering that in the code 99.8% of the computational time in the forward simulation is being spent for the numerical integration of the isotopomer equations, rejecting parameter vectors without performing the numerical integration for non-feasible parameter sets can reduce the computational cost notably. Therefore all mentioned constraints have been introduced as hidden constraints when *imfil* was used. Beside the already discussed constraints, also a constraint on R_c has been implemented, and this was typically set to two times the initial R_c and prevents that a numerical integration of flux functions is performed for parameter vectors with a very bad objective function value on the concentrations. It is also useful in the initial phase of the optimization, when R_x dominates the objective function, in order to constrain the redistribution of error into R_c .

Null-Space-Based Sampling

Hidden constraints can have the risk of compromising the identification of a decent gradient. Therefore, it is beneficial to enrich the stencil with additional directions that (1) are likely to lead to a better value in objective function and (2) are known to fulfill the given constraints together with the central finite differences. Feasible directions have been implemented in the algorithm using the *vstencil* property of the solver (see Figure 5.5).

A powerful set of such directions can be directly derived from Equation (7). Samples within this null space will have the same rhs u , and therewith the concentration profile as the current best iterate, but still a different solution of the isotopomer balances. Using this stoichiometry-based null space is further referred to as null-space-based sampling.

This approach allows iterating in the lower dimensional space of fluxes (parameters), which cannot be determined by the network stoichiometry and extracellular rates, but requiring ^{13}C labeling measurements to be identified. This parameter space reduction also eliminates the putative trade-off between concentration and enrichment observable residuals in the multi-objective function (i.e., the concentration residual remains equal). In order to find the best overall optimum on all observables the central finite differences for all parameters are sampled together with the null space basis vectors.

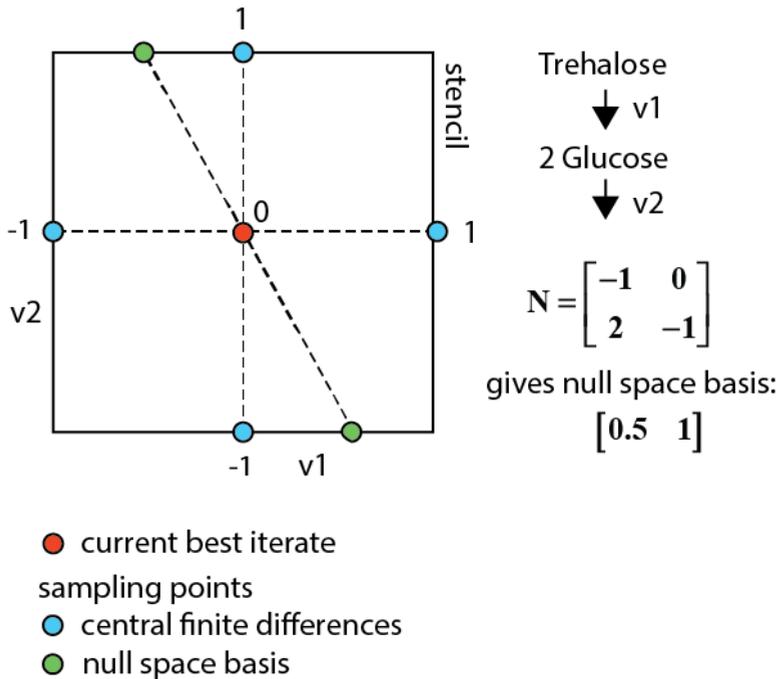


Figure 5.5. Sampling space for a network with two fluxes and one measured metabolite (glucose). When using central finite differences the blue points are sampled. Using the null space sampling only the two green points are sampled that fulfill the constraints and best reproduce the concentration measurements.

Results and Discussion

Computation of an Initial Set of Breakpoints Using the Concentration Measurements

In the piecewise affine flux estimation approach, the choice of the number and positions of the breakpoints is crucial for a successful approximation of the flux profile in time and they directly determine the achievable convergence of the flux approximation towards the real fluxes as well as the number of parameters to be identified from the observables.

The method requires that the breakpoints for all flux functions in the metabolic network have the same set of breakpoints. To best capture the different response of all the metabolites and fluxes a satisfying compromise of breakpoints has to be found. Changes in the (net) in- and out-fluxes of a metabolite pool influence the metabolite concentration and it is assumed that the major changes in metabolic (net)

fluxes are reflected if the flux approximation can reproduce the concentration profile well.

With this assumption, an optimization for the placement of breakpoints (number and times) can be performed using the less complex linear metabolite balances and the concentration observables leaving the ^{13}C labeling aside. If the reproduction of enrichment measurement is not satisfactory, additional domains can be introduced at any time in the optimization workflow. Moreover, a trade-off between the number of parameters to be estimated and the convergence that can be achieved with the PWA flux functions has to be found. Especially, a too high number of breakpoints can significantly increase the tendency of the system towards overfitting.

Using *in silico* data, the properties of the domain selection are studied. A small reaction network, further referred to as the spiral model, is simulated under feast-famine conditions and the approximation is compared to the noise-free measurements of the model with respect to:

- (1) The number of breakpoints and
- (2) The placement of breakpoints.

