



Resource Allocation Strategies and Adaptive Evolution: A Multi-Scale Modeling Approach in Cyanobacteria

Inaugural dissertation

for the attainment of the title of doctor
in the Faculty of Mathematics and Natural Sciences
at the Heinrich Heine University Düsseldorf

presented by

Sajjad Ghaffari Nasab Sharabiani

from Tehran, Iran

Düsseldorf, August 2025

from the Institute for Computer Science
at Heinrich Heine University Düsseldorf

Published by permission of the
Faculty of Mathematics and Natural Sciences at
Heinrich Heine University Düsseldorf

Supervisor: Prof. Dr. Martin Lercher
Co-supervisor: Prof. Dr. Oliver Ebenhöh

Date of the oral examination: 27.10.2025

Declaration

I declare under oath that I have produced my thesis independently and without any undue assistance by third parties under consideration of the “Principles for the Safeguarding of Good Scientific Practice at Heinrich Heine University Düsseldorf”.

Düsseldorf, August 28, 2025

Sajjad Ghaffari Nasab Sharabiani

Acknowledgements

I would like to thank my supervisor Prof. Dr. Martin Lercher, for giving me the opportunity to work on exciting and challenging projects during my PhD and for his invaluable guidance and continuous support throughout the course of my doctoral studies.

I am also thankful to Prof. Dr. Oliver Ebenhöf for his constructive feedback and for the opportunity to spend a short and inspiring stay in his lab. Furthermore, I would like to thank Prof. Dr. Daniel Ducat from Michigan State University, whose warm hospitality and inspiring research environment during my stay in his lab greatly broadened my scientific perspectives.

I would also like to thank Dr. Hugo Dourado, whose help and collaboration were of great importance during my PhD. His advice, discussions, and encouragement have been instrumental in shaping the direction of my research.

I gratefully acknowledge the iGrad-Plant program for funding my research and supporting my academic development.

I sincerely thank all my colleagues at the CCB group at Heinrich Heine University. The fruitful discussions, and supportive environment have made my time here both productive and enjoyable.

On a personal note, I owe my deepest gratitude to my loving wife, Mana, for her unwavering love, patience, and encouragement throughout this journey. I am also forever grateful to my family for their constant support. This thesis is lovingly dedicated to all of them, and especially to the memory of my mother, whose absence I deeply miss but whose love and values continue to guide and inspire me every day.

Contents

Introduction.....	1
1.1 General Introduction	2
1.2 Genome-Scale Metabolic Modeling	3
1.3 Flux Balance Analysis (FBA)	3
1.4 <i>Synechocystis</i> sp. PCC 6803: A Model Cyanobacterium for Metabolic and Biotechnological Insights.....	4
1.5 Flux Balance Analysis as a Tool to Investigate Metabolic Complexity and Evolutionary Strategies in Cyanobacteria.....	5
1.6 Conceptual and Computational Limitations of FBA	6
1.7 Beyond Traditional FBA: Linear Models of Resource Allocation and the Challenge of Integrating Proteome Constraints	6
1.8 Non-linear Framework for Understanding Cellular Resource Allocation	8
1.9 Growth Balance Analysis	8
1.10 Integrative Overview of the Three Studies	12
Metabolic Complexity and Its Role in Cyanobacterial Adaptation	14
2.1 Introduction.....	16
2.2 Results and Discussion	17
2.3 Conclusions.....	23
2.4 Materials and Methods.....	24
2.5 References.....	31
2.6 Supplementary Information for Metabolic Complexity and Its Role in Cyanobacterial Adaptation.....	34
Cyanobacterial resource allocation strategies explained by Growth Balance Analysis	53
3.1 Introduction.....	55
3.2 Results.....	59
3.3 Discussion	70
3.4 Materials and methods	72
3.5 References.....	79
3.6 Supplementary text for Cyanobacterial resource allocation strategies explained by Growth Balance Analysis	82
<i>Cell Growth Simulator</i> : a user-friendly web server for growth balance analysis.....	86
4.1 Introduction.....	90
4.2 Design and implementation	93
4.3 Results.....	97
4.4 Discussion	101

4.5	References.....	103
	Discussion and Outlook	106
5.1	Discussion	107
5.2	Future directions	111
	References.....	113

List of Figures

Introduction.....	1
Figure 1. Schematic representation of the Growth Balance Analysis (GBA) framework...	11
Metabolic Complexity and Its Role in Cyanobacterial Adaptation	14
Figure 1. The models included represent a broad range of metabolic complexity.	18
Figure 2. The fraction of viable environments differs widely across submodels	19
Figure 3. The number of additional reactions required for adaptation decreases with increasing genome size.	22
Figure S1. Networks sizes (number of reactions) of all models combined in the supermodel.	41
Figure S2. Additional properties of the supermodel.	42
Figure S3. Small metabolic networks tend to be less branched than large networks.	43
Figure S4. Fractions of viable environments for submodels in wet lab environments (seed).	44
Figure S5. The fraction of viable environments differs widely across submodels.....	45
Figure S6. The number of additional reactions required for adaptation decreases with increasing genome size.	46
Figure S7. The number of additional reactions required for adaptation decreases with increasing genome size.	47
Figure S8. Different <i>E. coli</i> strains show similar adaptabilities, despite variations in genome size.	48
Figure S9. Across species, generalists (blue) show a higher tendency for collateral adaptations than specialists (red), while this trend is reversed within <i>E. coli</i>	49
Figure S10. Specialists (red) tend to have higher exaptation indices than generalists (blue).	50
Figure S11. Practical application of ARM LP shows equally good performance as ARM MILP.	51
Figure S12. Growth for submodels and supermodel	52
Cyanobacterial resource allocation strategies explained by Growth Balance Analysis	53
Figure 1. Overview of the base GBA model for <i>Synechocystis</i>	62
Figure 2. Simulated growth curve of <i>Synechocystis</i> with photoinhibition.	63
Figure 3. The predicted proteome allocation in the GBA model.....	64
Figure 4. The mass fraction matrix for the extended GBA model.....	65
Figure 5. Overview of the extended GBA model for <i>Synechocystis</i>	66
Figure 6. Growth curves of <i>Synechocystis</i> in the extended GBA	68
Figure 7. <i>Synechocystis</i> resource allocation predicted by the extended GBA model.....	70
<i>Cell Growth Simulator</i> : a user-friendly web server for growth balance analysis.....	86

Fig 1. Schematic representation of the Growth Balance Analysis (GBA) framework	92
Fig 2. Analysis of an example GBA model	100
Discussion and Outlook	106
References.....	113

List of Tables

Introduction.....	1
Metabolic Complexity and Its Role in Cyanobacterial Adaptation	14
Table 1. Definition of variables of the ARM LP.	26
Table S1. Organism specific models and their properties	35
Table S2. Essential Reactions Added to Enable Growth on Fully Rich Medium	39
Table S3. Common Subsystems with Associated Gains in Simulations and Experiments .	40
Cyanobacterial resource allocation strategies explained by Growth Balance Analysis	53
Table S1. Parameters of the base model	83
Table S2. Parameters of the extended model.....	84

Summary

Cyanobacteria are the only known prokaryotes that perform oxygenic photosynthesis, serving as key primary producers across diverse habitats and playing a fundamental role in the earth's ecosystem. In addition to their ecological importance, cyanobacteria also hold great promise as sustainable resources for the production of industrially and medically valuable compounds. Because of their importance, an important aim is to optimize how these organisms allocate their limited cellular resources across various metabolic processes. The main goal of this thesis is to address this challenge by developing and applying advanced computational frameworks to study phototrophic growth and resource allocation in cyanobacteria. In line with this objective, Chapter 1 serves as an introduction, providing an overview of cyanobacteria and explaining the computational strategies used to study their growth and metabolism.

This thesis comprises three studies. In Chapter 2 (manuscript 1), we develop a pan-genome-scale metabolic model and use flux balance analysis (FBA), a linear approach to, investigate the adaptability of 102 unicellular organisms – including a range of cyanobacteria – to new nutrient sources under heterotrophic growth conditions. The analysis revealed a strong correlation between genome size and the number of reactions necessary for these adaptations.

In the second study in Chapter 3 (manuscript 2), we reformulate and extend an existing model of the cyanobacterium *Synechocystis*. We introduce Growth Balance Analysis (GBA) as an alternative mathematical framework for modeling phototrophic growth in cyanobacteria. Compared to the methodology in Faizi et al. (2018), GBA formulates equivalent models with simpler equations that allow more efficient calculations and easier model extensions through additional reactions and metabolites. As a proof of concept, we first present a GBA model inspired by Faizi et al. (2018), using identical parameter values. We then present a second, extended GBA model capable of predicting the optimal proteome allocation in more detail.

In the third study in Chapter 4 (manuscript 3), we present *Cell Growth Simulator*, a web-based application that enables efficient construction and analysis of GBA models through a user-friendly interface and interactive visualizations. *Cell Growth Simulator* uses an intuitive spreadsheet interface, eliminating the need for coding, and integrates data from the BRENDA enzyme database to facilitate the incorporation of kinetic parameters. *Cell Growth Simulator* helps users to interpret optimization results through customizable plots and dynamic metabolic pathway maps. The platform makes nonlinear modeling of resource allocation in coarse-grained cellular systems accessible to a broad scientific audience, providing an intuitive tool for advancing our understanding of cellular metabolism and growth and fostering interdisciplinary collaborations.

Overall, this work offers both theoretical and practical contributions: it sheds light on cyanobacterial adaptability and resource allocation while providing accessible computational frameworks to facilitate deeper explorations of phototrophic (and non-phototrophic) growth.

Zusammenfassung

Cyanobakterien sind die einzigen bekannten Prokaryoten, die eine oxygene Photosynthese durchführen. Sie fungieren als zentrale Primärproduzenten in vielfältigen Habitaten und spielen eine grundlegende Rolle im Ökosystem der Erde. Neben ihrer ökologischen Bedeutung gelten Cyanobakterien zudem als vielversprechende, nachhaltige Ressourcen für die Herstellung industriell und medizinisch wertvoller Verbindungen. Aufgrund ihrer Bedeutung ist es ein zentrales Ziel, zu optimieren, wie diese Organismen ihre begrenzten zellulären Ressourcen auf verschiedene Stoffwechselprozesse verteilen.

Das Hauptziel dieser Arbeit ist es, diese Herausforderung mithilfe der Entwicklung und Anwendung fortgeschrittener computergestützter Rahmenwerke anzugehen, um phototrophes Wachstum und Ressourcenallokation in Cyanobakterien zu untersuchen. Im Einklang mit diesem Ziel dient Kapitel 1 als Einführung, bietet einen Überblick über Cyanobakterien und erläutert die rechnerischen Strategien, die zur Untersuchung ihres Wachstums und Stoffwechsels eingesetzt werden.

Diese Dissertation umfasst drei Studien. In Kapitel 2 (Manuskript 1) entwickeln wir ein pan-genomisches metabolisches Modell und verwenden die Flux-Balance-Analysis (FBA), einen linearen Ansatz, um die Anpassungsfähigkeit von 102 einzelligen Organismen – darunter verschiedene Cyanobakterien – an neue Nährstoffquellen unter heterotrophen Wachstumsbedingungen zu untersuchen. Die Analyse zeigte eine starke Korrelation zwischen der Genomgröße und der Anzahl der für diese Anpassungen erforderlichen Reaktionen.

In der zweiten Studie in Kapitel 3 (Manuskript 2) reformulieren und erweitern wir ein bestehendes Modell des Cyanobakteriums *Synechocystis*. Wir führen die Growth Balance Analysis (GBA) als alternatives mathematisches Rahmenwerk zur Modellierung phototrophen Wachstums in Cyanobakterien ein. Im Vergleich zur Methodik von Faizi et al. (2018) formuliert GBA äquivalente Modelle mit einfacheren Gleichungen, die effizientere Berechnungen und eine leichtere Erweiterung der Modelle durch zusätzliche Reaktionen und Metaboliten ermöglichen. Als Machbarkeitsnachweis präsentieren wir zunächst ein GBA-Modell, das von Faizi et al. (2018) inspiriert ist und identische Parameterwerte verwendet. Anschließend stellen wir ein zweites, erweitertes GBA-Modell vor, das in der Lage ist, die optimale Proteom-Allokation detaillierter vorherzusagen.

In der dritten Studie in Kapitel 4 (Manuskript 3) stellen wir den *Cell Growth Simulator* vor, eine webbasierte Anwendung, die durch eine benutzerfreundliche Oberfläche und interaktive Visualisierungen eine effiziente Erstellung und Analyse von GBA-Modellen ermöglicht. Der *Cell Growth Simulator* nutzt eine intuitive Tabellenkalkulationsoberfläche, wodurch Programmieren überflüssig wird, und integriert Daten aus der BRENDA-Enzymdatenbank, um die Einbindung kinetischer Parameter zu erleichtern. Er unterstützt Nutzer bei der Interpretation von Optimierungsergebnissen durch anpassbare Diagramme und dynamische Stoffwechselwegkarten. Die Plattform macht die nichtlineare Modellierung der Ressourcenallokation in grobkörnigen zellulären Systemen einem breiten wissenschaftlichen Publikum zugänglich und bietet ein intuitives Werkzeug zur Vertiefung unseres Verständnisses von Zellstoffwechsel und Wachstum sowie zur Förderung interdisziplinärer Zusammenarbeit. Insgesamt leistet diese Arbeit sowohl theoretische als auch praktische Beiträge: Sie beleuchtet die Anpassungsfähigkeit und Ressourcenallokation von Cyanobakterien und stellt gleichzeitig zugängliche rechnerische Rahmenwerke bereit, die tiefere Untersuchungen des phototrophen (und nicht phototrophen) Wachstums ermöglichen.

Chapter 1

Introduction

1.1 General Introduction

Rising atmospheric CO₂ levels, now at their highest point in human history and a primary driver of climate change, together with accelerating population growth and expanding industrial activity, have put unprecedented pressure on natural resources. The overexploitation of these resources, combined with carbon emissions derived from fossil fuels, has already caused significant environmental damage. In response, there is an urgent need to develop sustainable and environmentally friendly strategies for the production of alternative fuels and chemicals (Luan & Lu, 2018). Microbial biomanufacturing is a promising solution, as both native and engineered microbes have been successfully used to produce a range of renewable biofuels and biochemicals. However, most of these microbial platforms – such as *Escherichia coli* and *Saccharomyces cerevisiae* – are heterotrophic and require organic carbon feedstocks (e.g., glucose or other sugar-based substrates) to sustain growth (Dodds, 2002; Jojima et al., 2010; Wang et al., 2019). In contrast, cyanobacteria use photosynthesis to directly convert atmospheric carbon into biomass, making them particularly attractive for sustainable bioproduction (Singh et al., 2016).

Cyanobacteria are a diverse and widespread group of prokaryotes known for their ability to perform oxygenic photosynthesis. Utilizing sunlight as an energy source and carbon dioxide as a feedstock, these photoautotrophic organisms can grow rapidly and have a relatively simple cellular structure. In addition, they are amenable to genetic manipulation, making them excellent model systems for studying photosynthesis and compelling hosts for biotechnological applications. In particular, engineered cyanobacteria are now well established as producers of several valuable chemicals (Knoot et al., 2018). However, their productivity and product titers have remained relatively low compared to those of heterotrophic hosts such as *Escherichia coli*, limiting the commercial potential of cyanobacterial bioproduction. Early research on cyanobacteria focused primarily on strains that were easily genetically tractable, laying the groundwork for more advanced strain engineering. Among these, *Synechocystis* sp. PCC 6803, *Synechococcus elongatus* PCC 7942, and *Synechococcus* sp. PCC 7002 have emerged as key unicellular models and remain widely studied for both fundamental and applied research (Mukherjee et al., 2020).

A promising approach toward understanding how these model systems can be made more efficient hosts for biotechnological applications is to take a systems biology perspective. Systems biology takes a holistic approach to understanding how genes, proteins, and metabolites interact within complex biological networks (Kitano, 2002). Over the past two decades, rapid advances in “omics” technologies have generated large, high-quality data sets that form the basis for increasingly sophisticated metabolic and regulatory models. These computational tools provide a powerful complement to laboratory experiments, which are often time-consuming and expensive. By simulating key aspects of wet-lab work, *in silico* methods can not only reduce the number of experiments required, but also allow researchers to explore broader experimental conditions and hypotheses.

Cyanobacteria became an early focus of systems biology because of their evolutionary importance: they pioneered oxygenic photosynthesis about 2.5 billion years ago, a trait that was subsequently transferred to eukaryotes through endosymbiosis, eventually giving rise to

algae and plants (Hohmann-Marriott & Blankenship, 2011). The close evolutionary relationship of cyanobacteria with plant plastids stimulated early genomic investigations, leading to the first publicly available complete genome sequence of *Synechocystis* sp. PCC 6803 (hereafter *Synechocystis*) more than two decades ago (Kaneko et al., 1996a). Equipped with robust genetic tools (Grigorieva & Shestakov, 1982), *Synechocystis* quickly became a widely used model organism among cyanobacteria. Today, due to their ecological relevance and industrial potential, more than 200 complete genome sequences of different cyanobacterial species have been made publicly available (Shih et al., 2013). These characteristics establish cyanobacteria as an important model organism for the development of microbial cell factories (Santos-Merino et al., 2023). Despite the abundance of high-throughput experimental data, including genomics, transcriptomics, and proteomics (Babele et al., 2019; Jahn et al., 2018; Matthias et al., 2014; Zavřel et al., 2019), achieving a fundamental mechanistic understanding of resource allocation in cyanobacteria remains a major challenge in biotechnology. The main goal of this thesis is to address this issue by developing and applying advanced computational frameworks to study phototrophic growth and resource allocation in cyanobacteria.

1.2 Genome-Scale Metabolic Modeling

The growing demand for quantitative insights into metabolic physiology and bioprocess optimization has led to extensive efforts in the mathematical modeling of metabolic function. One of the most powerful approaches in this area is the use of genome-scale metabolic models (GEMs), which use a stoichiometric matrix to represent the entire metabolic reaction network of an organism. GEMs are based on gene-protein reaction (GPR) associations that integrate annotated genomic data with experimentally derived information to create mass-balanced reconstructions of cellular metabolism (Bernstein et al., 2021).

A key advantage of GEMs is their ability to incorporate species-specific knowledge and complex 'omics data, providing a holistic framework for understanding cellular metabolism. Combined with constraint-based reconstruction and analysis (COBRA) techniques – most notably Flux Balance Analysis (FBA) – GEMs enable the translation of biological hypotheses into computational algorithms that can predict metabolic phenotypes. As tools for combining computational modeling with empirical data, GEMs have become indispensable for advancing our understanding of metabolic physiology and guiding the design of more efficient biotechnological processes (Gu et al., 2019).

1.3 Flux Balance Analysis (FBA)

Flux Balance Analysis (FBA) is a widely used constraint-based optimization framework that employs a stoichiometric representation of metabolic networks, ranging from simplified “core” models to comprehensive genome-scale reconstructions (Orth et al., 2010). The method requires the assumption of an optimality criterion (usually the maximization of the biomass yield in microbial systems (Feist & Palsson, 2010), often imprecisely referred to as growth rate maximization (Schuster et al., 2008)). This optimization in turn requires a limiting constraint, such as a limit on the uptake rate of an essential nutrient. Where available, kinetic parameters (*e.g.*, v_{max} values) can be incorporated as flux bounds to refine the model. In its

standard application, FBA finds the maximal biomass production rate $v_{biomass}$ and its corresponding flux distribution v for a given uptake constraint and stoichiometric matrix S . Formally, this defines the linear optimization problem:

$$\begin{aligned} &\text{Maximize } v_{biomass} \\ &\text{at Steady state condition } Sv = 0 \\ &\text{with } v_{min} \leq v \leq v_{max} \end{aligned} \tag{1}$$

Solving the FBA problem yields a set of metabolic flux distributions that satisfy the given objective. Due to its linear formulation and relatively low computational cost, FBA scales well to large metabolic networks with thousands of reactions. It is therefore particularly suitable for modeling genome-scale networks, studying microbial communities or multicellular tissues, and performing parameter scans under varying external conditions (Töpfer et al., 2020). However, by design, FBA does not inherently account for factors such as temporal dynamics, regulatory constraints, or experimentally measured transcription, protein, or metabolite abundances; these considerations typically require additional modeling frameworks or hybrid approaches (Sahu et al., 2021).

1.4 *Synechocystis* sp. PCC 6803: A Model Cyanobacterium for Metabolic and Biotechnological Insights

To date, *Synechocystis* is the best-characterized cyanobacterium and the model organism of choice for applications in biotechnology due to its robust growth characteristics and well-studied biology (Angermayr et al., 2009). This unicellular cyanobacterium, with a cell size of approximately 2 μm (Zavřel et al., 2017), was isolated in 1968 from a freshwater lake in California, USA (Stanier et al., 1971). *Synechocystis* is naturally transformable and genetically tractable. In 1996, it was the first phototrophic organism to have its genome fully sequenced, marking a milestone as the third bacterial genome ever sequenced (Kaneko et al., 1996b). Its genome consists of a single chromosome and seven plasmids, with a total size of approximately 3.95 megabases (Mb). A total of 3,507 protein-coding genes have been annotated in the UniProt database (Kaneko et al., 2003). The organism exhibits a maximum growth rate of approximately 0.135 $[\text{h}^{-1}]$, corresponding to a doubling time of 5.13 hours (Zavřel et al., 2015).

Building on this foundational knowledge, several GEMs have been developed for *Synechocystis*, reflecting its status as a model organism for the study of cyanobacterial metabolism and photosynthesis. One of the earliest GEMs, iSyn669, was published in 2011 and included 882 reactions, 690 metabolites, and 669 genes, providing a basic framework for understanding the metabolic network of *Synechocystis* (Montagud et al., 2011). This model was later refined and expanded into iSyn731, which incorporated additional experimental data and improved GPR associations, resulting in a more accurate representation of the organism's metabolic capabilities (Knoop et al., 2013). Another notable model, iJN678, was developed to incorporate detailed descriptions of phototrophic growth and carbon fixation pathways,

allowing simulations of light-driven metabolism and resource allocation under varying environmental conditions (Nogales et al., 2012). Höper et al. (2024) recently released an updated GEM for *Synechocystis* that spans 865 reactions and 783 genes, which, crucially, integrates explicit modeling of light absorption and thereby advances the quantitative analysis of phototrophic metabolism. This groundlaying work, in conjunction with the availability of its genome annotation on CyanoBase, positions *Synechocystis* as a suitable model for examining photoautotrophic growth and resource allocation, as explored in this study.

1.5 Flux Balance Analysis as a Tool to Investigate Metabolic Complexity and Evolutionary Strategies in Cyanobacteria

FBA provides a robust computational framework for investigating how metabolic fluxes and phenotypic behaviors respond to various genetic and environmental factors, making it particularly well-suited for assessing the impact of newly acquired genes on metabolic networks and organismal fitness. Horizontal gene transfer (HGT) – the transfer of genetic material across species boundaries – is a key evolutionary mechanism that has profoundly influenced microbial genome evolution, adaptation, and metabolic diversity. In particular, recent work has shown that *Escherichia coli* can acquire complex metabolic traits with relatively few new enzymatic functions, highlighting the remarkable adaptability of bacterial metabolism. However, it remains unclear whether this adaptability is unique to *E. coli* or represents a more general feature of bacterial lineages (Szappanos et al., 2016). Cyanobacteria offer an especially informative test case. These photoautotrophs thrive in diverse habitats, from nutrient-poor open oceans to hypersaline ponds and hot deserts (Oren, 2015). They adapt their metabolic strategies to meet the thermodynamic demands of photosynthetic electron transport, carbon-concentrating mechanisms and, often, nitrogen fixation (Bothe et al., 2010). Their genomes are highly modular, with large accessory gene pools – especially in freshwater *Synechocystis* – suggesting repeated instances of gene gain and loss that adapt metabolism to local light, CO₂ concentration, and nutrient availability (Jeong et al., 2021). Such rewiring can create pathways that alter redox balance via alternative electron sinks or reconfigure the ATP/NADPH ratio through cyclic electron flow. Through their integration of photosynthesis, cyanobacteria add a compelling dimension to the broader question of how microbes adapt to diverse environments. Despite their ecological significance, comparatively little is known about the genetic and network-level factors that enable or limit their metabolic flexibility – an important question that is addressed in manuscript 1 of this dissertation.

Several theoretical models have attempted to explain how new metabolic functions emerge and are integrated into existing networks. The “toolbox model” proposed by Maslov et al. (2009) posits that each new gene (*i.e.*, tool) can combine with the organism's existing “toolbox” of enzymes to generate additional metabolic pathways, creating a synergistic effect for organisms already equipped with a large toolbox. Another complementary hypothesis comes from (Wolf & Koonin, 2013), who introduced a “biphasic” model of genome evolution. Their phylogenetic analyses suggest alternating phases of genome expansion and contraction, potentially giving rise to generalist and specialist microbial lifestyles. While these models offer valuable perspectives on evolutionary dynamics, the underlying reasons that

drive certain organisms to evolve as generalists – able to thrive in a variety of environments – and others to evolve as specialists remain incompletely understood.

Manuscript 1 of this dissertation addresses this gap by examining 102 unicellular organisms, including several cyanobacterial strains, to determine how metabolic adaptability correlates with genome content. Using FBA simulations, the study shows that species with larger metabolic “toolboxes” (*i.e.*, gene repertoires) adapt more readily to novel nutrient conditions, consistent with the “toolbox model” of Maslov et al. (2009). Indeed, manuscript 1 classifies organisms as either generalists or specialists, and shows that generalists (including some cyanobacteria) have branching metabolic networks and derive multiple ancillary benefits from a single adaptive event. In contrast, specialists have more linear pathways in which adaptive gains often serve narrowly defined purposes. Crucially, this work sheds light on how HGT and intrinsic network architecture together govern the adaptive potential of cyanobacteria, a group whose photosynthetic capacity and evolutionary history make it uniquely relevant for both fundamental research and innovative biotechnological applications.

1.6 Conceptual and Computational Limitations of FBA

The linear optimization framework (Eq. 1) is computationally simple, but this simplicity comes with certain conceptual limitations. For example, the maximum flux towards biomass production ($v_{biomass}$) typically depends on the upper bounds of key uptake reactions (*e.g.*, limiting carbon sources). In contradiction to the notion of growth rate maximization, the solution may favor maximum yield (the ratio of biomass flux to uptake flux of the limiting nutrient) rather than maximum growth rate per se (Schuster et al., 2008). Moreover, because biomass production is a direct (and, in the absence of a maintenance energy term, linear) function of the maximal uptake flux of the limiting nutrient, accurate prediction of optimal biomass flux requires prior knowledge of substrate uptake rates.

A second challenge arises when large-scale metabolic reconstructions have more reactions than metabolites, which is typically the case. The resulting structure of the stoichiometric matrix leads to an underdetermined system and multiple valid solutions for the flux vector v at a given value of the objective function (Orth et al., 2010). As a result, additional criteria are needed to determine a single, biologically relevant optimal flux distribution. Importantly, these challenges are unproblematic for the work reported in manuscript 1: here, only the metabolic network’s ability to produce biomass from a set of available nutrients is important, while growth rate and detailed flux distributions are not considered. In contrast, the detailed and accurate study of molecular physiology – as examined in manuscript 2, see below – requires attention to these problems.

1.7 Beyond Traditional FBA: Linear Models of Resource Allocation and the Challenge of Integrating Proteome Constraints

Parsimonious FBA (pFBA) (Holzhütter, 2004) is one of the most common strategies to address the requirement to choose among multiple optimal FBA solutions. pFBA reflects the principle that cells can optimize growth by reducing enzymatic resource allocation wherever possible. Under the simplifying assumptions that the absolute values of fluxes are

proportional to the required enzyme investment and that the corresponding proportionality constants are identical across enzymes, pFBA finds the most resource-efficient solution by minimizing the sum of absolute fluxes at the maximal biomass production rate. Thus, pFBA can be seen as the simplest variant of resource allocation models – a class of models built on linear optimization frameworks to describe how cellular resources such as proteins, metabolites, and energy are allocated among different processes to maximize growth or other objective functions. Like pFBA, more advanced resource allocation models linearize the kinetic relationships between enzyme abundance and metabolic fluxes (and hence growth rate). These linear models provide a computationally tractable way to capture large-scale cellular behavior (Goelzer et al., 2011; Goelzer & Fromion, 2011).

At their core, resource allocation models generalize the idea of imposing explicit constraints on metabolic capacity. Two main approaches have emerged to address this "budgeting" problem: (i) protein budgeting, in which a fixed pool of protein must be allocated among different cellular tasks, and (ii) resource budgeting, in which protein budgeting is further coupled with descriptions of protein synthesis requirements. However, conventional protein budgeting approaches often assume that protein production itself is governed by a pre-existing budget, rather than dynamically feeding back into it.

Two examples of protein budgeting frameworks are FBA with molecular crowding (FBAwMC) (Vazquez et al., 2008) and Constrained-Allocation FBA (CAFBA) (Mori et al., 2016), which constrain enzyme concentrations for the complete system or individually for proteome sectors. Although these methods can improve predictive accuracy, they do so by imposing extra assumptions. Nevertheless, such approaches have found that growth-maximizing solutions, especially when coupled with genetic regulation, often match experimental phenotypes. Later resource allocation models have extended these frameworks by more fully integrating proteome partitioning and other capacity constraints (e.g., GECKO (Sánchez et al., 2017)).

Three major large-scale modeling strategies with resource budgeting are currently in use: Resource Balance Analysis (RBA) (Goelzer et al., 2011), Metabolic and Macromolecular Expression (ME) models (O'Brien et al., 2013), and Proteome Constrained (pc) models (Elseman et al., 2022). Each formulates an optimization problem at a fixed growth rate under linear constraints. RBA, introduced by (Goelzer et al., 2011), extends beyond metabolism to include other growth-related macromolecules (e.g., DNA, membranes) and the cellular translation apparatus, and, in some implementations, incorporates growth rate-dependent enzyme catalytic rates ($k_j(\mu) = a_j\mu + b_j$). This approach, calibrated against proteomic data, accurately predicts proteome partitioning in *Bacillus subtilis* (Goelzer et al., 2015). ME models, proposed by (Lerman et al., 2012), take a slightly different approach by embedding the gene expression machinery directly into the metabolic network and generating solutions that minimize ribosome production at the given growth rate. Unlike RBA, ME models do not impose explicit density constraints, but retain certain assumptions from the underlying metabolic framework, such as the use of fixed exchange flux limits.

Despite the success of these linearized resource allocation methods, they rely on important simplifying assumptions. All of the mentioned models assume that cells operate under an optimal state that can be found via linear optimization, thus ignoring the inherently non-linear

nature of enzymatic rate laws and the interplay between substrate concentrations and enzyme efficiency. This approximation is justified by the fact that fully nonlinear genome-scale models – involving thousands of variables – pose significant computational challenges with current technologies (Orth et al., 2010). To remain tractable, all methods discussed so far treat reaction fluxes (v_j) as proportional to enzyme abundance (e_j) multiplied by a catalytic rate (k_j) that is assumed to be either constant or, in some RBA implementations, a linear function of growth rate. As a result, the effects of substrate concentrations are ignored and other nonlinear phenomena are often overlooked.

1.8 Non-linear Framework for Understanding Cellular Resource Allocation

While all models discussed so far approximate cellular metabolism as a linear process, biological systems are inherently nonlinear. In contrast to linear cellular models, nonlinear approaches explicitly capture metabolite concentrations and their effects on reaction fluxes via kinetic rate laws. Molenaar et al. (2009) pioneered “self-replicator” models in which resource allocation emerges from optimizing cellular growth under key physiological constraints, including nonlinear kinetics in transport and enzymatic reactions, a fixed total protein concentration, and the allocation of ribosomes to produce all necessary proteins. Subsequent studies have extended this framework to photoautotrophic systems, incorporating processes such as photodamage and carbon cycling (Burnap, 2015; Faizi et al., 2018).

So far, nonlinear kinetic cell models have focused on small, coarse-grained representations of cellular physiology (Burnap, 2015; Faizi et al., 2018; Jahn et al., 2018; Molenaar et al., 2009). These models compress cellular complexity by representing multiple enzymes or pathways with a single catalytic unit. While lacking in molecular detail, these simplified models still provide valuable insights into overarching metabolic trade. Their limited scale reflects the computational difficulty of solving large nonlinear optimization problems (Wortel et al., 2018).

1.9 Growth Balance Analysis

Recently, growth balance analysis (GBA; Figure 1) has been introduced as a general framework for studying nonlinear resource allocation in growing cells subject to mass conservation, nonlinear reaction kinetics, and cell density constraints (Dourado et al., 2023; Dourado & Lercher, 2020). This work not only demonstrates the feasibility of large-scale nonlinear simulations, but also provides new perspectives on the mathematical properties of such models. In GBA, a cellular model is specified by a triple $(\mathbf{M}, \boldsymbol{\tau}, \rho)$. The matrix $\mathbf{M}$ is a mass-fraction form of the stoichiometric matrix, scaled by molecular weights and including a special ribosome reaction “r” that synthesizes proteins, as well as a row “p” that represents total protein concentration. The vector $\boldsymbol{\tau} = \boldsymbol{\tau}(\mathbf{c}, \mathbf{x})$ encodes turnover times for each reaction in the network; these kinetic functions are the inverses of the more customary kinetic rate laws, and they depend on internal metabolite concentrations $\mathbf{c}$ and on relevant external concentrations $\mathbf{x}$. Finally, ρ indicates the sum of all mass concentrations inside the cell, encompassing both metabolites and proteins, and is assumed to remain fixed.

Moreover, the balanced growth model at steady-state is specified by the following constraints:

$$\begin{aligned}
\sum_j M_j^i v^j &= \mu c^i && \text{(Mass conservation)} \\
v^j &= p^j / \tau^j(c, x) && \text{(Reaction kinetics)} \\
\sum_j p_j &= c^p && \text{(Protein density)} \\
\sum_i c^i &= \rho && \text{(Cellular density)} \\
c^m &\geq 0 && \text{(Non-negativity of metabolite concentration)} \\
p^j &\geq 0 && \text{(Non-negativity of protein concentration)}
\end{aligned}$$

Because $\mathbf{M}$ is scaled in mass units, both metabolite concentrations c^i and protein concentrations p_j appear in units of $[\text{g L}^{-1}]$, and fluxes v^j have units of $[\text{g L}^{-1} \text{ h}^{-1}]$. Kinetic parameters – such as Michaelis constants (K_m) ($[\text{g L}^{-1}]$) and turnover numbers (k_{cat}) (the amount of product per unit of protein per unit of time, resulting in units of $[\text{h}^{-1}]$) – should also be expressed in mass units. By definition, ρ (in $[\text{g L}^{-1}]$) is the sum of the mass concentrations of every intracellular component, a quantity taken to be constant under balanced growth, consistent with observations in *E. coli* (Bremer & Dennis, 2008).

Within GBA, the optimal state of the cell emerges as a solution to an optimization problem built on two core algebraic expressions, formulated in terms of flux fractions $\mathbf{f} := \mathbf{v}/(\mu\boldsymbol{\rho})$. Here, $\mathbf{v}$ denotes the mass reaction fluxes, μ is the growth rate, ρ is the constant total density of intracellular components, and the ratio is taken element-wise. The first core equation,

$$\mu(\mathbf{f}, \mathbf{x}) = \frac{M_r^p f^r}{\mathbf{f}^T \boldsymbol{\tau}(\rho \mathbf{M} \mathbf{f}, \mathbf{x})} \quad (2)$$

focuses on how the ribosome reaction contributes to the growth rate, where M_r^p is the relevant entry in $\mathbf{M}$. The second equation,

$$\sum_{i,j} M_j^i f^j = 1 \quad (3)$$

ensures a strict balance on total cellular density. Here, the summation runs over the internal reactants (i) and the reactions (j) in $\mathbf{M}$. Apart from this explicit density condition, all other constraints – such as reaction stoichiometry, nutrient limitations, and enzyme capacities – are encoded in the first equation or in the expressions for $\boldsymbol{\tau}$ (Dourado et al., 2023). Solving for $\mathbf{f}$ at optimal growth also determines key cellular properties, including metabolite levels $\mathbf{c}_m$, total protein concentration $\mathbf{c}_p$, and each individual protein concentration p^j via

$$\mathbf{c} = \rho \mathbf{M} \mathbf{f} \quad (4)$$

$$p^j = \mu \rho f^j \tau^j(\rho \mathbf{M} \mathbf{f}, \mathbf{x}) \quad (5)$$

The turnover times τ are defined by reaction rate laws (here, irreversible Michaelis–Menten with inhibition (eq. 6)), so that τ values depend systematically on the intracellular metabolite pool $\mathbf{c}_m$ and, for transport reactions, on the external environment $\mathbf{x}$.

$$\tau_j(c, x) = \frac{1}{k_{catf}^j} \prod_m \left(\frac{c_m}{K_m^j + c_m} \right)^{-1} \left(\frac{c_m}{KA_m^j + c_m} \right)^{-1} \left(\frac{KI_m^j}{KI_m^j + c_m} \right)^{-1} \prod_n \left(\frac{x_n}{K_n^j + x_n} \right)^{-1} \quad (6)$$

Here,

- k_{catf}^j is the forward turnover number;
- K_m^j is the Michaelis constant for internal metabolite m in reaction j
- KA_m^j is the activation constant for metabolite m in reaction j ;
- KI_m^j is the inhibition constant for metabolite m in reaction j ;
- K_n^j is the Michaelis constant for external reactant n in reaction j ;
- products over m run across internal metabolites involved in reaction j ;
- products over n run across external reactants transported by j .