In the spiral network, six metabolite pools are observed (extracellular: A_{ex}, intracellular A, B, C, D, E) at 13 different time points during a feast/famine cycle of 140 s. The first and last breakpoints are fixed (0 and 140 s, the start and end of the feast-famine cycle). With only these breakpoints, no reasonable approximation can be obtained (see Figure 5.6) and the sum of squares is high (579.7, see Table 5.1). Introducing one additional breakpoint (best at $t = 126$ s) reduces the sum of squares about fourfold (134.1). Nevertheless, the approximation of the concentration profiles is visually not yet sufficient. To compare the goodness of fit we use the R^2 and adjusted with reference to the scenario without any free breakpoint (thus only the fixed points $t_0 = 0\text{s}$ and $t_{\text{end}} = 140\text{s}$). Further, the Akaike information criterion (AIC) is calculated and summarized in Table 5.1. All criteria equally suggest a not yet sufficient reproduction ($R^2 = 0.77$).

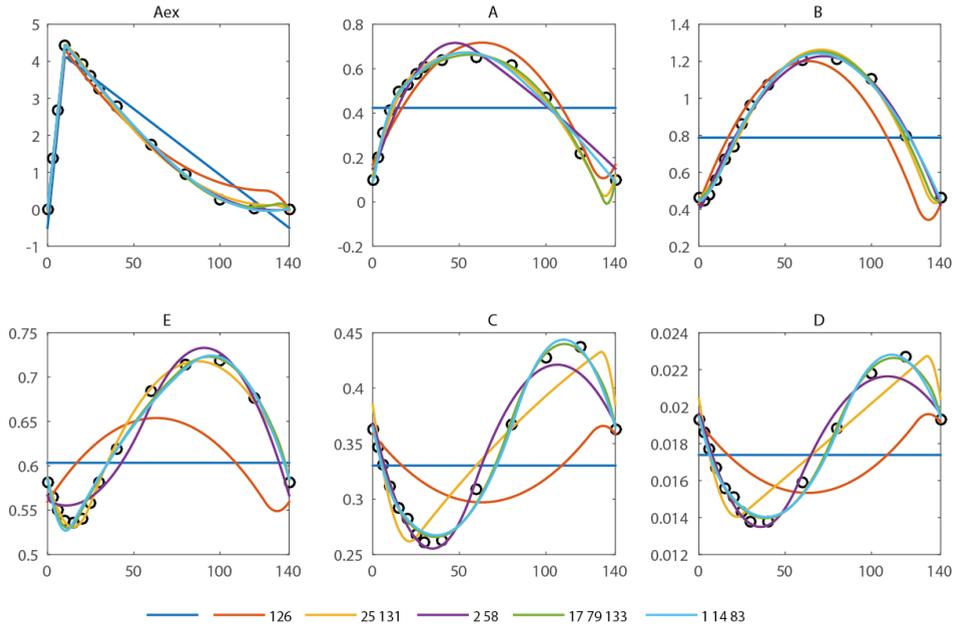


Figure 5.6. Approximation of the concentration measurements using piecewise affine derivatives with 0 (blue), 1 (orange), 2 (yellow and magenta) and 3 (green and light blue) free breakpoints. The respective residual sum of squared errors can be found in Table 5.1.

Table 5.1. Comparison of goodness of fit for different number of breakpoints and placement. In all cases 98 observations (measurements) are present, the number of parameters corresponds to $n-p$ (n = number of fluxes, p = number of breakpoints).

Break Points (s)	#p	RSS	R^2	\bar{R}^2	AIC
-	8	579.7	-	-	91.7
126	16	134.6	0.77	0.72	45.5
25, 131	24	15.1	0.97	0.97	-31.6
2, 58	24	22.2	0.96	0.95	-15.2
17, 79, 133	32	2.3	1.00	0.99	-95.7
1, 14, 83	32	4.7	0.99	0.99	-65.5

With a second free breakpoint (best choice $t_2 = 25$ s, $t_3 = 131$ s), a sum of squares of 15.1 ($R^2 = 0.97$) is reached and the concentration measurements can be reproduced within the expected error range of 10%. Addition of another, third free breakpoint further reduces the RSS to 2.3 ($R^2 = 0.99$). Next to these global optima, local minima are observed.

The territory of approximation error (RSS) also exhibits local minima. The second best combination of breakpoints was obtained with $t_2 = 2s$ and $t_3 = 58s$ (RSS 22.2). For three breakpoints, two clusters with low residual are obtained. The global minimum (RSS = 2.3, $R^2 = 0.996$) is found for a combination with $t = (17s, 79s, 133s)$ whereas the second optimum has a RSS of 4.7 at $t_2 = 1s$, $t_3 = 14s$, $t_4 = 83s$. In all cases 98 observations (measurements) are present, the number of parameters $\#p$ corresponds to $n-p$ (n = number of fluxes, p = number of breakpoints). These local minima differ in sum of squares about two-fold; nevertheless, based on expected experimental noise both should be taken into account and compared when labeling data is incorporated (isotopomer simulation).