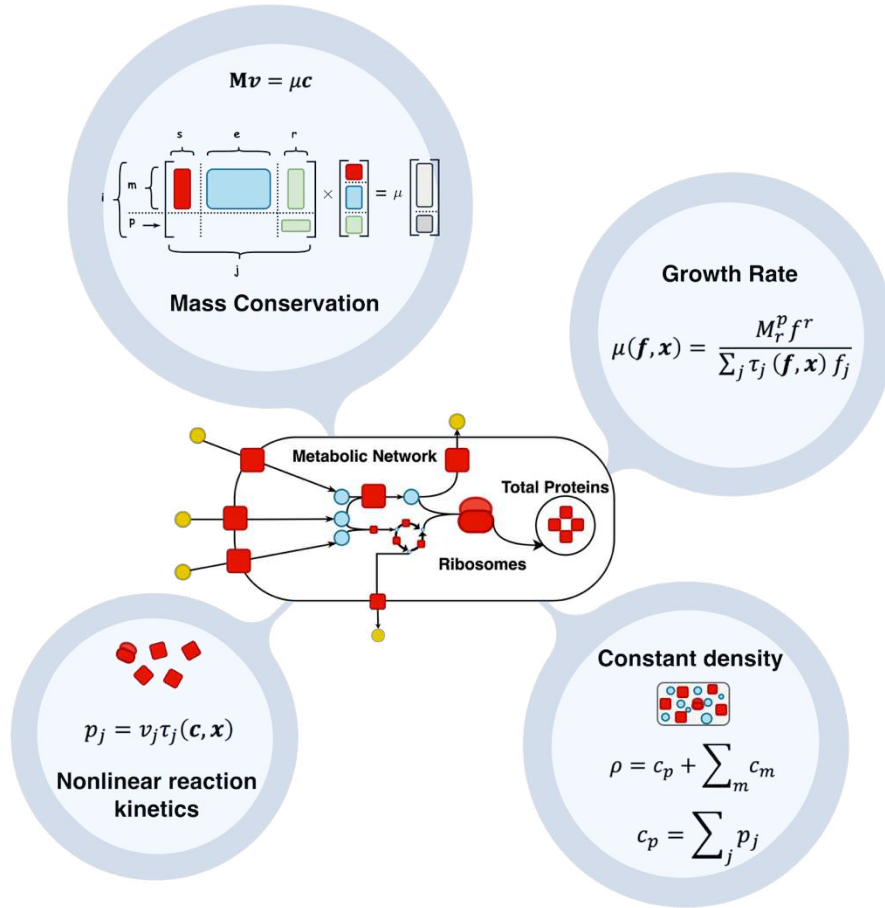


Figure 1. Schematic representation of the Growth Balance Analysis (GBA) framework (adapted from manuscript 1)

In manuscript 2, we present a proof-of-concept GBA model for the cyanobacterium *Synechocystis*, inspired by a previous self-replicator model (Faizi et al., 2018) and using identical parameter values to validate the approach against established work. We then extend the GBA framework to predict optimal proteome allocation at a more detailed level, demonstrating the ability of the GBA framework to move toward self-replicator models of complex cellular systems.

In manuscript 3, to increase the accessibility of Growth Balance Analysis (GBA) beyond the expert community, we describe *Cell Growth Simulator*, a user-friendly web platform built with R and Shiny. R is a widely used programming language and environment for statistical analysis and data visualization, while Shiny provides a lightweight framework for building interactive web applications. There are currently numerous web tools for linear metabolic modeling and visualization—such as CNApy (Thiele et al., 2022), Escher (King et al., 2015), ModelExplore (Martyushenko & Almaas, 2019), CAVE (Mao et al., 2023), and Fluxer (Hari & Lobo, 2020)—all of which use linear optimization approaches such as flux balance analysis (FBA) (Orth et al., 2010c). In addition, platforms like COPASI (Hoops et al., 2006) can handle constant catalytic rates (k_{cat}), which is consistent, *e.g.*, with Resource Balance Analysis (RBA) models (Goelzer & Fromion, 2011). However, no existing, easy-to-use web application facilitates the solution of cellular models with nonlinear kinetics. *Cell Growth Simulator* fills this gap by providing the first dedicated web-based solution for building,

simulating, and analyzing coarse-grained GBA models with up to ten reactions. This upper limit on model complexity is chosen deliberately to allow users to capture fundamental principles of cellular resource allocation and proteome efficiency with minimal computational cost (Doan et al., 2022; Dourado et al., 2023; Ghaffarinasab et al., 2023; Hui et al., 2015; Molenaar et al., 2009; Scott et al., 2010; Weiße et al., 2015).

1.10 Integrative Overview of the Three Studies

Manuscript 1 focused on the role of metabolic complexity in shaping the adaptive potential of bacteria, particularly of cyanobacteria. Using a pan-genome-scale metabolic model and flux balance analysis (FBA), we showed that species with larger metabolic networks (generalists) adapt more readily to novel nutrient environments than species with smaller metabolic networks (specialists). This finding supports the “toolbox model” (Maslov et al., 2009), in which a large existing set of enzymes (“tools”) provides multiple evolutionary stepping stones that facilitate rapid adaptation through horizontal gene transfer (HGT). Our results highlight a positive feedback loop between complexity and evolvability: organisms with large and branched metabolic networks benefit disproportionately from each gene acquisition, accelerating adaptation. These findings contrast with more traditional evolutionary models (e.g., Fisher's geometric model), which suggest that complexity may hinder adaptive evolution (Fisher, 1930). Instead, we show that metabolic complexity can be a facilitator of evolutionary diversification.

Building on this large-scale understanding of metabolic adaptability, manuscript 2 focused on the resource allocation strategies in the cyanobacteria *Synechocystis* sp. PCC 6803 using growth balance analysis (GBA). Linear constraint-based models, such as FBA, generally do not capture how enzymes and other cellular resources (e.g., proteins, metabolites) are quantitatively allocated in response to changing environmental conditions. By contrast, GBA explicitly accounts for metabolite concentrations and their influence on reaction fluxes through kinetic rate laws (Dourado et al., 2023; Dourado & Lercher, 2020). Compared to a previous coarse-grained nonlinear model of *Synechocystis* (Faizi et al., 2018), the GBA formulation proved to be mathematically simpler yet equally effective in reproducing observed growth trends and proteome allocation under different light intensities. This streamlined approach is particularly advantageous for exploring more sophisticated phototrophic behaviors and extending the model to include additional cellular processes, such as major photosynthetic complexes or carbon fixation. Both the base model – equivalent to the model of Faizi et al. (2018) – and our extended GBA model successfully reproduced experimental observations such as the average proteome fractions and growth rate-dependent trends (Zavřel et al., 2019). Specifically, as the growth rate shifts from light-limited to light-saturated conditions, the extended GBA model predicts a decrease in the proteome fractions of photosystem I (PSI) and photosystem II (PSII) along with an increase in ATP synthase and cytochrome b6, mirroring experimental observations (Faizi et al., 2018; Zavřel et al., 2019).

Finally, manuscript 3 addressed a practical challenge: the computational and technical barriers associated with nonlinear modeling of cellular systems. Although GBA offers a robust approach for modeling self-replicating cells – explicitly incorporating metabolite concentrations and their effects on reaction fluxes – its implementation traditionally demands

significant programming expertise and computational resources. To address these challenges, the *Cell Growth Simulator* was developed as a user-friendly web application designed for small-scale, self-replicating cell models. *Cell Growth Simulator* uses an intuitive spreadsheet interface, integrates kinetic parameter retrieval from the BRENDA enzyme database (Chang et al., 2021), and offers interactive visualization tools. This platform not only makes nonlinear modeling of resource allocation in coarse-grained cellular systems accessible to researchers with limited programming skills, but it also serves as a valuable tool for fostering interdisciplinary collaboration and enriching our understanding of cellular metabolism and growth.

Chapter 2

Metabolic Complexity and Its Role in Cyanobacterial Adaptation

C.J.F. constructed the pipeline that integrates the previously collected metabolic network reconstructions into a single pan-genome model. S.G. extended and updated this pipeline to incorporate newly constructed metabolic models. F.L. and C.J.F. collaborated on developing the accelerated algorithm for active reaction minimization. C.J.F. and S.G. conducted the adaptation calculations and performed data analysis. The initial manuscript draft and figures were prepared by C.J.F. and S.G., and the final version was completed by C.J.F., S.G., and M.J.L.

Manuscript status: In preparation for submission

Metabolic Complexity and Its Role in Cyanobacterial Adaptation

Claus Jonathan Fritzemeier^{1§}, Sajjad Ghaffarinasab^{1§}, Felix Lieder², Balázs Szappanos³, Florian Jarre², Balázs Papp³, Csaba Pál³, Martin J. Lercher^{1*}

¹ Institute for Computer Science and Department of Biology, Heinrich Heine University, Universitätsstraße 1, D-40225 Düsseldorf, Germany

² Institute for Mathematics, Heinrich Heine University, Universitätsstraße 1, D-40225 Düsseldorf, Germany

³ Synthetic and Systems Biology Unit, Institute of Biochemistry, Biological Research Centre of the Hungarian Academy of Sciences, Temesvári krt. 62, Szeged H-6726, Hungary

§ Shared first authors

* To whom correspondence should be addressed

Abstract

Cyanobacteria exemplify the remarkable ability of bacteria to adapt to new environments through horizontal gene transfer (HGT), a key driver of genome innovation and ecological success. However, anecdotal observations suggest that not all bacteria adapt equally: while certain species frequently diversify into new niches, others remain more specialized. To investigate the factors governing this variation, we used pan-genomic modeling to examine how 102 unicellular organisms, including several cyanobacterial strains, respond to over 5,000 different nutrient environments. Our analysis revealed that highly specialized endosymbionts typically require more than 50 additional metabolic reactions to establish growth in a novel environment, while generalist species such as *E. coli* require fewer than five. Strikingly, several cyanobacteria also exhibit higher adaptability, driven by their larger metabolic repertoires, which allow for more efficient evolutionary stepping stones via HGT. Thus, there is a positive feedback between metabolic complexity and evolvability, contrary to prior theoretical expectations that organismal complexity hinders adaptive evolution.

2.1 Introduction

Cyanobacteria have played a pivotal role in shaping Earth's biosphere by pioneering oxygenic photosynthesis roughly 2.5 billion years ago, fundamentally altering the planet's atmosphere and paving the way for the evolution of eukaryotes. Like many unicellular organisms, they exhibit an extraordinary capacity to adapt to diverse and changing environments (Brooks, Turkarslan, Beer, Lo, & Baliga, 2011). Different phylogenetic lineages differ widely in the frequency with which they give rise to new strains or even new species, but it is currently unclear what determines these differences. The splitting off of new lineages will often be adaptive, with the new lineage specializing to a different life style or environment. Among bacteria, such specialization is typically accompanied by the loss of now superfluous genes from the genome and the acquisition of additional genes via horizontal gene transfer (HGT, also termed lateral gene transfer) (Koonin & Wolf, 2008; Pal, Papp, & Lercher, 2005). While classic examples include the massive gene loss in the endosymbiotic bacterium *Buchnera* (Pal et al., 2006) or the gain of niche-specific pathways by pathogenic *Escherichia coli* strains (Alteri, Smith, & Mobley, 2009), cyanobacteria stand out for their evolutionary longevity and metabolic diversity – traits that have helped them persist in varied habitats, from open oceans to extreme terrestrial environments (Chen et al., 2021). As a consequence of these evolutionary dynamics, bacterial pan-genomes can be partitioned into core genes (found in almost all strains), shell genes (found in several strains), and cloud genes (restricted to a single strain) (Koonin & Wolf, 2008). Understanding how HGT can shape the remarkable adaptability of cyanobacteria offers insights not only into their evolutionary history, but also into broader questions about the mechanisms underpinning microbial specialization and innovation.

Bacterial strains of the same species often differ widely in their metabolic capabilities. For example, a study on *E. coli* found that individual strains could grow in between 437 and 624 of the tested environments (Monk et al., 2013). Based on such differences, lineages can be categorized as metabolic generalists or specialists. A prolonged reduction in environmental complexity – such as experienced by a generalist bacterium becoming a permanent endosymbiont – causes a corresponding reduction in metabolic complexity, which can be predicted quantitatively from genome-scale metabolic modeling (Pal et al., 2006). That bacterial evolution appears to organize itself into short bursts of innovation followed by long phases of genome reduction (Wolf & Koonin, 2013) indicates that the inverse process – a specialist evolving into a generalist – is comparatively rare.

In previous work (Szappanos et al., 2016), we utilized metabolic simulations to show that the standard lab strain *E. coli* K-12 can adapt to most previously unviable nutritional environments by acquiring at most three additional enzymes and/or transporters via HGT. In many cases, different new environments required the acquisition of overlapping gene sets. We found that complex metabolic innovations requiring multiple enzyme-coding genes can evolve via the successive acquisition of individual biochemical reactions, where each confers an additional benefit for the utilization of specific nutrients. This observation indicates an important role of exaptations in metabolic evolution, where stepwise metabolic niche expansion can lead to a substantial acceleration of adaptation (Szappanos et al., 2016).

However, multiple genes can also be acquired simultaneously via horizontal gene transfer. Successful transfer events of DNA in *E. coli* appear to co-transfer at most 30kb of DNA (Pang & Lercher, 2017). A reconstruction of the ancestral metabolic networks of 53 *E. coli* strains showed that all metabolic innovations identifiable *in silico* in this lineage indeed arose through the acquisition of a single DNA segment <30kb on one of the branches of the phylogeny. At the same time, around 10% of innovations relied on the exaptation of acquisitions on earlier branches of the strain phylogeny (Pang & Lercher, 2019).

These findings demonstrate that complex metabolic innovations can evolve – and have indeed evolved in *E. coli* – without the need to resort to neutral explorations of phenotype space. Such neutral explorations had been suggested earlier as an important facilitator of adaptation (Barve & Wagner, 2013), but the corresponding non-adaptive evolution of metabolic networks is expected to be extremely slow, and no direct empirical support has been identified for this scenario in bacteria (Szappanos et al., 2016). Thus, theoretical, computational, and comparative genomics considerations indicate that bacterial evolution of metabolic networks can be understood purely from a consideration of adaptive processes (Szappanos et al., 2016).

It has been suggested that complex organisms adapt more slowly than simple ones owing to increasing pleiotropic constraints (Fisher, 1930; Orr, 2005). Here, we argue that at least when metabolic networks are considered, the opposite may be true. More complex networks provide more raw material for adaptation to novel environments. Prompted by previous evidence for a broad adaptability of the generalist *E. coli* (Pang & Lercher, 2019; Szappanos et al., 2016) and simulations of abstract representations of artificial reaction networks (Maslov, Krishna, Pang, & Sneppen, 2009), we hypothesize that bacteria with more complex metabolic systems might be more adaptable than specialists.

Here, we explore this hypothesis in cyanobacteria along with other bacterial organisms by investigating how the size of metabolic networks affects the adaptability of their metabolic systems. Using pan-genome-scale metabolic simulations, we show that the ease with which microbes adapt to novel environments varies widely among species, with metabolic specialists typically requiring an order of magnitude more gene acquisitions than generalists adapting to the same environment. The increased adaptability of generalists is highlighted by their much higher potential for collateral adaptation, *i.e.* the ability to grow in additional, unselected environments due to ecologically unrelated previous adaptations. Specialist species, on the other hand, have largely lost their adaptive potential. When they do adapt, however, they show a stronger tendency toward exaptation, *i.e.*, they are more likely to re-use previously acquired enzymes and transporters for subsequent adaptations.

2.2 Results and Discussion

Construction of a pan-genome scale metabolic supermodel from organism-specific models

To allow coherent simulations of metabolic network expansion through HGT, we first created a pan genome-scale metabolic supermodel that contained all examined organism-specific metabolic networks as submodels. The supermodel built from 102 organismal metabolic models contains 16,018 unique reactions and 7,551 unique metabolites. **Fig. 1** shows the sizes of the organism submodels included. Most metabolites are assigned to the compartments

extracellular (e), periplasm (p), and cytosol (c) (**Suppl. Fig. S2**). Several additional compartments in the supermodel originate from the contributions of four eukaryotic organisms (*Chlamydomonas* (iRC1080), *Saccharomyces cerevisiae* (iMM904, iND750), *Phaeodactylum tricornutum* (iLB1027_lipid)) and the cyanobacterium *Synechocystis* sp. PCC 6803 (iJN678). For the well-studied *Escherichia coli* str. K-12 substr. MG1655, five different models were included. We also included metabolic models for 55 other *E. coli* and *Shigella* strains (Monk et al., 2013). Further details about the organisms and metabolic models included are listed in **Suppl. Table S1**.

Branching points in metabolic networks occur when multiple reactions produce and/or consume the same metabolite. One function of such branching points is to link alternative pathways to central metabolism. As expected, we find that larger metabolic networks tend to be less linear, *i.e.*, they contain a lower proportion of metabolites that are consumed and produced by less than three reactions (Spearman's $\rho=-0.42$, $P=0.04$; **Suppl. Fig. S3**).

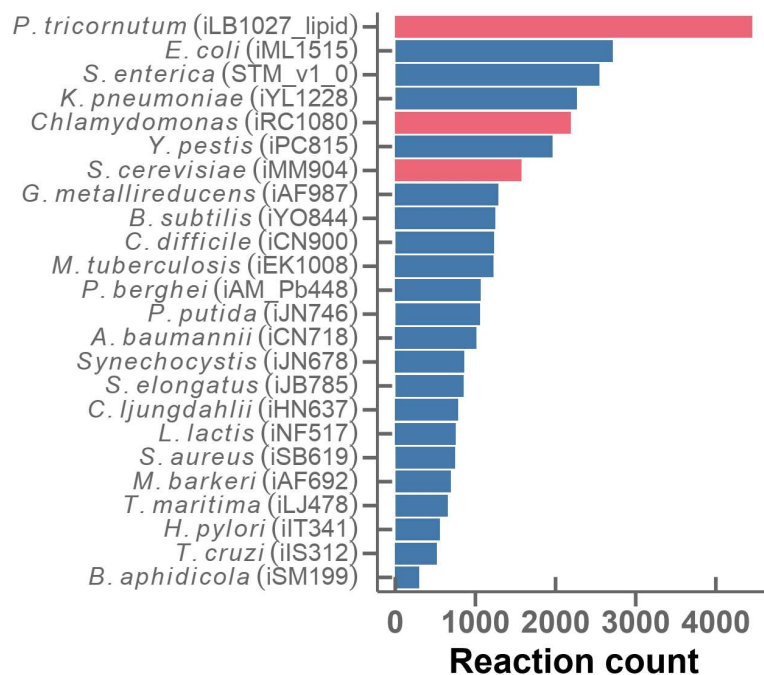


Figure 1. The models included represent a broad range of metabolic complexity.

The bars show the number of metabolic reactions for each model that contributed to the pan-genome-scale supermodel. Only one representative strain and model is shown for each species. Red bars indicate eukaryotic models. Suppl. Fig. S1 shows the corresponding information for additional strains and models.

We used flux balance analysis (FBA) (Orth, Thiele, & Palsson, 2010; Watson, 1984) to estimate the ability of each submodel to grow in each of a large number of nutritional environments. To make the results comparable, we used the same general biomass reaction for all organism-specific submodels, *i.e.*, each metabolic system was required to produce the same metabolic precursors for cellular growth (Methods). We examined two sets of nutritional environments: one set that largely contains typical wet lab growth media (Henry et

al., 2010), including those assayed in the Biolog phenotyping system; and another set of random minimal media, each comprising a combination of carbon, nitrogen, sulfur, and phosphorus sources plus trace elements.

As most models cannot grow in any of the random minimal environments, we checked whether all models can grow in a medium that supplies all possible nutrients. Only three models are not viable in this maximally rich condition: the hyperthermophilic bacterium *Thermotoga maritima* (iLJ478), the parasitic protozoon *Trypanosoma cruzi* Dm28c (iLS312), and the endosymbiotic bacterium *Buchnera aphidicola* (iSM199). This is because the general biomass objective function contains more amino acids than the original biomass functions of these models. Therefore, we included the missing essential reactions (1, 13, and 5 reactions, respectively; **Suppl. Table S2**) to enable these models to grow on this fully rich medium. We chose not to exclude these models from further analyses, as the ability of extreme specialists to adapt to new environments is one of the questions we aim to explore.

As shown in **Fig. 2**, the minimal random environments are too restricted for most modeled organisms and hence provide limited insights into the growth of the submodels in the real world. In contrast, almost all submodels can grow in at least some of the wet lab environments, with the most versatile model – *E. coli* – growing in 36% of wet lab media, while the cyanobacterium *Synechocystis* sp. PCC 6803 achieves growth in 8% of these conditions (**Fig. 2**). The distribution of the fraction of viable wet lab environments across submodels is bimodal (**Suppl. Fig. S4**), naturally dividing these organisms into generalists and specialists; we set the dividing line at growth in 20% of assayed media. As expected, the same three organisms unable to grow in the full medium are also unable to grow in any wet lab environment. To guard against any biases introduced by the general biomass function, we repeated this analysis with using the generation of energy (conversion of ADP to ATP) as the objective function, with qualitatively similar results (**Suppl. Fig. S5**).

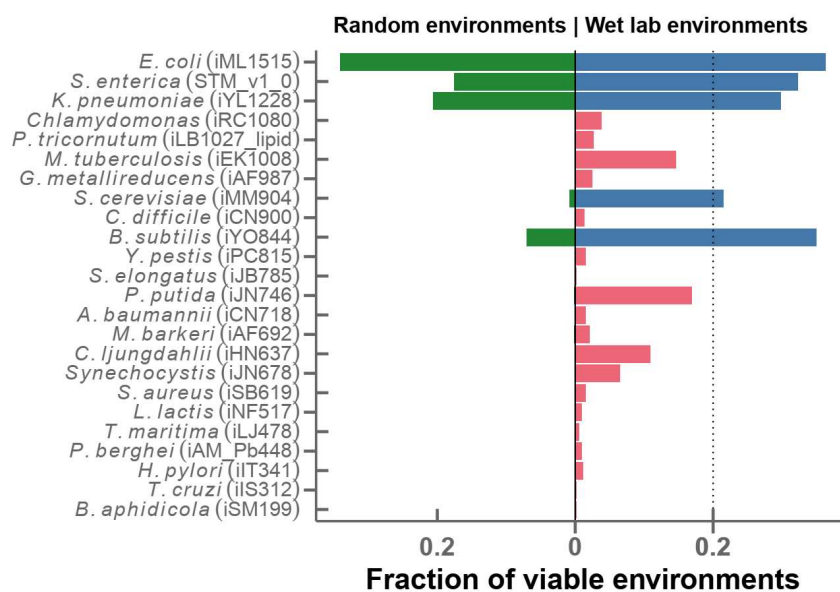


Figure 2. The fraction of viable environments differs widely across submodels, both for random minimal environments (green bars to the left) and for common wet lab environments

(blue and red bars to the right). The dotted vertical line at 0.2 indicates the threshold for partitioning metabolic systems into generalists (blue) and specialists (red). Models are ordered top to bottom by decreasing genome size. **Suppl. Fig S6** shows the corresponding results when using energy production instead of biomass production as the objective function.

More complex networks are more adaptable

We next quantify the difficulty for an organism to adapt to new environments. For each submodel and each environment in which it is currently unable to grow, we identified the minimal number of reactions that have to be added to produce biomass; below, we refer to this number as the *added reactions*. The distribution of added reactions per wet lab environment varies widely across organisms (**Fig. 3a**, including only one representative for *E. coli*). Results are quantitatively similar when considering random instead of wet lab environments (**Suppl. Fig. S6**), and qualitatively similar when using energy generation instead of biomass production as the objective function (**Suppl. Fig. S7**).

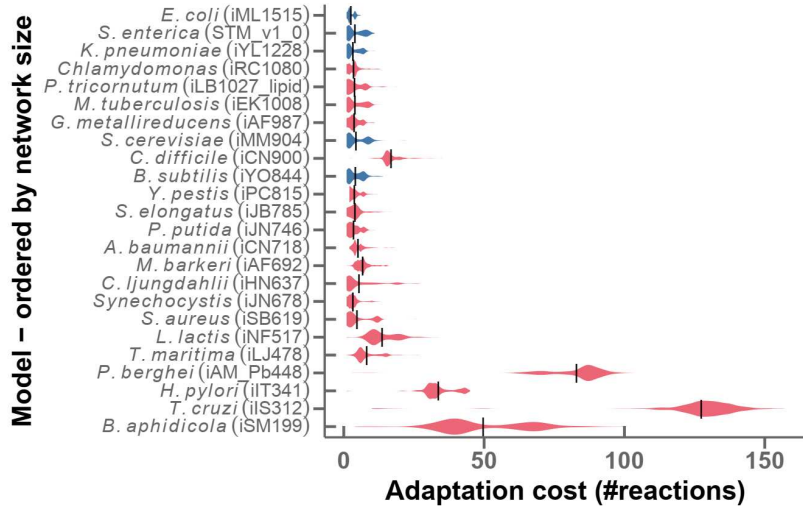
The four smallest and most specialized metabolic networks require the largest number of added reactions to adapt to new environments. The endosymbiont *Buchnera aphidicola* needs to add on average 51.7 reactions to reach new environments. Similarly, the pathogen *Helicobacter pylori*, which exclusively lives in human stomachs, needs on average 33.7 additional reactions. *Plasmodium berghei*, which is a protozoan parasite that causes malaria in rodents, requires on average 83.0 reactions to be viable in a new environment. Finally, the parasite *Trypanosoma cruzi* Dm28c requires on average 129.1 reactions. All four organisms are highly specialized to one or a few specific, stable environments. Accordingly, their metabolisms show very little flexibility, reflected in very small numbers of metabolic genes (*B. aphidicola*: 199 metabolic genes out of a total of 517 genes (Shigenobu, Watanabe, Hattori, Sakaki, & Ishikawa, 2000); *H. pylori*: 341 metabolic genes out of 1590 total genes (Tomb et al., 1997); *T. cruzi*: 312 metabolic genes out of 1430 (De Pablos & Osuna, 2012); *P. berghei*: 448 metabolic genes out of 5216 (Otto et al., 2014)). At the other end of the spectrum in **Fig. 3a** is *E. coli*: the standard lab strain K12 (iML1515) requires on average 2.55 and at most 6 reactions to adapt to any of the tested environments. The cyanobacteria *Synechocystis* sp. PCC 6803 and *Synechococcus elongatus* PCC 7942 require similarly low numbers of extra reactions on average – 3.27 and 3.99, respectively.

Although it is likely that many properties influence the ability of a metabolic system to adapt to new nutritional environments, network size alone explains 56% of the variance across all assayed models (**Fig. 3b**; Spearman's $\rho = -0.75$, $P = 3.57 \times 10^{-5}$). The solid line in **Fig. 3b** shows the best-fitting power law, $\text{added reactions} = a \times (\text{gene count})^b$. The best-fitting exponent is $b = 2.87$ (95% CI = [3.47, 2.28]), which is slightly larger than the $b = 2$ expected from abstract models of metabolic network expansion (Maslov et al., 2009). The two cyanobacterial strains, *Synechocystis* sp. PCC 6803 and *Synechococcus elongatus* PCC 7942, fall below the fitted curve in **Fig. 3b**. In other words, they exhibit negative residuals; they require fewer added reactions than the power law predicts based on their gene counts. This deviation is consistent with their photoautotrophic physiology. External energy from light and reliance on CO₂ fixation lessen the need to recruit new catabolic modules when the medium changes. Thus, adaptation mainly hinges on a small set of transport and cofactor steps rather than on the addition of entire pathways.

The dataset contains *E. coli* models of various sizes, with between 904 and 1,516 metabolic genes. These mostly differ only marginally in their adaptability (**Suppl. Fig. S8**): the average number of added reactions for generalist *E. coli* (including nine strains with auxotrophies) lies between 2.40 and 2.97, while the average number of added reactions for specialist *E. coli* ranges from 2.30 to 5.22. The outlier requiring the largest number of additional reactions is *E. coli* DH1 (iEcDH1_1363; **Suppl. Fig. S8**), which is auxotrophic for thiamine (Meselson & Yuan, 1968) due to the loss of a complete operon (Monk et al., 2013). Similar to the picture across species (**Fig. 3b**), and despite the low variation in adaptability, we find a substantial negative correlation between the average number of added reactions and network size when comparing different *E. coli* strains (Spearman's $\rho = -0.60$, $P = 6.3 \times 10^{-6}$, excluding strains with auxotrophies).

Do different metabolic subsystems contribute unequally to the adaptations observed in the simulations? To answer this question, we utilized the high-quality subsystem annotations available for the 55 strain-specific *E. coli* models. We repeated the analysis of environment-specific adaptation for each *E. coli* strain, but restricting gene acquisitions to genes from other *E. coli* strains. For this purpose, we created a second pan-genome-scale metabolic supermodel exclusively for the 55 *E. coli* strains (see Methods, “Analysis of *E. coli* metabolic subsystems involved in adaptations”), containing 1,644 unique metabolites and 2,493 unique reactions from a total of 69 metabolic subsystems. Our simulations predicted the involvement of genes from 27 subsystems in environment-specific adaptations (**Suppl. Table S3**). 19 of these subsystems coincided with those observed in a comparative genomics analysis of adaptive gene acquisitions across *E. coli* strains (Pang & Lercher, 2019) ($P = 0.026$, $OR = 3.43$, Fisher's exact test; **Suppl. Table S3**). However, the odds ratios for observing gene acquisitions from a specific subsystem (compared to all other subsystems) differ markedly between our simulations and those inferred from comparative genomics (**Suppl. Table S3**). This discrepancy indicates that the simulated nutritional environment changes may not be directly representative of those experienced by real *E. coli* strains, possibly because *E. coli* strains tend to transition between relatively similar, host-associated environments.

a



b

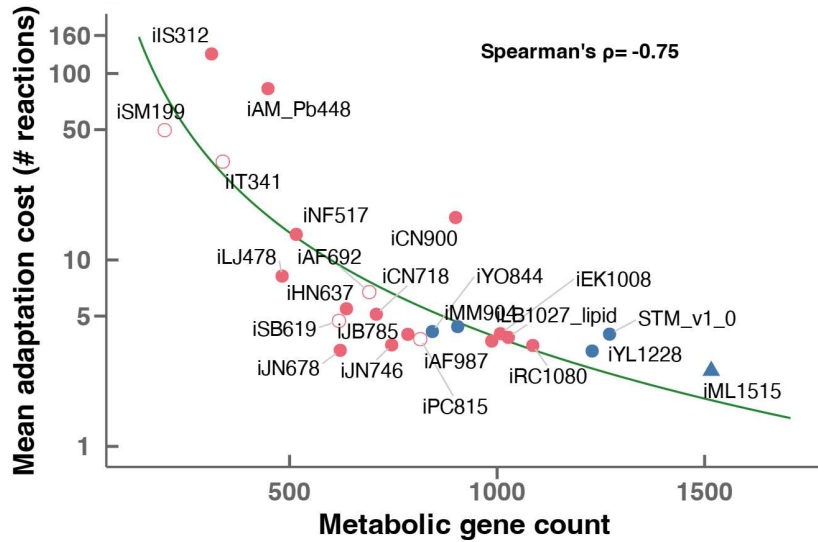


Figure 3. The number of additional reactions required for adaptation decreases with increasing genome size. (a) Distributions of added reactions, summarized as violin plots. The height at each point in a “violin” indicates the local density of the distribution for the given model. Models are ordered top-down by decreasing size. (b) The average number of added reactions (log scale) plotted against metabolic gene count for each model. The green line shows the best fitting power law, $added\ reactions = a \times (gene\ count)^b$. In both panels, colors distinguish specialists (red) and generalists (blue). Organisms with known auxotrophies are shown as open circles. The 55 *E. coli* strains are represented by the iML1515 model (blue triangle) only. For this figure, only wet lab environments are considered.

Specialists often re-use gained reactions in later adaptations

If an organism adapts to a given environment by acquiring a matching set of metabolic reactions, it can happen that the same reactions now also facilitate growth in another environment, where the organism was unviable before. With few exceptions, such collateral,

unselected-for adaptation happens more frequently for generalists than for specialists across species; this trend is reversed when comparing different *E. coli* strains, possibly because the repair of auxotrophies facilitates growth in multiple environments (**Suppl. Fig. S9**; see also Refs. (Barve & Wagner, 2013; Hosseini & Wagner, 2016)).

But even if the reactions acquired to adapt to environment *A* do not provide immediate access to environment *B*, they may still provide a subset of the reactions required to adapt later to this second environment. We quantify the propensity to profit from adaptations in this way with an *exaptation index* (see Methods). One might hypothesize that while specialists show little collateral adaptation, they may show a high potential for step-wise exaptation, for example if added reactions remove an auxotrophy. As expected from this hypothesis, **Suppl. Fig. S10** shows that we indeed tend to find higher exaptation indices for specialists than for generalists; moreover, the propensity for such exaptations is higher for specialists with small genomes compared to specialists with larger genomes. The cyanobacteria *Synechocystis* sp. PCC 6803 and *Synechococcus elongatus* PCC 7942 behave more like generalists. The few reactions they acquire to enter one medium are often reused in others through collateral adaptation, while they rely modestly on stepwise exaptation.

2.3 Conclusions

Adaptations arise by extensions of existing phenotypes and genotypes. In specialists with small genomes, adaptation to new ecological niches is typically difficult, as it demands the simultaneous acquisition of multiple mutations or genes. As a consequence, specialists with simple genomes may often be evolutionary dead-ends. The smallest and most specialized metabolic systems, those of *Buchnera aphidicola* (an endosymbiont of aphids), *Trypanosoma cruzi* Dm28c (an internal human pathogen), and *Helicobacter pylori* (an endosymbiont of the human stomach), are trapped in their endosymbiotic life style, having all but lost their adaptive potential. The opposite is true for organisms with complex genomes – such as *Synechocystis* or *E. coli* – whose larger “toolboxes” (Maslov et al., 2009) can more easily be extended for novel tasks.

The observed relationship between metabolic network size and adaptability leads to a positive feedback between complexity and evolvability. This conclusion is the exact opposite of what is suggested by Fisher’s geometric model (Fisher, 1930; Orr, 2005). Fisher’s model supports the idea that more complex systems are less likely to adapt through natural selection.

Specifically, it has been argued that a mutation of a given size is less likely to be favorable in complex than in simple organisms because it affects many phenotypic traits simultaneously in the former. In support of this idea, it has been observed that genes encoding proteins involved in many protein-protein interactions are less likely to be horizontally transferred than genes encoding less highly-connected proteins (Cohen, Gophna, & Pupko, 2011; Jain, Rivera, & Lake, 1999). This effect might be expected, as the interaction between two proteins requires the co-evolution of the amino acid sequences directly involved in the contact, and hence a protein encoded by a newly acquired gene may not bind sufficiently strongly to existing proteins of the host. Conversely, different enzymes that interact in a metabolic network perform their molecular functions independently, and their amino acid sequences do not need to be finetuned with respect to each other. This line of argument suggests that metabolic genes

with high connectivity may be integrated easily into an existing network, while genes with high connectivity in the protein-protein interaction network may not.

However, while the amino acid sequences of different enzymes may be independent of each other, their expression has to be coordinated precisely. Thus, finetuning is necessary also for the integration of metabolic genes into an existing network, although the adjustments must occur in terms of regulatory changes rather than amino acid sequence changes. That metabolic complexity and protein-protein interaction complexity appear to have opposite effects on adaptability might then be explained by faster adaptive evolution of gene expression compared to functional gene properties (Lenski, 2017; Lozada-Chavez, Janga, & Collado-Vides, 2006).

Exaptation – the utilization of metabolic genes acquired in previous adaptations for adaptive purposes in a new environment – plays an important role in the adaptation of both generalists and specialists, although in different ways. Generalist species, but not specialist species, show a high degree of collateral adaptation, *i.e.*, previous adaptations often enable growth in environments other than those experienced by the organisms’ ancestors (Barve & Wagner, 2013). Conversely, specialist species that acquire new metabolic genes in the adaptation to one environment are more likely to re-use (exapt) these genes in later adaptations to other environments; thus, stepwise metabolic niche expansion will play an even stronger role in the adaptation of specialists than previously observed for the generalist *E. coli* (Szappanos et al., 2016), and might thus be the facilitator of rare genome expansions (Koonin & Wolf, 2008).

2.4 Materials and Methods

Supermodel generation

We started with 109 genome scale models (GSMs) downloaded from the BiGG database (Schellenberger, Park, Conrad, & Palsson, 2010). We removed seven models of multicellular eukaryotes. As we are specifically interested in variations in metabolic model size and as the BiGG database contains only few species with very small metabolic systems, we added the model for *Buchnera aphidicola* str. APS (Macdonald, Lin, Russell, Thomas, & Douglas, 2012). Thus, 102 GSMs (termed “submodels” in this work) contributed to the supermodel (**Suppl. Table S1**). As a preprocessing step, we checked whether reactions and metabolites from different submodels but with the same IDs represented the same biochemical reaction; if not, we renamed them. Reactions were compared on the basis of stoichiometry and reversibility, while metabolites were compared on the basis of their chemical formulas if these were available.

A preliminary supermodel was formed as the union of the reactions and metabolites from all submodels. As detailed below, we then curated this preliminary model by ensuring mass balance and by removing energy generating cycles (EGCs) (Fritzemeier, Hartleb, Szappanos, Papp, & Lercher, 2017). While each individual model passes these quality checks, the reactions in the merged supermodel may be combined in ways that violate thermodynamic laws or the mass balance. Mass balance is considered first, because proper mass balance is a requirement for the EGC removal. The final supermodel was thus mass-balanced and had no energy-generating cycles.

Correction of mass balance

Mass balance of a reaction is generally ensured by contrasting all atoms of the educts and all atoms of the products. However, due to incomplete data, the mass balance for many reactions is not known; removing all reactions with uncertain mass balance would render most of the models non-functional. To circumvent this problem, only the mass balance of the exchange reactions was considered: the number of atoms of the same kind (*e.g.*, carbon) entering the model has to equal the number of corresponding atoms leaving the model. The only reactions that allow exchange of molecules with the model environment are exchange reactions and biomass reactions. At the same time, these are the only reactions in a network that are allowed to be imbalanced. We first removed exchange reactions and biomass objective functions that contain a metabolite of unknown composition from the model, as for these we cannot guarantee mass balance. To identify potentially imbalanced reactions, we fixed the net exchange of atoms to zero. We then removed all reactions that are blocked in this situation.