Clearly, the breakpoint optimization will lead to an optimal sequence of breakpoints for the flux functions in terms of the chosen objective function R_c , but likely never the optimal one, subject to the complete set of observables. Still, the domain selection approach often gives a good sequence of breakpoints in practice. As seen, the implementation of the domain selection in the flux estimation leads to a highly non-convex optimization landscape, and this makes incorporation of the domain selection into the main optimization (using the enrichment information) laborious and computationally hardly feasible on standard computer hardware. From a practical standpoint, it is desirable to start with a simple model, as extension of a model is usually more straightforward than a model reduction, moreover, a smaller number of parameters can often be better identified from the observables²⁰.

Here the use of objective functions penalizing the number of parameters, like adjusted R^2 or AIC, can help in finding a minimal set of breakpoints. However, also using those criteria does not necessarily guarantee a sufficient representation of the concentrations, i.e., occurrence of overfitting and no negative concentrations. Having in mind those practical limitations of the domain selection, we implemented constraints in the approach that allow computing flux functions, even with a suboptimal choice of breakpoints and shape-prescriptive constraints.

Introduction of Shape-Prescriptive Constraints

As an example the sequence of knots $t = (10 s, 79 s, 133 s)$ is discussed. Two major flaws can be observed (Figure 5.7): (1) Negative concentrations have been computed for metabolites Aex and A, and (2) Metabolite A shows overfitting in the last domain. Negative concentrations are undoubtedly not feasible and cannot be accepted; moreover they compromise the numerical integration of the isotopomer balances being detrimental to the numerical robustness of the forward simulation.

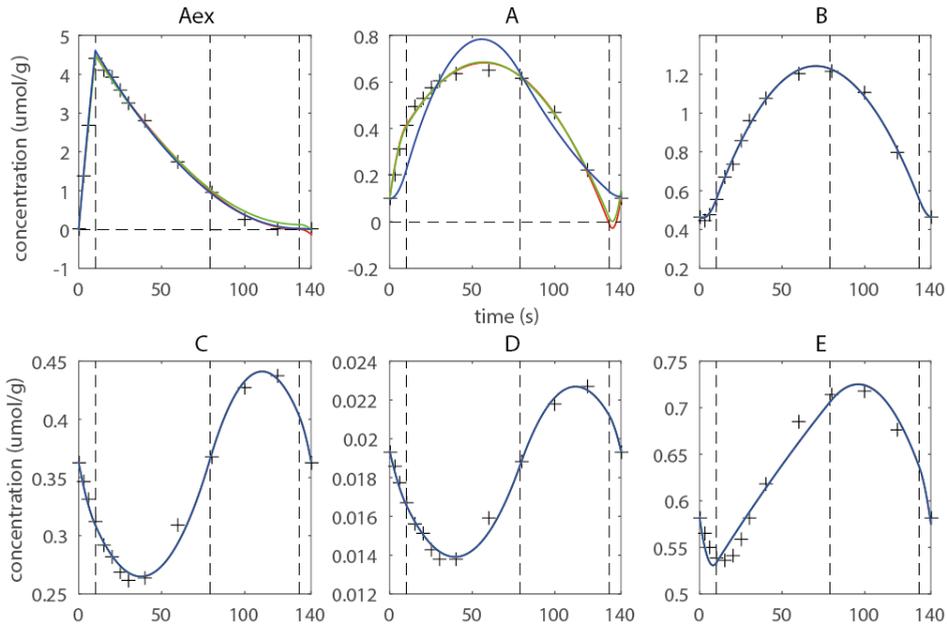


Figure 5.7. Estimated concentration profile using three free breakpoints $t = [0s\ 10s\ 79s\ 133s\ 140s]$ (red), non-negativity constraints (green) and shape constraints for metabolite A (blue), the corresponding R_c are 4.06 (red) 4.60 (green) and 72.79 (blue).

To enforce non-negativity quadratic inequality, constraints are introduced. This measure prevented all metabolites to reach negative concentrations (Figure 5.3, green approximation), but does not improve the overfitting behavior observed for metabolite A. Therefore, a shape constraint was introduced enforcing the concentration in the last domain to be monotonous decreasing. The constraint eliminates the overfitting, but leads to higher deviations from the observables in the first and second domain, whereas the fit on A_{ex} is also improved in the last domain with respect to overfitting. Thus, the example illustrates that the introduction of additional constraints is essential as it ensures a feasible concentration profile for the subsequent ^{13}C -based estimation of fluxes. Moreover shape constraints are a very efficient measure to prevent overfitting, and this can also help to smoothen noise in the optimization landscape which leads to a better-posed optimization problem. Moreover they can help to achieve a higher convergence of the flux functions towards the real fluxes, as overfitting in the concentration space can be efficiently eliminated. Considering that the flux estimation problem is usually high-dimensional, the constraints also reduce the parameter space that has to be searched, which can lead to an increase in convergence speed. This example was kept simple for demonstration, an extensively constrained network using

experimental data can be found in the next subchapter (also refer to Supporting Material for more details on the introduced constraints).

Estimating Flux Functions Using the Implicit Filtering Algorithm

To demonstrate the practical performance of the optimization approach, a previously described dynamic labeling experiment with *P. chrysogenum* was re-evaluated. The metabolic network for flux identification, further referred to as PenG model, comprises 17 balanced metabolites and 28 fluxes.

Five free breakpoints were determined, leading to a total of seven breakpoints, i.e., $t = (0 \text{ s } 18 \text{ s } 36 \text{ s } 90 \text{ s } 185 \text{ s } 230.5 \text{ s } 360 \text{ s})$. Because of the dependency of the last and first domain (feast/famine setup),

$28 \times 6 = 168$ flux values need to be estimated.