Removing erroneous energy-generating cycles

Another problem occurring when combining multiple GSMs is the formation of erroneous energy generating cycles (EGCs) (Fritzemeier et al., 2017; Szappanos et al., 2016). In GSMs, such thermodynamically impossible cycles can produce energy equivalents (*e.g.*, by synthesizing ATP) in infinite amounts without the consumption of nutrients (Fritzemeier et al., 2017). Thermodynamics are strongly influenced by metabolite concentrations. However, GSMs consider thermodynamics only approximately through the directionality of reactions. Thus, combining two networks can cause the formation of EGCs even if the individual networks are EGC-free.

Based on a previously published algorithm (Fritzemeier et al., 2017), we constructed a greedy approach to build organism-specific EGC-free supermodels. We chose not to build one supermodel for all analyses, as the order of adding metabolic networks to the growing supermodel can affect the final model, and as we wanted to study the adaptability of each organism starting from a model from which none (or only a few) reactions had been removed.

From the preliminary, mass-balanced supermodel, we first considered the set of reactions of the focal organism and removed any EGCs present. We then iteratively added the remaining submodels, each time removing all EGCs before proceeding to the next one. The order of adding organisms was determined by the initial number of EGCs; models with fewer EGCs were always added first.

To remove EGCs, we first determined the smallest set of reactions capable of producing energy equivalents in the model. This problem was solved in previous work with the ARM MILP algorithm, but here we instead used the ARM LP algorithm (see Methods, “Active reaction minimization”). We randomly chose one reaction in this cycle; we deleted the reaction if it was irreversible, and constrained it to be irreversible in the opposite direction if it was reversible. We repeated this process until no more EGCs were present. This procedure resulted in one mass-balanced, EGC-free supermodel for each organism-specific model in our dataset.

Active reaction minimization

Mixed integer linear programs (MILP) are frequently used to extend FBA, *e.g.*, in ROOM (Satish Kumar, Dasika, & Maranas, 2007), gapfind, and gapfill (Shlomi, Berkman, & Ruppin, 2005). In many of these problems, the objective is active reaction minimization (ARM). The pan-genome-scale model in this work is much bigger than any genome-scale models. Current methods of minimizing the number of active reactions under flux balance constraints cannot be applied due to the exponential complexity of this problem. We here use an approximate method that leads to major speedups and minor inaccuracies. A corresponding linear approximation has also been used in combination with the Gapfill algorithm (Thiele, Vlassis, & Fleming, 2014).

We relax the following ARM MILP problem into a sequence of ARM LP^k for $k \in \{1, \dots, n\}$ problems. We use the property of the simplex algorithm to find sparse solution vectors.

ARM MILP:

$$\begin{aligned} & \min \left(\sum_{i \in B} b_i \right) \\ & s. t. : \\ & S * v = 0 \\ & l_i \leq v_i \leq u_i \quad \forall i \in R \\ & v_i \neq 0 \Rightarrow b_i = 1 \quad \forall i \in B \end{aligned}$$

ARM LP^k:

$$\begin{aligned} & \min \left(\sum_{i \in B} \left| v_i * \frac{1}{\max(\varepsilon_i^k, |v_i^{k-1}|)} \right| \right) \\ & s. t. : \\ & S * v = 0 \\ & l_i \leq v_i \leq u_i \quad \forall i \in R \end{aligned}$$

Table 1. Definition of variables of the ARM LP.

$S \in \mathbb{R}^{ M } \times \mathbb{R}^{ R }$	stoichiometric matrix
$v \in \mathbb{R}^{ R }$	Vector of fluxes
$l \in \mathbb{R}^{ R }$	Vector of lower bounds
$u \in \mathbb{R}^{ R }$	Vector of upper bounds
$b \in \{0,1\}^{ B }$	Vector of binary variables
$R \in \{1, \dots, m\}$	Set of m reaction indices
$B \subseteq R$	Set of indices that are objective of the optimization
v_i^k	Flux of the i -th reaction in the k -th optimization (0 if undefined)
k	Optimization step counter

n	Total number of optimization steps
ε_i^k	i -th upper bound of weight factor in optimization round k

In this sequence of linear problems, the optimization function of the $(k + 1)$ -th problem is reweighted with the solution of the k -th problem. The initial values for ε_i^0 are either set to one or to some positive random values. For the $(k + 1)$ -th optimization, we recalculated $\varepsilon_i^{k+1} := \varepsilon_i^k * 1/10$.

In order to show the practical application of our linear approximation of active reaction minimization, we show the comparison between the MILP result and the LP approximation. To limit the computation time to a reasonable span, we allowed the solver for the MILP eight parallel threads per problem and a maximum time of two minutes per problem. Thus, some results are suboptimal, but the gap value accounts for the maximal possible difference to the optimal value. For the ARM LP calculations, the linear problem was solved twelve times and the best solution was kept. After every fourth optimization, ε was reinitialized with random values and v is set as undefined.

Suppl. Fig. S11a shows both results in direct comparison for a total of 2830 problems we solved with the *E. coli* model iAF1260 and the standard biomass reaction. The ARM LP performs better for some problems with the non-optimal MILP solutions, *i.e.*, with a gap greater zero. This is also the case for the exact solutions. We suspect the solver to have some numerical issues and thus to give a non-optimal solution in four MILP cases. All results were successfully verified with FBA. The differences between the pairwise results are shown in **Suppl. Fig. S11b**. For over 50% (1587 of 2830) of the problems, the ARM LP found a better or equally good optimal value. For 75% of the problems, the ARM LP solution differed by at most two reactions from the MILP ARM solution. Without a pre-specified time limit, the MILP ARM computation times vary widely (from seconds to hours), while the ARM LP problems are always solved in split seconds.

Adaptation cost

The adaptation cost is the minimal number of reactions an organism has to obtain in order to survive in an environment that did not support growth beforehand. The environment is defined as the set of nutrients available for growth, and viability is defined as the ability to produce biomass at a rate above $0.01 \text{ mmol gDW}^{-1} \text{ h}^{-1}$. This calculation depends on two major factors: the definition of the environments, *i.e.*, the growth media, and the choice of the biomass objective function for a model.

Sets of environments

All molecule types that can be taken up by the supermodel are potential nutrients. Environments differ by which of these potential nutrients are present. We analyzed two sets of environments. The first set (“wet lab media”) is taken from the Seed database (Henry et al., 2010) and represents wet lab growth media. The environments in the second set (“random media”) are derived from a minimal growth medium for the *E. coli* model iAF1260. Each of these environments consists of one carbon, one nitrogen, one sulfur, and one phosphorous

(CNPS) source, accompanied by trace elements essential for growth (Szappanos et al., 2016). The complete set is generated by randomly choosing 5000 such combinations.

Biomass objective functions

Each model includes the definition of at least one biomass reaction. For the further analyses, we selected one of these. The biomass reactions of 16 models were blocked in the mass balance step. These models did not have a valid biomass reaction anymore and were excluded from the analyses that were based on the model-specific biomass functions. To make the submodels comparable, we defined a general biomass reaction (based on the iAF1260 biomass reaction) that contains only a set of core metabolites shared by all organisms (ribose nucleotides; deoxyribose nucleotides; amino acids; water) and an energy dissipation term (converting ATP to ADP + Phosphate + H⁺). However, for a subset of three models that were unable to grow on maximally rich medium using this general objective function, we performed the adaptation cost analysis to identify essential reactions. Subsequently, we supplemented these models with the essential reactions to allow them to grow on maximally rich medium. In addition, we also considered a biomass reaction that consists only of the energy dissipation term, thus indicating if a model is able to produce energy from the nutrients. **Suppl. Fig. S12** shows the percentages of environments (wet lab or random) in which the individual submodels can produce a non-zero flux through the different biomass reactions.

Calculation of adaptation cost with ARM LP

The mass-balanced and EGC-free, organism-specific model is formally a submodel of the organism-specific supermodel (see above). For a given environment and both the submodel and the supermodel, we used standard FBA to test if biomass can be produced above the threshold of 0.01 mmol gDW⁻¹ h⁻¹ (“growth”). Adaptation cost were calculated for environments that support growth of the supermodel but not of the submodel. For each such environment, we used ARM LP to estimate the minimal number of reactions that have to be added from the supermodel to the submodel to facilitate growth. This procedure was performed for each combination of organism-specific model, environment, and biomass reaction.

Analysis of *E. coli* metabolic subsystems involved in adaptations

To examine the contribution of specific metabolic subsystems to adaptability, we constructed a second pan-genome-scale metabolic supermodel restricted to the 55 *E. coli* strains. This reduced supermodel was built using the same procedure described above, ensuring that simulated gene acquisitions were limited to genes present in other *E. coli* strains. For each strain-specific submodel, we calculated the adaptation cost in the SEED wet-lab media, using the model-specific biomass objective function. Each acquired gene was assigned to a metabolic subsystem based on the consistent subsystem annotations provided with the models.

For each gene, we computed its gain frequency per simulation as the total number of acquisitions across all simulations, divided by the number of model-environment combinations in which the gene was both absent from the model and required for growth (*i.e.*, where the model was non-viable prior to gene acquisition). Aggregating across all models and

environments, we then calculated the mean gain frequency for each subsystem. The odds ratio for a given subsystem was defined as the mean gain frequency within the subsystem, divided by the mean gain frequency across all other subsystems. This odds ratio quantifies whether genes from a particular subsystem were disproportionately acquired during simulated adaptations.

We then calculated an empirical P -value to test if gene acquisitions from a given subsystem were statistically significantly higher than expected under the null hypothesis that genes from all subsystems were equally likely to be gained. In each of $N=100,000$ randomizations, we randomly permuted the subsystem labels of all genes and recalculated the odds ratio. For each subsystem, the empirical P -value was estimated as the fraction of simulations where the odds ratio from the random assignment was equal or higher than the odds ratio observed in the original simulations.

To compare our simulation results to metabolic gene acquisitions inferred from comparative genomics of sequenced *E. coli* strains, we used data from the study by Pang et al. (Pang & Lercher), which inferred gene acquisitions along the branches of a phylogenetic tree of 55 *E. coli* strains. We calculated odds ratios and empirical P -values as described above.

Collateral Adaptation Index

We defined a collateral adaptation index to quantify the probability that adaptation to one environment would lead to a “collateral” adaptation to other, unselected-for environments. For each submodel, we first identified the n random environments in which it cannot produce biomass (unviable environments). For each of these environments in turn, we identified the smallest set of reactions from the supermodel that have to be added to enable biomass production; these reactions define the adaptation cost. We then determined in how many of the $n-1$ remaining previously unviable environments this extended model can grow. If we denote this number m , then the collateral adaptation index is defined as the corresponding fraction, $m / (n-1)$. Thus, an index of 1 indicates collateral adaptation to all environments, while an index of 0 indicates no collateral adaptation. This definition was similarly described elsewhere (Barve & Wagner, 2013). To make sure that the adaptations considered are indeed collateral and not selected in the initial environment, we considered only environments that had no overlap with the source environment, *i.e.* none of the carbon, nitrogen, sulfur or phosphate source from the adapted environment was contained in the tested environments.

Exaptation Index

Assume that to adapt to grow in a new environment m_1 , an organism needs the additional reaction set r_1 . To grow in a second distinct environment m_2 , the same organism may need the reaction set r_2 . The fraction of preadapted reactions can be defined as $f_{m_1, m_2} := \frac{|r_1 \cap r_2|}{|r_2|}$.

We define the exaptation index as

$$e_{m_j} := \text{mean} \left(f_{m_i, m_j} \right) ,$$

Where the mean is calculated across all $m_i \in M$, and M is the set of environments distinct from m_j in which the un-adapted submodel was unviable. For example, an exaptation index $e_{m_j} = 0.5$ means that on average, the organism already acquired half of the reactions needed

to adapt to further environments, while an exaptation index of $e_{m_j} = 1$ indicates collateral adaptation.

Hardware, Software

All calculations were computed with the constraint-based modelling package “sybil” in GNU R, using IBM ILOG CPLEX as the solver. Calculations were done on a compute cluster with a peak usage of about 600 CPUs. The whole process is implemented as a pipeline reducing human interaction to a minimum. Frequent control points ensure data integrity and correctness of calculations. The code used in our simulations, as well as the corresponding results, are available on a GitLab repository at the following link: <https://github.com/Sijr73/Supermodel>

Acknowledgements

We are grateful for computational support through the Zentrum für Informations- und Medientechnologie (ZIM) at Heinrich Heine University Düsseldorf. We gratefully acknowledge financial support by the German Research Foundation through DFG grants IRTG 1515 supporting CJF; CRC 680, CRC 1310 and, under Germany’s Excellence Strategy, grant EXC 2048/1 (Project ID:390686111) to MJL; as well as through an ENORM graduate school fellowship to CJF.

2.5 References

- Alteri, C. J., Smith, S. N., & Mobley, H. L. (2009). Fitness of *Escherichia coli* during urinary tract infection requires gluconeogenesis and the TCA cycle. *PLoS Pathog*, 5(5), e1000448. doi:10.1371/journal.ppat.1000448
- Barve, A., & Wagner, A. (2013). A latent capacity for evolutionary innovation through exaptation in metabolic systems. *Nature*, 500(7461), 203-206. doi:10.1038/nature12301
- Brooks, A. N., Turkarslan, S., Beer, K. D., Lo, F. Y., & Baliga, N. S. (2011). Adaptation of cells to new environments. *Wiley Interdiscip Rev Syst Biol Med*, 3(5), 544-561. doi:10.1002/wsbm.136
- Cohen, O., Gophna, U., & Pupko, T. (2011). The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. *Mol Biol Evol*, 28(4), 1481-1489. doi:10.1093/molbev/msq333
- Chen, M. Y., Teng, W. K., Zhao, L., Hu, C. X., Zhou, Y. K., Han, B. P., ... & Shu, W. S. (2021). Comparative genomics reveals insights into cyanobacterial evolution and habitat adaptation. *The ISME Journal*, 15(1), 211-227. doi:10.1038/s41396-020-00775-z
- De Pablos, L. M., & Osuna, A. (2012). Multigene families in *Trypanosoma cruzi* and their role in infectivity. *Infect Immun*, 80(7), 2258-2264. doi:10.1128/IAI.06225-11
- Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford,: The Clarendon press.
- Fritzemeier, C. J., Hartleb, D., Szappanos, B., Papp, B., & Lercher, M. J. (2017). Erroneous energy-generating cycles in published genome scale metabolic networks: Identification and removal. *PLoS Comput Biol*, 13(4), e1005494. doi:10.1371/journal.pcbi.1005494
- Henry, C. S., DeJongh, M., Best, A. A., Frybarger, P. M., Lindsay, B., & Stevens, R. L. (2010). High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol*, 28(9), 977-982. doi:10.1038/nbt.1672
- Hosseini, S. R., & Wagner, A. (2016). The potential for non-adaptive origins of evolutionary innovations in central carbon metabolism. *BMC Syst Biol*, 10(1), 97. doi:10.1186/s12918-016-0343-7
- Jain, R., Rivera, M. C., & Lake, J. A. (1999). Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci U S A*, 96(7), 3801-3806. doi:10.1073/pnas.96.7.3801
- Koonin, E. V., & Wolf, Y. I. (2008). Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res*, 36(21), 6688-6719. doi:10.1093/nar/gkn668
- Lenski, R. E. (2017). Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. *ISME J*, 11(10), 2181-2194. doi:10.1038/ismej.2017.69
- Lozada-Chavez, I., Janga, S. C., & Collado-Vides, J. (2006). Bacterial regulatory networks are extremely flexible in evolution. *Nucleic Acids Res*, 34(12), 3434-3445. doi:10.1093/nar/gkl423
- Macdonald, S. J., Lin, G. G., Russell, C. W., Thomas, G. H., & Douglas, A. E. (2012). The central role of the host cell in symbiotic nitrogen metabolism. *Proc Biol Sci*, 279(1740), 2965-2973. doi:10.1098/rspb.2012.0414

- Maslov, S., Krishna, S., Pang, T. Y., & Sneppen, K. (2009). Toolbox model of evolution of prokaryotic metabolic networks and their regulation. *Proc Natl Acad Sci U S A*, 106(24), 9743-9748. doi:10.1073/pnas.0903206106
- Meselson, M., & Yuan, R. (1968). DNA restriction enzyme from *E. coli*. *Nature*, 217(5134), 1110-1114. doi:10.1038/2171110a0
- Monk, J. M., Charusanti, P., Aziz, R. K., Lerman, J. A., Premyodhin, N., Orth, J. D., . . . Palsson, B. O. (2013). Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. *Proc Natl Acad Sci U S A*, 110(50), 20338-20343. doi:10.1073/pnas.1307797110
- Orr, H. A. (2005). The genetic theory of adaptation: a brief history. *Nat Rev Genet*, 6(2), 119-127. doi:10.1038/nrg1523
- Orth, J. D., Thiele, I., & Palsson, B. O. (2010). What is flux balance analysis? *Nat Biotech*, 28(3), 245-248.
- Otto, T. D., Bohme, U., Jackson, A. P., Hunt, M., Franke-Fayard, B., Hoeijmakers, W. A., . . . Janse, C. J. (2014). A comprehensive evaluation of rodent malaria parasite genomes and gene expression. *BMC Biol*, 12, 86. doi:10.1186/s12915-014-0086-0
- Pal, C., Papp, B., & Lercher, M. J. (2005). Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat Genet*, 37(12), 1372-1375. doi:10.1038/ng1686
- Pal, C., Papp, B., Lercher, M. J., Csermely, P., Oliver, S. G., & Hurst, L. D. (2006). Chance and necessity in the evolution of minimal metabolic networks. *Nature*, 440(7084), 667-670. doi:10.1038/nature04568
- Pang, T. Y., & Lercher, M. J. (2017). Supra-operonic clusters of functionally related genes (SOCs) are a source of horizontal gene co-transfers. *Sci Rep*, 7, 40294. doi:10.1038/srep40294
- Pang, T. Y., & Lercher, M. J. (2019). Each of 3,323 metabolic innovations in the evolution of *E. coli* arose through the horizontal transfer of a single DNA segment. *Proc Natl Acad Sci U S A*, 116(1), 187-192. doi:10.1073/pnas.1718997115
- Satish Kumar, V., Dasika, M. S., & Maranas, C. D. (2007). Optimization based automated curation of metabolic reconstructions. *BMC Bioinformatics*, 8, 212. doi:10.1186/1471-2105-8-212
- Schellenberger, J., Park, J. O., Conrad, T. M., & Palsson, B. O. (2010). BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, 11, 213. doi:10.1186/1471-2105-11-213
- Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y., & Ishikawa, H. (2000). Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature*, 407(6800), 81-86. doi:10.1038/35024074
- Shlomi, T., Berkman, O., & Ruppin, E. (2005). Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc Natl Acad Sci U S A*, 102(21), 7695-7700. doi:10.1073/pnas.0406346102
- Szappanos, B., Fritzscheier, J., Csorgo, B., Lazar, V., Lu, X., Fekete, G., . . . Papp, B. (2016). Adaptive evolution of complex innovations through stepwise metabolic niche expansion. *Nat Commun*, 7, 11607. doi:10.1038/ncomms11607
- Thiele, I., Vlassis, N., & Fleming, R. M. (2014). fastGapFill: efficient gap filling in metabolic networks. *Bioinformatics*, 30(17), 2529-2531. doi:10.1093/bioinformatics/btu321

- Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., . . . Venter, J. C. (1997). The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature*, 388(6642), 539-547. doi:10.1038/41483
- Watson, M. R. (1984). Metabolic maps for the Apple II. *Biochemical Society Transactions*, 12(6), 1093-1094. doi:10.1042/bst0121093
- Wolf, Y. I., & Koonin, E. V. (2013). Genome reduction as the dominant mode of evolution. *Bioessays*, 35(9), 829-837. doi:10.1002/bies.201300037
- Zanghellini, J., Ruckerbauer, D. E., Hanscho, M., & Jungreuthmayer, C. (2013). Elementary flux modes in a nutshell: properties, calculation and applications. *Biotechnol J*, 8(9), 1009-1016. doi:10.1002/biot.201200269

2.6 Supplementary Information for Metabolic Complexity and Its Role in Cyanobacterial Adaptation

Claus Jonathan Fritzemeier, Sajjad Ghaffarinasab, Felix Lieder, Balázs Szappanos, Florian Jarre, Balázs Papp, Csaba Pál, Martin J. Lercher

Table S1. Organism specific models and their properties. The models in this table are sorted in ascending order by the number of metabolic genes. The column “55 *E.coli*” indicates, whether models from the publication Monk et al. (2013).