The simulations have been performed over three consecutive feast/famine cycles. As a starting point for the parameter optimization, a feasible parameter set derived from the best fit for the metabolite concentration measurements (including the respective constraints) was used.

We first applied a damped Quasi Newton optimization algorithm as reference convergence performance (see Figure 5.8). We found a very slow convergence which could originate from (1) stringent hidden constraints or (2) parameter correlations. The slow performance was likely not originating from putative over-stringent constraints, as the speed was even reduced when no constraints were applied.

The slow convergence suggests that the PWA optimization problem itself is ill-posed and gradient-based solvers underperform. One reason for the failure could be high correlation of the parameters, which can be taken as an advantage when the correlation is identified and used to determine search directions as implemented in the null-space-based sampling approach.

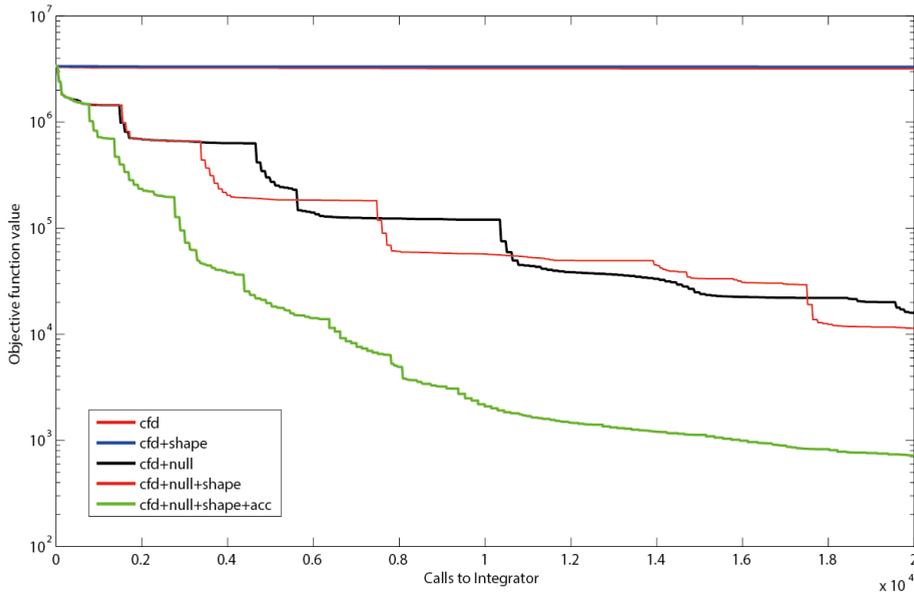


Figure 5.8. Objective function as a function of executed integrations. Using only central finite differences (cfd), resembling the search directions of a quasi-Newton solver or cfd together with shape constraints (shape) leads to a slow decrease in the residual sum of squares. The use of null-space-based sampling (null) significantly increases the convergence speed. Combining the null-space-based sampling with shape constraints leads to a slight improvement. Even better results are obtained when thresholds for improvement within the current stencil are introduced. Using a threshold of minimally 1% improvement in objective function (acc) to keep the stencil reduces inefficient iterations (green line).

Null-space-based sampling indeed significantly increases the convergence and shape constraints even increase the speed further. This increase supports the hypothesis that the shape constraints in the example were not overly stringent nor compromised the optimization landscape for the optimizer, although they have been implemented as hidden constraints in the implicit filtering algorithm. This means for the practical application that all PWA domains can be constrained without loss in performance as long as the constraints are not overly stringent.

It can be observed that the implicit filtering algorithm leads to a step-wise convergence (see Figure 5.8).

This behavior can be explained by the stencil rescaling when no parameter set leading to a better value in the objective function is sampled on a particular stencil. Step descents are observed right after a shrinking of the stencil. After a step decent a plateau is reached as the stencil is sampled, until no decent direction can be identified anymore, i.e., the gradients become very small. This inefficiency could

be reduced by the implementation of a threshold for the improvement between subsequent iterations. Indeed, this measure (further referred to as accelerator; acc) improves the convergence significantly, especially for the first couple of iterations. However, it has to be noted that the thresholds can increase the risk of obtaining a local optimum. The strategy that has been implemented to prevent such behavior is to restart the complete sequence of stencil scales after failure of a stencil.

Conclusions

Dynamic flux identification requires advanced experimental and computational approaches, especially experimental setups that allow for the simultaneous, respective serial measurement of intracellular concentrations and labeling enrichments. For the evaluation and interpretation of this experimental data, modeling and parameter estimation are crucial and require attention to obtain a good approximation of metabolic network fluxes.

Metabolic networks show an inherent correlation of (flux) parameters and common optimization approaches fail to identify the proper search direction. To solve this issue, we applied PWA flux functions and a two-staged optimization that resulted to be feasible also for larger networks. Especially, the implementation of implicit filtering-based optimization algorithms increased the convergence speed and basically eliminates manual intervention by frequent re-initialization of the optimization.

The proposed implementation allowed for the obtaining of a reasonable estimate for 168 parameters within 20,000 function evaluations. Key for fast convergence were (1) implementation of null space sampling and (2) shape constraints that prevent overfitting and reduce the search space.

Once the flux functions in time have been identified, these together with the respective effector concentration measurements allow for testing of different biological hypothesis on the regulatory mechanism. In contrast to the classical reaction kinetic approach, the identification of kinetic parameters can now be decoupled from the overall metabolic network, i.e., each approximated flux profile can be reproduced and tested using different kinetics formats for a small network as previously described¹⁶.

Supplementary Materials

Supplementary materials can be accessed at: <http://www.mdpi.com/2218-1989/5/4/697/s1>

Symbols and Abbreviations

Subscript

0	Initial condition
asp	Aspiration value
e	Metabolite identifier
end	End of feast famine cycle
i	index of flux
j	index of breakpoint
k	index of domain
m	Index of measurement
M	All measurements

Superscript

inp	Enrichment of used substrate for labelling
b	Known flux
n	Unknown flux
f	Number of unknown fluxes

General

a	Scaling factor
e	Enzyme activity
R, RSS	(weighted) residual sum of squares
R^2	Coefficient of determination
θ	Flux function parameters
α	Kinetic parameters
c	(metabolite) concentration
x	(C-molar) enrichment
v	Flux
u	Right-hand-side (dc/dt)
W	Weight matrix
$\hat{}$	Optimal estimate
t	Time
λ	Lagrange multiplier
w	Translation vector of the null space
N	Stoichiometry matrix
σ	Standard deviation
imfil	Implicit filtering algorithm
PWA	Piecewise affine
acc	Accelerator for imfil

DMFA Dynamic metabolic flux analysis
AIC Aikaie information criterion

References

- 1 Postma, P. W., Lengeler, J. W. & Jacobson, G. R. Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiol Rev* 57, 543-594 (1993).
- 2 Lapin, A., Klann, M. & Reuss, M. in *Biosystems Engineering II* Vol. 121 *Advances in Biochemical Engineering / Biotechnology* (eds Christoph Wittmann & Rainer Krull) Ch. 53, 23-43 (Springer Berlin Heidelberg, 2010).
- 3 Herring, C. D. *et al.* Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. *Nat Genet* 38, 1406-1412 (2006).
- 4 Kresnowati, M. T. *et al.* When transcriptome meets metabolome: fast cellular responses of yeast to sudden relief of glucose limitation. *Molecular Systems Biology* 2, 49 (2006).
- 5 Teusink, B. *et al.* Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. *European journal of biochemistry / FEBS* 267, 5313-5329 (2000).
- 6 Theobald, U., Balthes, M., Rizzi, M. & Reuss, M. Structured metabolic modelling applied to dynamic simulation of the Crabtree-and Pasteur-effect in baker's yeast. *Biochemical Engineering-Stuttgart*, 361-364 (1991).
- 7 Heijnen, J. J. Approximative kinetic formats used in metabolic network modeling. *Biotechnology and Bioengineering* 91, 534-545 (2005).
- 8 Jia, G., Stephanopoulos, G. & Gunawan, R. Incremental parameter estimation of kinetic metabolic network models. *BMC Systems Biology* 6, 142 (2012).
- 9 Chou, I.-C., Martens, H. & Voit, E. Parameter estimation in biochemical systems models with alternating regression. *Theoretical Biology and Medical Modelling* 3, 25 (2006).
- 10 Wahl, S. A., Haunschild, M. D., Oldiges, M. & Wiechert, W. Unravelling the regulatory structure of biochemical networks using stimulus response experiments and large-scale model selection. *IEE Proceedings - Systems Biology* 153, 275-285 (2006). <http://digital-library.theiet.org/content/journals/10.1049/ip-syb_20050089>.
- 11 Tran, L. M., Rizk, M. L. & Liao, J. C. Ensemble Modeling of Metabolic Networks. *Biophysical Journal* 95, 5606-5617 (2008).
- 12 Link, H., Kochanowski, K. & Sauer, U. Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo. *Nat Biotech* 31, 357-361 (2013).
- 13 Jia, G., Stephanopoulos, G. & Gunawan, R. Ensemble Kinetic Modeling of Metabolic Networks from Dynamic Metabolic Profiles. *Metabolites* 2, 891-912 (2012).
- 14 Goel, G., Chou, I. C. & Voit, E. O. System estimation from metabolic time-series data. *Bioinformatics* 24, 2505-2511 (2008).