Model ID	Organism	Gene count	Metabolite count	Reaction count	55 <i>E.coli</i>	Taxonomy ID	PubMed ID
e_coli_core	<i>Escherichia coli str. K-12 substr. MG1655</i>	137	72	95		511145	26443778
iSM199	<i>Buchnera aphidicola str. APS</i>	199	298	297			22513857
iIS312	<i>Trypanosoma cruzi Dm28c</i>	312	606	519		1416333	
iIS312_Epimastigote	<i>Trypanosoma cruzi Dm28c</i>	312	606	519		1416333	
iIS312_Amastigote	<i>Trypanosoma cruzi Dm28c</i>	312	606	519		1416333	
iIS312_Trypomastigote	<i>Trypanosoma cruzi Dm28c</i>	312	606	519		1416333	
iIT341	<i>Helicobacter pylori 26695</i>	339	485	554		85962	16077130
iAM_Pb448	<i>Plasmodium berghei</i>	448	903	1067		5821	29300748
iAM_Pc455	<i>Plasmodium cynomolgi strain B</i>	455	907	1074		1120755	29300748
iAM_Pk459	<i>Plasmodium knowlesi strain H</i>	459	909	1079		5851	29300748
iAM_Pv461	<i>Plasmodium vivax Sal-I</i>	461	909	1078		126793	29300748
iAM_Pf480	<i>Plasmodium falciparum 3D7</i>	480	909	1083		36329	29300748
iLJ478	<i>Thermotoga maritima MSB8</i>	482	570	652		243274	19762644
iNF517	<i>Lactococcus lactis subsp. cremoris MG1363</i>	516	650	754		416870	23974365
iSB619	<i>Staphylococcus aureus subsp. aureus N315</i>	619	655	743		158879	15752426
iJN678	<i>Synechocystis sp. PCC 6803</i>	622	795	863		1148	22308420
iHN637	<i>Clostridium ljungdahlii DSM 13528</i>	637	698	785		748727	24274140
iNJ661	<i>Mycobacterium tuberculosis H37Rv</i>	661	826	1025		83332	17555602
iAF692	<i>Methanosarcina barkeri str. Fusaro</i>	692	628	690		269797	16738551
iCN718	<i>Acinetobacter baumannii AYE</i>	709	888	1015		509173	29692801
iJN746	<i>Pseudomonas putida KT2440</i>	746	909	1056		160488	18793442
iND750	<i>Saccharomyces cerevisiae S288c</i>	750	1059	1266		559292	15197165
iJB785	<i>Synechococcus elongatus PCC 7942</i>	785	768	849		1140	27911809
iPC815	<i>Yersinia pestis CO92</i>	815	1552	1961		214092	21995956

Model ID	Organism	Gene count	Metabolite count	Reaction count	55 <i>E.coli</i>	Taxonomy ID	PubMed ID
iSynCJ816	<i>Synechocystis sp. PCC 6803</i>	816	928	1044		1148	–
iYO844	<i>Bacillus subtilis subsp. subtilis str. 168</i>	844	991	1250		224308	17573341
iYS854	<i>Staphylococcus aureus subsp. aureus USA300_TCH1516</i>	866	1335	1455		451516	30625152
iCN900	<i>Clostridioides difficile 630</i>	900	885	1229		272563	
iJR904	<i>Escherichia coli str. K-12 substr. MG1655</i>	904	761	1075		511145	12952533
iMM904	<i>Saccharomyces cerevisiae S288c</i>	905	1226	1577		559292	19321003
iAF987	<i>Geobacter metallireducens GS-15</i>	987	1109	1285		269799	24762737
iEK1008	<i>Mycobacterium tuberculosis H37Rv</i>	1008	998	1226		83332	29499714
iLB1027_lipid	<i>Phaeodactylum tricornutum CCAP 1055/1</i>	1027	2172	4456		556484	27152931
iSDY_1059	<i>Shigella dysenteriae Sd197</i>	1059	1890	2540	X	300267	24277855
iRC1080	<i>Chlamydomonas</i>	1086	1706	2191		3052	21811229
iSBO_1134	<i>Shigella boydii Sb227</i>	1134	1910	2592	X	300268	24277855
iSbBS512_1146	<i>Shigella boydii CDC 3083-94</i>	1147	1912	2592	X	344609	24277855
iSFxv_1172	<i>Shigella flexneri 2002017</i>	1169	1918	2639	X	591020	24277855
iSFV_1184	<i>Shigella flexneri 5 str. 8401</i>	1184	1917	2622	X	373384	24277855
iS_1188	<i>Shigella flexneri 2a str. 2457T</i>	1188	1914	2620	X	198215	24277855
iSF_1195	<i>Shigella flexneri 2a str. 301</i>	1195	1917	2631	X	198214	24277855
iYL1228	<i>Klebsiella pneumoniae subsp. pneumoniae MGH 78578</i>	1229	1658	2262		272620	21296962
iSSON_1240	<i>Shigella sonnei Ss046</i>	1240	1938	2694	X	300269	24277855
iAF1260	<i>Escherichia coli str. K-12 substr. MG1655</i>	1261	1668	2382		511145	17593909
iAF1260b	<i>Escherichia coli str. K-12 substr. MG1655</i>	1261	1668	2388		511145	19840862
iECH74115_1262	<i>Escherichia coli O157:H7 str. EC4115</i>	1262	1918	2695	X	444450	24277855
STM_v1_0	<i>Salmonella enterica subsp. enterica serovar Typhimurium str. LT2</i>	1271	1802	2545		99287	21244678
iECED1_1282	<i>Escherichia coli ED1a</i>	1279	1929	2707	X	585397	24277855
iECUMN_1333	<i>Escherichia coli UMN026</i>	1332	1935	2741	X	585056	24277855

Model ID	Organism	Gene count	Metabolite count	Reaction count	55 <i>E.coli</i>	Taxonomy ID	PubMed ID
iG2583_1286	<i>Escherichia coli</i> O55:H7 str. CB9615	1283	1919	2705	X	701177	24277855
iE2348C_1286	<i>Escherichia coli</i> O127:H6 str. E2348/69	1284	1919	2704	X	574521	24277855
iECSP_1301	<i>Escherichia coli</i> O157:H7 str. TW14359	1299	1920	2713	X	544404	24277855
iECNA114_1301	<i>Escherichia coli</i> NA114	1301	1927	2719	X	1033813	24277855
iECs_1301	<i>Escherichia coli</i> O157:H7 str. Sakai	1301	1923	2721	X	386585	24277855
iLF82_1304	<i>Escherichia coli</i> LF82	1302	1940	2727	X	591946	24277855
iECOK1_1307	<i>Escherichia coli</i> IHE3034	1304	1943	2730	X	714962	24277855
iECS88_1305	<i>Escherichia coli</i> S88	1305	1944	2730	X	585035	24277855
ic_1306	<i>Escherichia coli</i> CFT073	1307	1938	2727	X	199310	24277855
iZ_1308	<i>Escherichia coli</i> O157:H7 str. EDL933	1308	1923	2722	X	155864	24277855
iECP_1309	<i>Escherichia coli</i> 536	1309	1943	2740	X	362663	24277855
iUTI89_1310	<i>Escherichia coli</i> UTI89	1310	1942	2726	X	364106	24277855
iNRG857_1313	<i>Escherichia coli</i> O83:H1 str. NRG 857C	1311	1945	2736	X	685038	24277855
iAPECO1_1312	<i>Escherichia coli</i> APEC O1	1313	1944	2736	X	405955	24277855
iEC042_1314	<i>Escherichia coli</i> 042	1314	1926	2715	X	216592	24277855
iUMN146_1321	<i>Escherichia coli</i> UM146	1319	1944	2736	X	869729	24277855
iECABU_c1320	<i>Escherichia coli</i> ABU 83972	1320	1944	2732	X	655817	24277855
iEcHS_1320	<i>Escherichia coli</i> HS	1321	1965	2754	X	331112	24277855
iECIAI39_1322	<i>Escherichia coli</i> IAI39	1321	1957	2722	X	585057	24277855
iECO103_1326	<i>Escherichia coli</i> O103:H2 str. 12009	1327	1958	2759	X	585395	24277855
iECSF_1327	<i>Escherichia coli</i> SE15	1327	1951	2743	X	431946	24277855
iECDH10B_1368	<i>Escherichia coli</i> str. K-12 substr. DH10B	1327	1947	2743	X	316385	24277855
iBWG_1329	<i>Escherichia coli</i> BW2952	1328	1949	2742	X	595496	24277855
iECO111_1330	<i>Escherichia coli</i> O111:H- str. 11128	1328	1959	2761	X	585396	24277855
iECB_1328	<i>Escherichia coli</i> B str. REL606	1329	1953	2749	X	413997	24277855
iEC55989_1330	<i>Escherichia coli</i> 55989	1330	1953	2757	X	585055	24277855

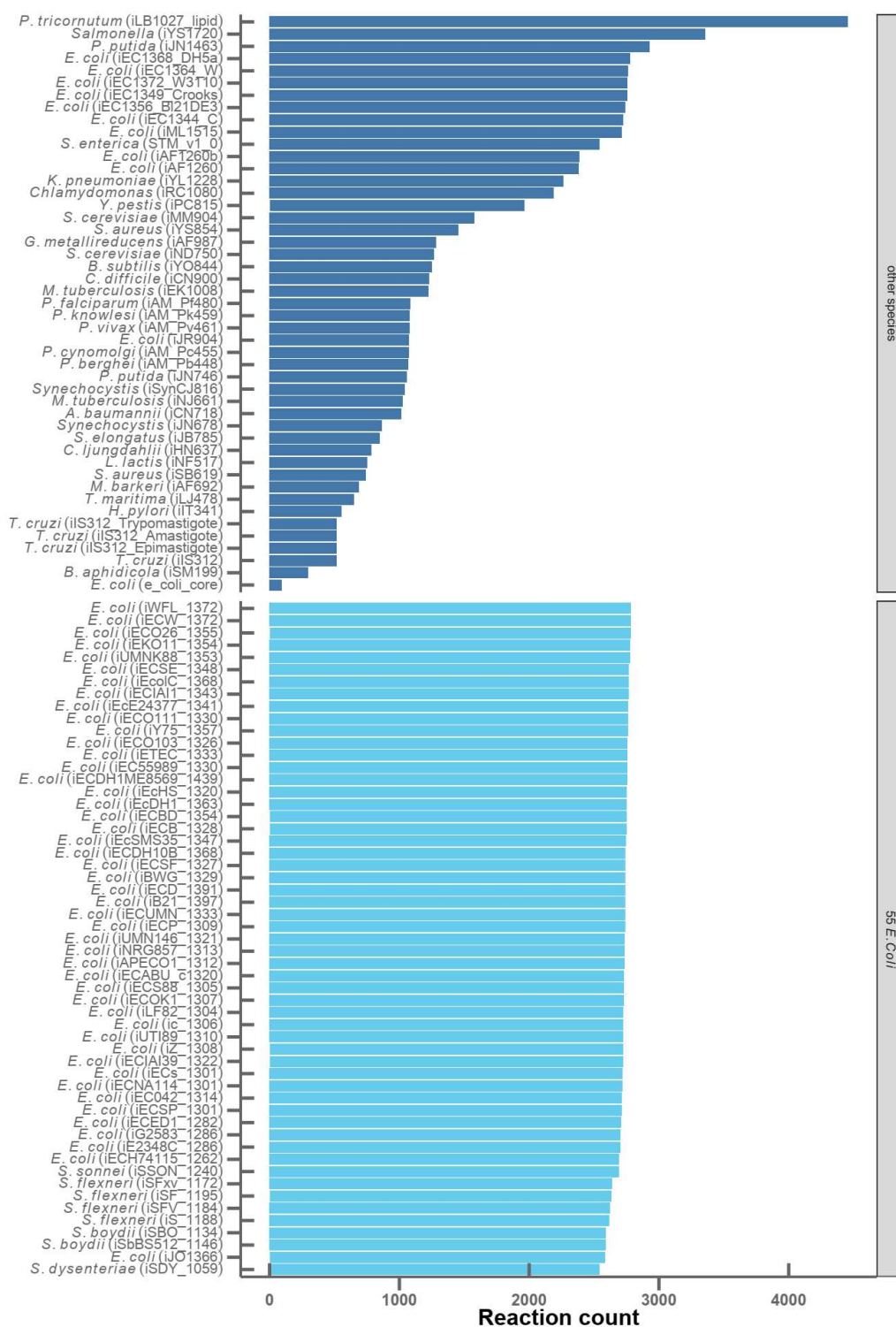
Model ID	Organism	Gene count	Metabolite count	Reaction count	55 <i>E.coli</i>	Taxonomy ID	PubMed ID
iECD_1391	<i>Escherichia coli BL21(DE3)</i>	1333	1945	2742	X	469008	24277855
iETEC_1333	<i>Escherichia coli ETEC H10407</i>	1333	1964	2757	X	316401	24277855
iB21_1397	<i>Escherichia coli BL21(DE3)</i>	1337	1945	2742	X	469008	24277855
iEcE24377_1341	<i>Escherichia coli E24377A</i>	1341	1974	2764	X	331111	24277855
iECIAI1_1343	<i>Escherichia coli IAI1</i>	1343	1970	2766	X	585034	24277855
iEC1344_C	<i>Escherichia coli C</i>	1344	1934	2726		498388	27667363
iEcSMS35_1347	<i>Escherichia coli SMS-3-5</i>	1347	1949	2747	X	439855	24277855
iECSE_1348	<i>Escherichia coli SE11</i>	1348	1957	2769	X	409438	24277855
iEC1349_Crooks	<i>Escherichia coli ATCC 8739</i>	1349	1946	2756		481805	27667363
iUMNK88_1353	<i>Escherichia coli UMNK88</i>	1353	1971	2778	X	696406	24277855
iECBD_1354	<i>Escherichia coli BL21-Gold(DE3)pLysS AG</i>	1354	1954	2749	X	866768	24277855
iEKO11_1354	<i>Escherichia coli KO11FL</i>	1354	1974	2779	X	595495	24277855
iECO26_1355	<i>Escherichia coli O26:H11 str. 11368</i>	1355	1965	2781	X	573235	24277855
iEC1356_BI23DE3	<i>Escherichia coli BL21(DE3)</i>	1356	1918	2740		469008	27667363
iY75_1357	<i>Escherichia coli str. K-12 substr. W3110</i>	1358	1953	2760	X	316407	24277855
iEcDH1_1363	<i>Escherichia coli DH1</i>	1363	1949	2751	X	536056	24277855
iEC1364_W	<i>Escherichia coli W</i>	1364	1927	2764		566546	27667363
iJO1366	<i>Escherichia coli str. K-12 substr. MG1655</i>	1367	1805	2583	X	511145	21988831
iEcolC_1368	<i>Escherichia coli ATCC 8739</i>	1368	1971	2769	X	481805	24277855
iEC1368_DH5a	<i>Escherichia coli DH5[alpha]</i>	1368	1951	2779		668369	27667363
iEC1372_W3110	<i>Escherichia coli str. K-12 substr. W3110</i>	1372	1918	2758		316407	27667363
iECW_1372	<i>Escherichia coli W</i>	1372	1975	2783	X	566546	24277855
iWFL_1372	<i>Escherichia coli W</i>	1372	1975	2783	X	566546	24277855
iECDH1ME8569_1439	<i>Escherichia coli DH1</i>	1439	1950	2756	X	536056	24277855
iJN1463	<i>Pseudomonas putida KT2440</i>	1452	2153	2927		160488	
iML1515	<i>Escherichia coli str. K-12 substr. MG1655</i>	1516	1877	2712		511145	29020004
iYS1720	<i>Salmonella pan-reactome</i>	1707	2436	3357			30218022

Table S2. Essential Reactions Added to Enable Growth on Fully Rich Medium

Model ID	Organism	Added reaction	Name of the reaction
iLS312	<i>Trypanosoma cruzi</i> Dm28c	UMPK	UMP kinase
		ADKd	Adenylate kinase
		DADNt2	Deoxyadenosine transport
		DGSNt2	Deoxyguanosine transport
		NTD8	5'-nucleotidase (dGMP)
		TYRabc	polar amino acid transport system via abc system (tyrosine)
		TRPabc	polar amino acid transport system via abc system (tryptophan)
		AMPt6	AMP transport inout
		CMPt6	CMP transport inout
		DTMPt6	DTMP transport inout
		DURIK1_1	Deoxyuridine kinase
		DURIt2	Deoxyuridine transport
		MANpts	D-mannose transport via PEP:Pyr PTS
iLJ478	<i>Thermotoga maritima</i>	ASNt2r	L asparagine reversible transport
iSM199	Buchnera aphidicola	LEUTAi	Leucine transaminase
		ILEabc	L-isoleucine transport via ABC system
		VALt2r	L valine reversible transport
		ILETA2	Branched-chain-amino-acid transaminase
		PHEabc	L phenylalanine reversible transport via abc system