- 15 Liu, Y. & Gunawan, R. Parameter estimation of dynamic biological network
models using integrated fluxes. *BMC Syst Biol* 8, 127 (2014).
- 16 Abate, A., Hillen, R. C. & Aljoscha Wahl, S. Piecewise affine approximations of
fluxes and enzyme kinetics from in vivo ¹³C labeling experiments.
International Journal of Robust and Nonlinear Control 22, 1120-1139
(2012).
- 17 de Jonge, L. *et al.* Flux response of glycolysis and storage metabolism during
rapid feast/famine conditions in *Penicillium chrysogenum* using dynamic ¹³C
labeling. *Biotechnology Journal* 9, 372-385 (2014).
- 18 Leighty, R. W. & Antoniewicz, M. R. Dynamic metabolic flux analysis (DMFA):
a framework for determining fluxes at metabolic non-steady state. *Metab
Eng* 13, 745-755 (2011).
- 19 Bonarius, H. P. J., Schmid, G. & Tramper, J. Flux analysis of underdetermined
metabolic networks: the quest for the missing constraints. *Trends in
Biotechnology* 15, 308-314 (1997).
- 20 Raue, A. *et al.* Structural and practical identifiability analysis of partially
observed dynamical models by exploiting the profile likelihood.
Bioinformatics 25, 1923-1929 (2009).
- 21 Isermann, N. & Wiechert, W. Metabolic isotopomer labeling systems. Part II:
structural flux identifiability analysis. *Mathematical biosciences* 183, 175-
214 (2003).
- 22 Degenring, D., Froemel, C., Dikta, G. & Takors, R. Sensitivity analysis for the
reduction of complex metabolism models. *Journal of Process Control* 14,
729-745 (2004).
- 23 Wiechert, W., Möllney, M., Petersen, S. & de Graaf, A. A. A Universal
Framework for ¹³C Metabolic Flux Analysis. *Metab Eng* 3, 265-283 (2001).
- 24 Wu, L. *et al.* Quantitative analysis of the microbial metabolome by isotope
dilution mass spectrometry using uniformly ¹³C-labeled cell extracts as
internal standards. *Anal Biochem* 336, 164-171 (2005).
- 25 Van Kleeff, B. H. A., Kuenen, J. G. & Heijnen, J. J. Heat flux measurements
for the fast monitoring of dynamic responses to glucose additions by yeasts
that were subjected to different feeding regimes in continuous culture.
Biotechnology Progress 12, 510-518 (1996).
- 26 Suarez-Mendez, C., Sousa, A., Heijnen, J. & Wahl, A. Fast "Feast/Famine"
Cycles for Studying Microbial Physiology Under Dynamic Conditions: A Case
Study with *Saccharomyces cerevisiae*. *Metabolites* 4, 347 (2014).
- 27 Lange, H. C. *et al.* Improved rapid sampling for in vivo kinetics of
intracellular metabolites in *Saccharomyces cerevisiae*. *Biotechnology and
Bioengineering* 75, 406-415 (2001).
- 28 van Heerden, J. H. *et al.* Lost in Transition: Startup of Glycolysis Yields
Subpopulations of Nongrowing Cells. *Science* (2014).

- 29 Suarez-Mendez, C. A., Ras, C. & Wahl, S. A. Metabolic adjustment upon repetitive substrate perturbations using dynamic ^{13}C -tracing in yeast. *Microbial Cell Factories* 16, 161 (2017).
- 30 Antoniewicz, M. R., Kelleher, J. K. & Stephanopoulos, G. Elementary metabolite units (EMU): A novel framework for modeling isotopic distributions. *Metab Eng* 9, 68-86 (2007).
- 31 Klamt, S., Schuster, S. & Gilles, E. D. Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism in purple nonsulfur bacteria. *Biotechnol Bioeng* 77, 734-751 (2002).
- 32 Vaz, A. I. F. & Vicente, L. N. A particle swarm pattern search method for bound constrained global optimization. *J Glob Optim* 39, 197-219 (2007).
- 33 Akaike, H. in *Selected Papers of Hirotugu Akaike Springer Series in Statistics* (eds Emanuel Parzen, Kunio Tanabe, & Genshiro Kitagawa) Ch. 16, 215-222 (Springer New York, 1998).
- 34 Montgomery, D. C. & Runger, G. C. *Applied Statistics and Probability for Engineers*. (John Wiley & Sons, 2010).
- 35 Kelley, C. T. *Implicit filtering*. (Society for Industrial and Applied Mathematics, 2011).

Chapter 6: General Conclusions and Outlook

The three most important characteristics for industrial fermentations are typically summarized as 'TRY': Titer, Rate and Yield.

Titer is an important size determining the cost for extraction of the product from the (aqueous) fermentation broth and also determines the necessary scale of the production equipment to produce an amount of product in batch fermentation. Rate translates to the (volumetric) productivity that can be achieved and therewith determines the size of the equipment for fermentation and downstream processing and is in consequence an important criterion to determine the capital cost of a process. Yield refers to the amount of carbon that is ending up from the educt in the commercial product to be produced and typically determines the production cost of the final product, especially when producing low value commodity chemicals.

All of these characteristics are governed by thermodynamics. Maximizing yield, i.e. high free energy conservation is counteracting rate and titer, which require high thermodynamic driving forces. The metabolic network that determines a specific stoichiometric yield influences the obtainable yields and titers. Therefore, network fluxes inside a microbial cell factory are not independent of each other, leading to complex metabolic trade-offs.

For that reason, it is crucial to explore and understand the governing principles of free energy conservation and how these determine the resulting phenotypes in order to make ideal compromises for strain development and also the design of microbial cell factories.

To demonstrate this theoretical concept in practice, in Chapter 2 a model system (low pH lactic acid fermentation) was evaluated with respect to the attainable product yield on substrate, where the versatility of stoichiometric modelling coupled to kinetic equations was demonstrated. This methodology is not limited to the presented case and can be directly applied to the production of other components that lead to uncoupling e.g. succinic acid that is a bio-based precursor for 1,4 butanediol. The model can describe the relation between operating conditions and cellular physiology globally to obtain the best process conditions and identify metabolic engineering targets on a rational basis. One additional advantage of stoichiometric modelling is that it gives a rich output of information about cellular physiology, without requiring a great deal of experimental observables for parametrization. This also means that such models can be applied early in the development process. At the same time they can be used to reconcile experimental

data and serve as a basis to derive new hypothesis and is therefore an underutilized tool in (academic) research.

Whereas targets for increased free energy conservation can be identified rationally and improved phenotypes can be simulated readily *in silico*, this does mean that they can also be engineered rationally in practice. In cases where implementation via genetic engineering fails, directed evolution could be applied. Directed evolution relies on creating an environment that is selective for the desired phenotype and therewith leads to an enrichment of the phenotype allowing for its subsequent isolation and analysis (Chapter 3). The challenges involved in directed evolution of phenotypes with increased free energy conservation are mainly two; (1) finding the appropriate experimental setup to enrich the desired phenotypes and (2) understanding the interrelation between the applied environmental conditions and the favored phenotypes from an intrinsic level.

Both challenges were addressed, where first an existing approach using incubation of cells in droplet was adapted for *S. cerevisiae* and validated for a system with known difference in free energy conservation. A detailed mathematical description was developed to derive the performance and limitations of the approach. In future the droplet cultivation system can be employed to isolate phenotypes with increased free energy conservation from microbial populations within the now known limits of the approach. Genotypic changes can be reverse engineered, characterized and incorporated into production strains in order to increase the anabolic product yield on substrate. One interesting experiment to carry out with the system may also be the characterization of a single yeast deletion library¹, where the relative enrichments could be assessed and compared using deep barcode sequencing². This could potentially lead to new discoveries on connections between metabolic stoichiometry and biomass (number) yield and consequently new targets for increased free energy conservation.

Furthermore, the approach can be used for transporter engineering (e.g. maltose-proton uniporter to maltose uniporter, see Chapter one) to screen larger mutant libraries. The created transporter can be used to realize an increase in free energy conservation compared to glucose and would at the same time create knowledge that may be used to improve the toolbox for rational engineering of membrane transporters in the future.

Secondly it was attempted to describe the cellular physiology from 'inside-out' by starting with the constraints posed by the cellular morphology on the enzyme activity that can be expressed in the cell applying constraint network modelling and kinetic modelling (Chapter 4). The model computes optimal protein compositions in

the cell to reach the maximal growth rate for a selected condition. With these simulations on hand selective conditions for directed evolution can be derived. Overall, it was found that the model can predict a wide range of metabolic phenomena like overflow metabolism and catabolite repression making it useful to devise experimental strategies for strain improvement. As the model predicts optimal enzyme make ups with respect to the objective function it may also be interesting to compare the predictions to quantitative proteomic data to derive new hypotheses about metabolic functions from the 'non-idealities' of cellular proteome allocation under the respective conditions. This is a very promising approach with respect to the increasing availability of quantitative omics data, especially fluxome and proteome data that can be used to refine and extend the parametrization of such models. However it also becomes apparent from the work that there is a lack of basic quantitative physiological data for different microorganisms, a path of research that has been largely abandoned in the post-genomic era and may in future pose limitations on the development of useful models.

Even for the described steady-state conditions, kinetic information is required. To improve the quality of *in vivo* kinetics, a new methodology was presented allowing the reconstruction of dynamic fluxes in metabolic instationary conditions together with a strategy to solve the arising large parameter identification problem. In comparison to hitherto existing approach the incorporation of ^{13}C tracing also allows for the identification of fluxes in underdetermined metabolic networks and therewith lays the foundation for the resolution of energy consuming futile cycles within the cells. This is in particular relevant with respect to the inevitable dynamics that cells are exposed to in large-scale bioreactors. The research in this field is still at its beginning, but is likely to turn more important especially with the demand for robust and efficient host organisms for cost-effective large-scale fermentation processes from renewable feedstock. So far this approach has not received a lot of attention, which may be due to the high analytical requirements but also due to the fact that the development of metabolic models describing the metabolic state under large-scale conditions are still at the beginning³. However, with increasing availability of computational power the interest to develop integrated models, in particular to link them with CFD techniques⁴ is increasing⁵ and will certainly require tools that can resolve transient fluxes in metabolic networks.

References

- 1 Pierce, S. E. *et al.* A unique and universal molecular barcode array. *Nature methods* 3, 601 (2006).
- 2 Smith, A. M. *et al.* Quantitative phenotyping via deep barcode sequencing. *Genome research*, gr. 093955.093109 (2009).
- 3 Wang, G. *et al.* Comparative performance of different scale-down simulators of substrate gradients in *Penicillium chrysogenum* cultures: the need of a biological systems response analysis. *Microbial Biotechnology* 11, 486-497 (2018).
- 4 Wang, G. *et al.* Integration of microbial kinetics and fluid dynamics toward model-driven scale-up of industrial bioprocesses. *Engineering in Life Sciences* 15, 20-29 (2015).
- 5 Haringa, C. *et al.* Computational fluid dynamics simulation of an industrial *P. chrysogenum* fermentation with a coupled 9-pool metabolic model: Towards rational scale-down and design optimization. *Chemical Engineering Science* 175, 12-24 (2018).