Table S3. Common Subsystems with Associated Gains in Simulations and Experiments

Subsystem	Total gains in Simulations	Total gains in Experiments	Odds ratio Simulations	Odds ratio Experiments	P-value Simulations	P-value Experiments
Alternate Carbon Metabolism	659	67	0.48	0.69	0.83	0.98
Arginine and Proline Metabolism	6	4	0.15	1.23	0.58	0.27
Cell Envelope Biosynthesis	16	15	0.21	1.73	0.51	0.02
Cofactor and Prosthetic Group Biosynthesis	2040	2	13.9	0.25	0	0.99
Cysteine Metabolism	4	1	0.69	2.81	0.19	0.10
Exchange	2	1	0.05	0.31	0.65	0.48
Fatty Acid Metabolism	11	2	0.41	0.68	0.19	0.24
Folate Metabolism	652	4	22.8	1.71	0.02	0.2
Inorganic Ion Transport and Metabolism	5	52	0.01	1.5	0.99	0.09
Methionine Metabolism	189	2	3.82	0.33	0.07	0.68
Nucleotide Salvage Pathway	49	7	0.4	1.00	0.56	0.45
S_Alternate_Carbon_source	31	3	0.68	1.51	0.21	0.23
S_Aromatic_Acid_Breakdown	17	2	0.05	0.8	0.86	0.62
S_Transport_Outer_Membrane_Porin	1	5	0.01	1.41	0.83	0.27
Transport Inner Membrane	246	56	0.26	1.29	0.95	0.04
Transport Outer Membrane	9	15	0.08	2.93	0.89	0.001
Tyrosine Tryptophan, and Phenylalanine Metabolism	379	2	16.8	0.74	0.01	0.61
Urea Cycle	49	3	0.94	1.18	0.14	0.19
Valine Leucine, and Isoleucine Metabolism	760	2	23.3	0.66	0.006	0.4



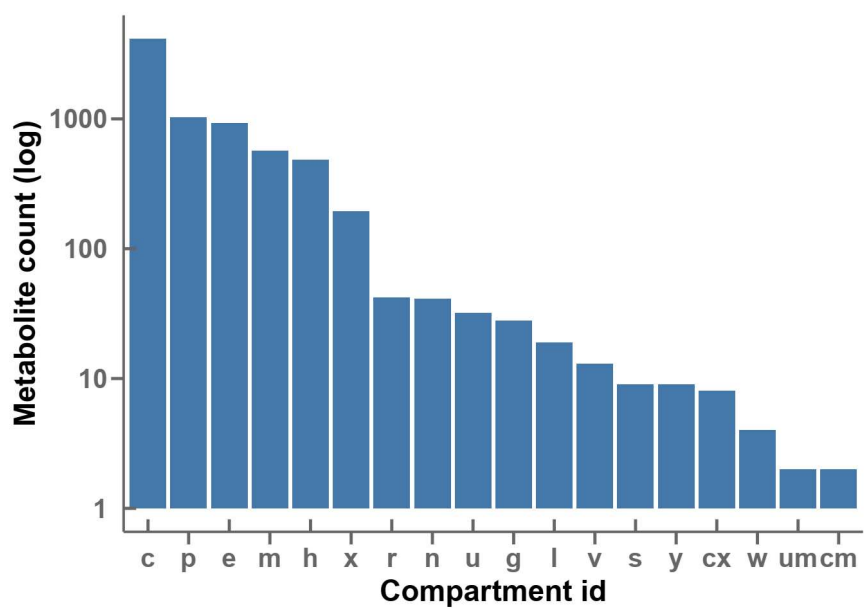


Figure S2. Additional properties of the supermodel. a) Number of metabolites in the compartments. The compartment with the most reactions is the cytosol (c) followed by the extracellular (e) and periplasm (p). The Remaining compartments originate from the eukaryotic models used.

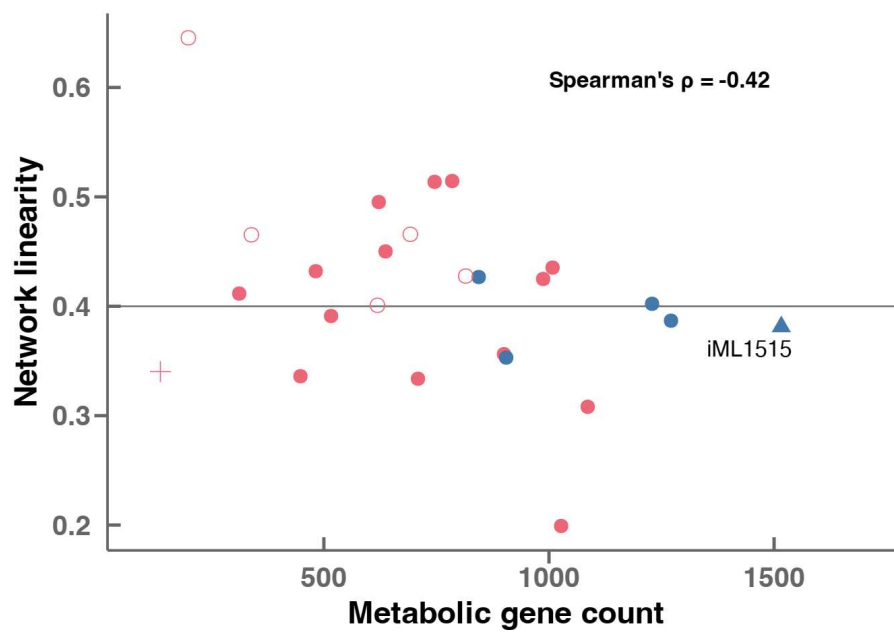


Figure S3. Small metabolic networks tend to be less branched than large networks. Network linearity is defined as the fraction of metabolites that participate in only two reactions, *i.e.*, metabolites that are intermediates in unbranched pathways. The colors of circles and points distinguish specialists (red) and generalists (blue). The 55 *E. coli* strains are represented by the iML1515 model (blue triangle) only. Organisms with known auxotrophies are shown as open circles. The highly branched *E. coli* core metabolism is shown with a red plus sign. Spearman correlation between network linearity and gene count: $\rho = -0.42$, using only iML1515 as representative for the 55 *E. coli*.

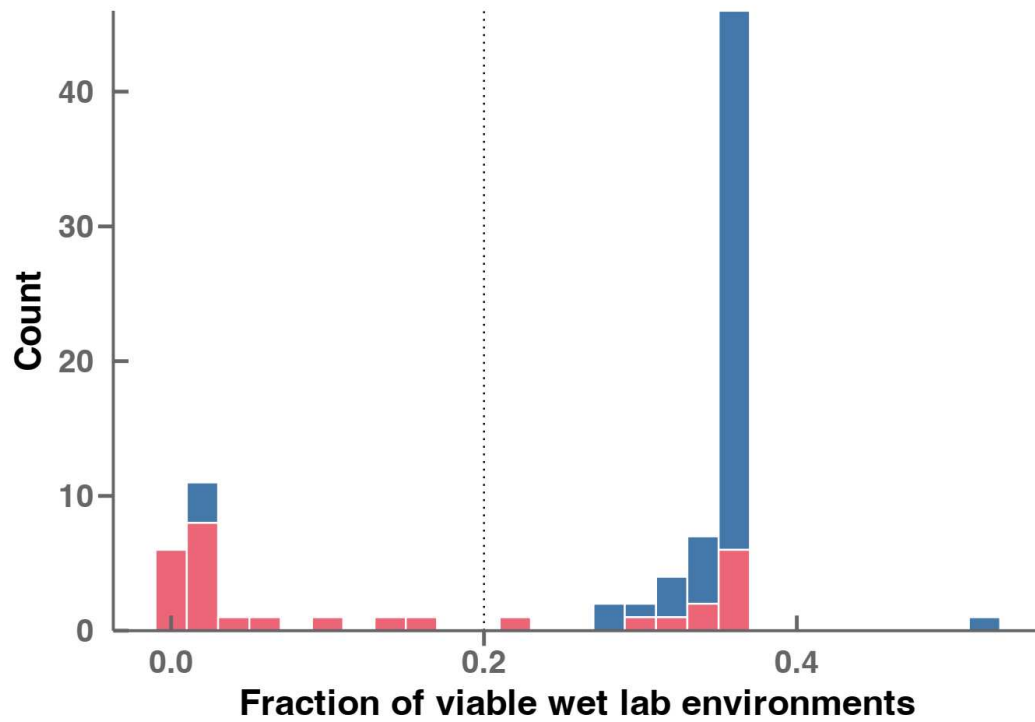


Figure S4. Fractions of viable environments for submodels in wet lab environments (seed). Bars show a single stacked histogram combining 55 *E. coli* strains (blue) and all other species (red). The dotted vertical line marks the threshold used to classify models as specialists (left) versus generalists (right).

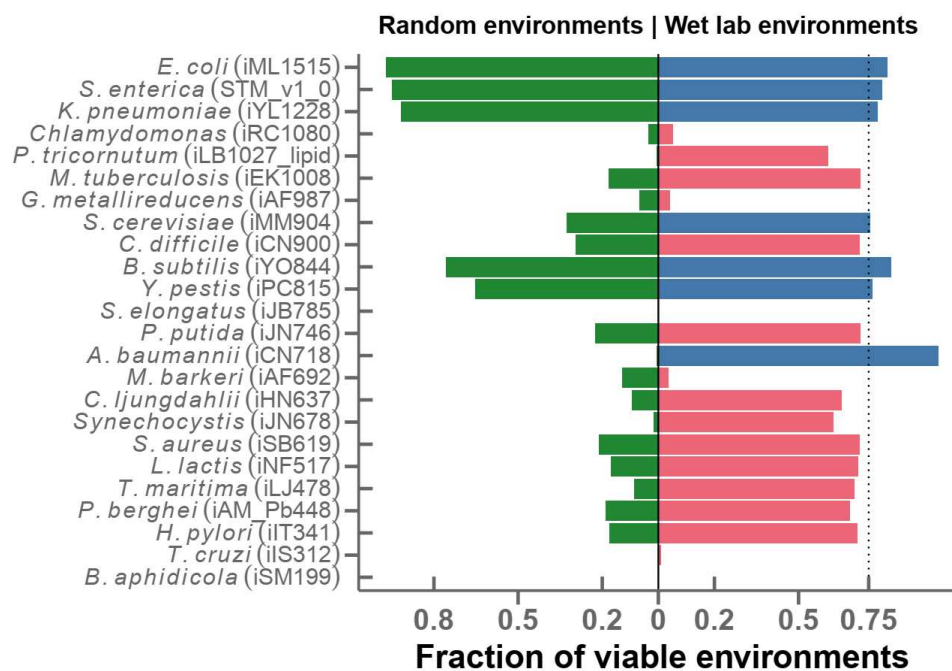


Figure S5. The fraction of viable environments differs widely across submodels, both for random minimal environments (green bars to the left) and for common wet lab environments (blue and red bars to the right), here energy generation as the objective function.

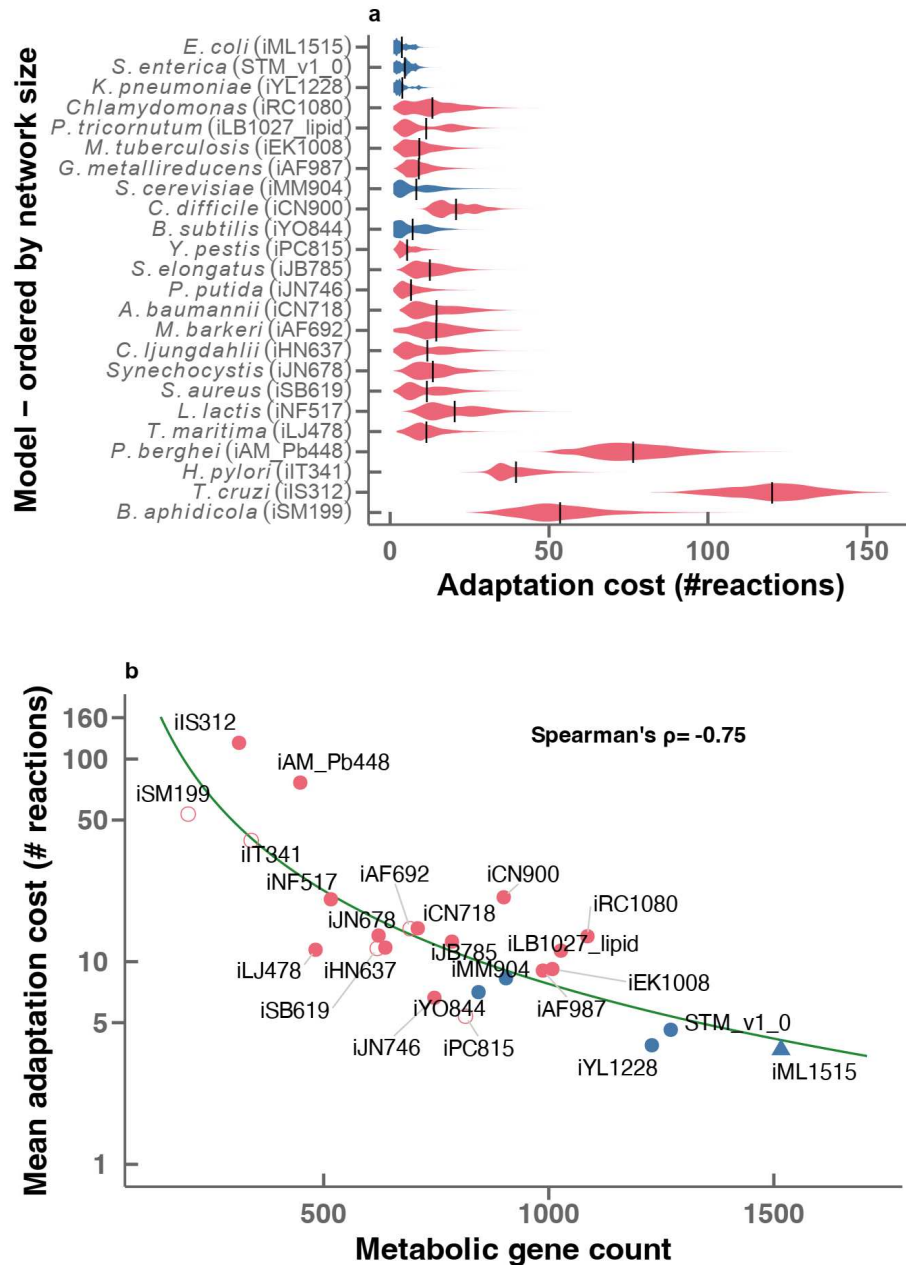


Figure S6. The number of additional reactions required for adaptation decreases with increasing genome size. Same as Fig. 3 of the main text, but considering random environments instead of wet lab environments. (a) Distributions of added reactions, summarized as violin plots. The height at each point in a “violin” indicates the local density of the distribution for the given model. Models are ordered top-down by decreasing size. (b) The average number of added reactions (log scale) plotted against metabolic gene count for each model. The solid line shows the best fitting power law, $added\ reactions = a \times (gene\ count)^b$, with the best-fitting exponent $b=2.54$. In both panels, colors distinguish specialists (red) and generalists (blue). Organisms with known auxotrophies are shown as open circles. The 55 *E. coli* strains are represented by the iML1515 model (blue triangle) only.

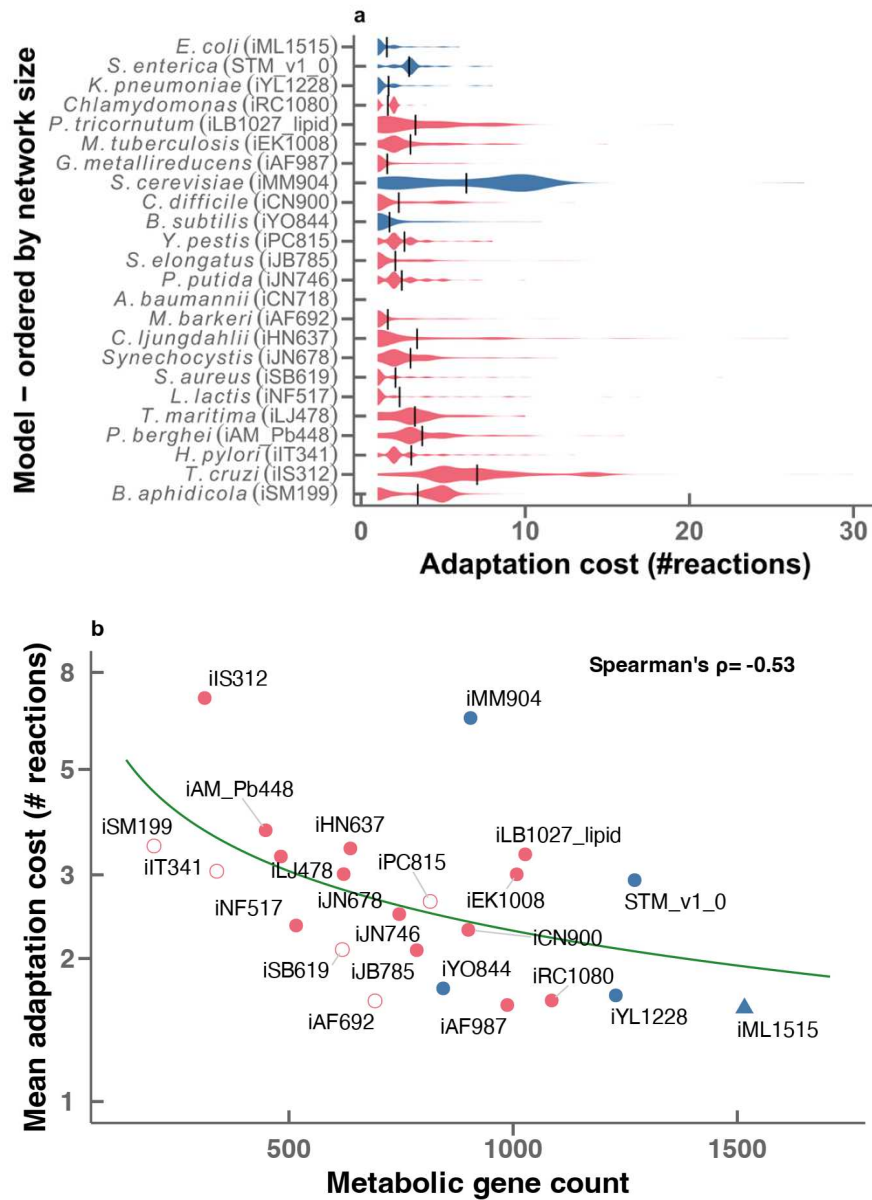


Figure S7. The number of additional reactions required for adaptation decreases with increasing genome size. Same as Fig. 3 of the main text, but considering the generation of energy as the objective function (instead of biomass production). (a) Distributions of added reactions, summarized as violin plots. The height at each point in a “violin” indicates the local density of the distribution for the given model. Models are ordered top-down by decreasing size. (b) The average number of added reactions (log scale) plotted against metabolic gene count for each model. The solid line shows the best fitting power law, $\text{added reactions} = a \times (\text{gene count})^b$, with the best-fitting exponent $b=1.4$. In both panels, colors distinguish specialists (red) and generalists (blue). Organisms with known auxotrophies are shown as open circles. The 55 *E. coli* strains are represented by the iML1515 model (blue triangle) only.