Curriculum vitae

Robin Schumacher was born on January 1st 1988 in Albstadt, Germany. He finished his secondary education at the humanistic *Hohenzollern Gymnasium Sigmaringen* with the *Abitur* in 2006 and started his diploma degree in Bioengineering at the *Technische Universitaet Carolo-Wilhelmina zu Braunschweig* in Germany in the same year. During his studies Robin was selected to pursue Erasmus studies abroad at the *Eidgenoessische Technische Hochschule Zuerich* where he pursued studies in the fields of process engineering and chemical engineering. Moreover Robin joined the Morbidelli group at ETH to perform a research project with the title: "Effect of pH and Sparging on Growth, Energy Metabolism and Productivity of Monoclonal Antibodies in Mammalian Cell Culture".

After leaving ETH, he performed an internship in pharmaceutical manufacturing at Pfizer Germany in Freiburg im Breisgau, before he joined *Technische Universiteit Delft*. There he wrote his diploma thesis abroad with the title: „Energy and storage metabolism in *Saccharomyces cerevisiae* under dynamic feast-famine conditions" with which he graduated from *Technische Universitaet Carolo-Wilhelmina zu Braunschweig* in June 2013.

In February 2013 he joined the Bioprocess Technology group (currently Cell Systems Engineering group) at *Technische Universiteit Delft* for his PhD training under supervision of Dr. S.A. Wahl and Prof. Dr. ir. J.J. Heijnen, where he worked on different aspects related to increased free energy conservation as presented in this thesis. In February 2018 he joined the Agency for Science, Technology and Research in Singapore where he currently holds a position as Research Fellow at the Biotransformation Innovation platform (BioTrans).

List of publications

Schumacher, R.; Wahl, S.A. Effective Estimation of Dynamic Metabolic Fluxes Using ^{13}C Labeling and Piecewise Affine Approximation: From Theory to Practical Applicability. *Metabolites* 2015, 5, 697-719.

The remaining chapters of this thesis will be submitted for publication shortly

Acknowledgements

This thesis is the result of the inspiration, guidance and support of many people, therefore I want to express my gratitude to everyone who has been with me on the incredible thin line between failure and success; to everyone that shared joy and sorrows with me; to all the memorable moments we were able to share in the last years. To all the Delftian Prometheuses that rekindled the spark and kept the flame burning within me.

First of all I would like to thank my promotor and copromotor, Sef Heijnen and Aljoscha Wahl for accompanying this scientific journey. I want to thank you for sharing your knowledge, expertise and experience. I only realized now after leaving TU Delft how deeply this time has shaped my thinking. I am glad to have had you as my mentors and am grateful to have had the honor to learn from the best in the field. I will happily carry on your values.

I want to thank the committee members for their time and effort in evaluating this work and all other people who contributed directly or indirectly to it.

I also want to thank my parents and siblings for the unconditional love and support. Melanie I want to thank you for your love and support.

Leonor and Simon, I want to thank you for being my paranymphs and all the effort you put in the organization of the defense, moreover I want to thank you for your friendship, support and guidance.

The office mate cluster

Jinrui Zhang

David Mendez

Katelijne Bekers

Eleni Vasilakou

Shima Saffarian

Thank you very much for sharing an office with me, which is probably a burden sometimes.

The CSE cluster

Mihir Shah

Francisca Lameiras

Yi Song

Christina Bernal

Sushil Gaykawad

Anisha Goel

Hugo Cueto-Rojas

Camilo Suarez-Mendez

Joana Gomes

Yaya Liu

Florence Lip

Alexandre Carnet

Angela ten Pierick

Reza Maleki Seifar

Cor Ras

For all that have not been in the unfortunate position of sharing an office with me, it was a great pleasure working with you.

The EBT/Tango Tuesday cluster

Marissa Boleij

Eveline van der Berg

Danny de Graaf

Bart Joosse

Jules Rombouts

Monica Conthe

Emmaouela Korkakaki

Ingrid Pinel

Thank you for fraternizing with me although bearing the stigma of working with monoseptic systems.

The Nantia Bikaki cluster
Nantia Bikaki

What can I say, you are special.

The BPT cluster
Sylvia Pirrung
Joana Pereira
Alex Hanke
Susanna Pedraza de la Cuesta
Carlos Cabrera-Rodríguez

Thank you guys, still trying to figure out where they moved you in the new building.

The IMB cluster
Jasmine Bracher
Angel Sevilla
Ioannis Papapetridis

Thanks for bearing with me, although I do not know how to clone genes.

The unclusterizable cluster
Sharzad Naghibzadeh
Ana Malianu
Isabelle Duijnste
Joud Hazineh
Daniela Quintanilla
Manuel Salvador
Rik Schuffelen
Dennis Ruetze
Cees Haringa
Fatima Anisi
Weiwei Li

Thank you very much for making the time in Delft enjoyable and memorable.

The project cluster

Robert Mans

Ryan Henderson

Bert Poolman

Ton van Maris

This project was incredible complex and I believe we can be proud of what we have learned, thank you very much for your contributions.

The students cluster

Joerian Droog


Heleen Ouboter

Michael Rodrian

Francisco Pinheiro

Margo van Berkum

Thank you very much for your effort and dedication contributing to this work.



ISBN 978-94-6375-156-8