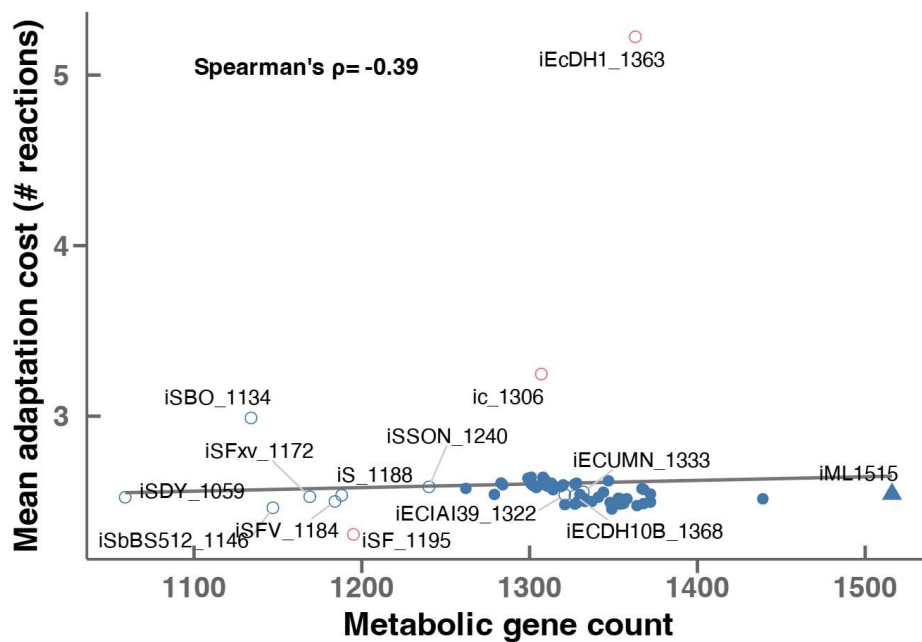


Figure S8. Different *E. coli* strains show similar adaptabilities, despite variations in genome size. Analogous to Fig. 3b of the main text, but showing all *E. coli* submodels. Colors distinguish specialists (red) and generalists (blue). Organisms with known auxotrophies are shown as open circles.

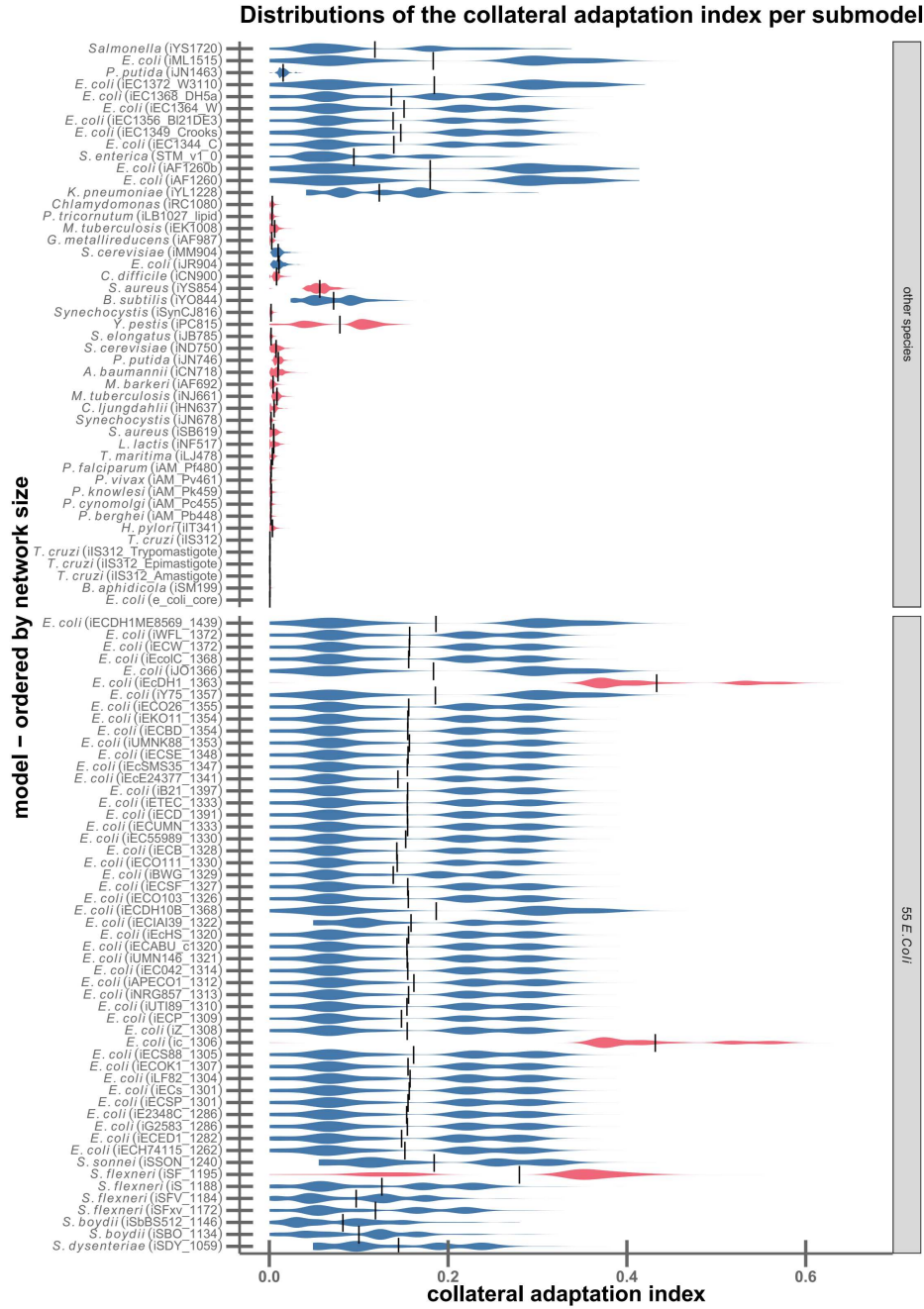


Figure S9. Across species, generalists (blue) show a higher tendency for collateral adaptations than specialists (red), while this trend is reversed within *E. coli*. For each submodel, we first identified the n random environments in which it cannot produce biomass (unviable environments). For each of these environments in turn, we then identified the smallest set of reactions from the supermodel that have to be added to enable biomass production. The collateral adaptation index is then the fraction of the $n-1$ remaining previously unviable environments in which this extended model can grow. Each “violin” summarizes the distribution of the collateral adaptation indices for one submodel. Models in each of the two groups on the y-axis (top: one representative per species; bottom: *E. coli* strains) are sorted by gene count. The mean of each distribution is marked with a vertical line.

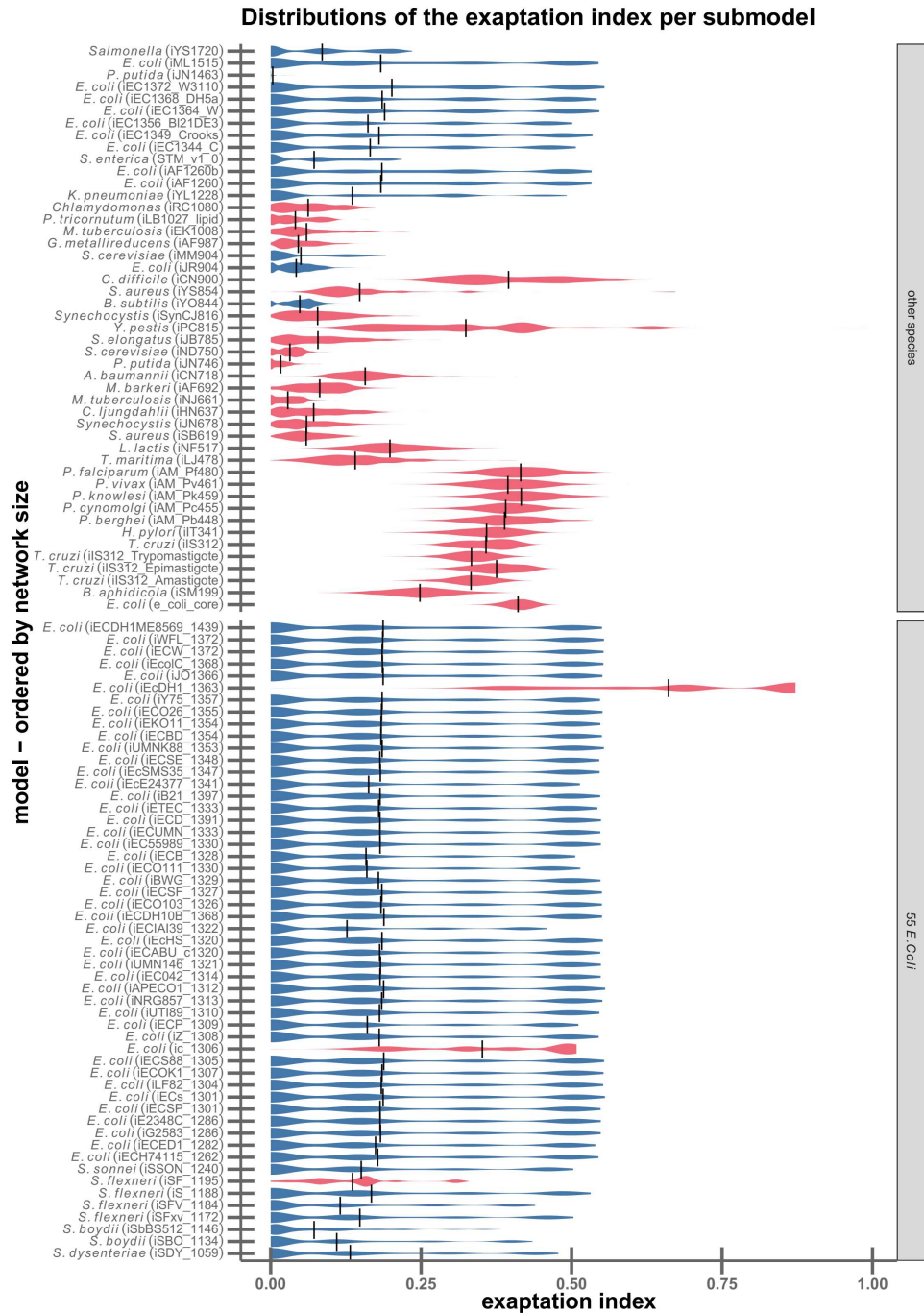


Figure S10. Specialists (red) tend to have higher exaptation indices than generalists (blue). Each “violin” summarizes the distribution of the exaptation index for one submodel. Models in the groups on the y-axis (top: one representative per species; bottom: *E. coli* strains) are sorted by gene count. The mean of each distribution is marked with a vertical line.

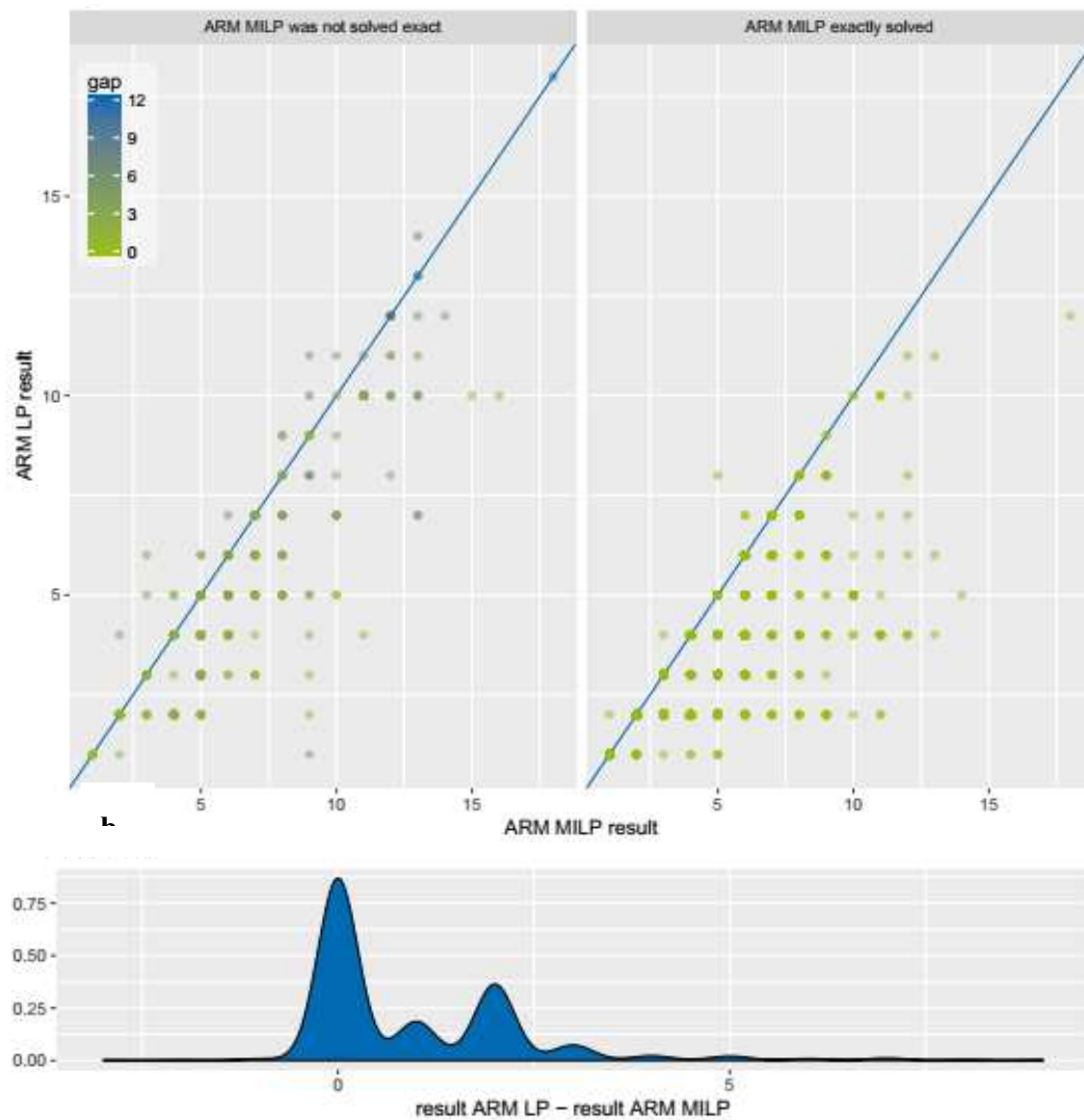


Figure S11. Practical application of ARM LP shows equally good performance as ARM MILP. a) Result (objective value) comparison of ARM MILP and ARM LP. Dot color indicates the gap size (smaller is better). In the left panel, ARM MILP solutions are suboptimal due to the limited computation time. Results shown in the right panel could be solved exactly within the time limit. The blue lines indicate equal objective values. b) Distribution of the difference between ARM LP and ARM MILP results

Chapter 3

Cyanobacterial resource allocation strategies explained by Growth Balance Analysis

S.G. conceived and implemented the GBA models for *Synechocystis* sp. PCC 6803, carried out the simulations, analyzed the data, and drafted the initial manuscript. M.J.L. contributed to the conceptual development, assisted in data interpretation, and critically reviewed the manuscript. O.E. offered theoretical insights into resource-allocation modelling. H.D. supervised the project, supported model development, contributed to data interpretation, and coordinated the final writing of the manuscript.

Manuscript status: Under consideration at the journal of iScience

Cyanobacterial resource allocation strategies explained by Growth Balance Analysis

Sajjad Ghaffarinasab¹, Oliver Ebenhöf², Martin J. Lercher¹, Hugo Dourado^{1*}

¹ Institute for Computer Science and Department of Biology, Heinrich Heine University, Universitätsstraße 1, D-40225 Düsseldorf, Germany

² Institute of Quantitative and Theoretical Biology, Heinrich Heine University Düsseldorf, Düsseldorf 40225, Germany

* To whom correspondence should be addressed

Abstract

Cyanobacteria have emerged as attractive microbial cell factories because of their ability to convert atmospheric CO₂ and sunlight into valuable chemicals. To increase their growth and productivity, one should aim to optimize the allocation of limited cellular resources among different metabolic processes. Here, we developed two growth balance analysis (GBA) models for the cyanobacterium *Synechocystis* sp. PCC 6803. The first model reformulates an existing coarse-grained, non-linear model of *Synechocystis* in the GBA framework, thereby drastically simplifying the mathematical formulation. The streamlined GBA formulation facilitates simple model extensions, such as the inclusion of additional reactions and reactants. We exploit this capability by extending the *Synechocystis* GBA model with a more detailed description of photosynthesis. The extended model captures the main trends of proteome partitioning across environmental conditions. Our findings demonstrate that GBA models provide a useful and easily extensible toolbox for understanding the physiological capabilities of cyanobacteria, their allocation of cellular resources, and the potential of their bioengineering for optimized biomass production.

Keywords:

Growth balance analysis, Optimal Resource allocation, Cyanobacteria, Photodamage, Systems Biology

3.1 Introduction

Photosynthetic cyanobacteria are the only prokaryotes capable of oxygenic photosynthesis, converting CO₂ and sunlight into biomass. Compared to plants and microalgae, cyanobacteria exhibit higher photosynthetic efficiency as well as faster growth, and they are more accessible to genetic manipulations. These features make them an important model organism for designing microbial cell factories (Santos-Merino et al., 2023). While a vast amount of experimental high-throughput data – including genomics, transcriptomics, and proteomics – is available (Babele et al., 2019; Jahn et al., 2018; Matthias et al., 2014; Zavřel et al., 2019), a mechanistic understanding of cyanobacterial resource allocation from simple principles remains an ongoing challenge in biotechnology.

Such mechanistic understanding can be sought through computational models, which also allow the study of phenotypes that are not easily accessible to experiments. Recently, various linear approaches in computational biology have been used to study resource allocation of organisms at balanced growth, a steady-state condition where concentrations of cellular components are constant in time (Dourado et al., 2023; Molenaar et al., 2009). These linear approaches include genome-scale Models of Metabolism and Macromolecular Expression, ME-models (O’Brien et al., 2013); Resource Balance Analysis, RBA (Goelzer et al., 2011); and genome-scale models with enzymatic constraints using kinetic and omics data, GECKO (Sánchez et al., 2017). These methodologies consider the production cost of macromolecules for catalyzing each reaction by approximating the kinetic rate laws as a linear relationship between fluxes and the concentration of their catalysts, ignoring metabolite concentrations and how they influence fluxes via the saturation of the catalysts (Dourado & Lercher, 2020).

In contrast to these linear cellular models, alternative nonlinear models can explicitly account for metabolite concentrations and their influence on reaction fluxes through kinetic rate laws. Molenaar et al. (2009) introduced nonlinear “self-replicator” models, where resource allocation patterns follow directly from the optimization of the cellular growth rate under basic physiological constraints. These constraints include a fixed total protein concentration; nonlinear kinetic rate laws for all modeled reactions, including transporters that exchange mass with the environment, enzymatic reactions converting internal reactants, and a “ribosome” reaction producing the total amount of protein required to catalyze all cellular reactions; and mass conservation accounting for the dilution by growth of all components. Later work (Burnap, 2015; Faizi et al., 2018) extended this self-replicator approach to simulate photoautotrophic resource allocation, including physiological processes specific to photoautotrophs, such as photodamage and carbon cycling.

Cellular models with nonlinear kinetics have been limited to small, coarse-grained descriptions of cellular physiology (Burnap, 2015; Faizi et al., 2018; Jahn et al., 2018; Molenaar et al., 2009). The main purpose of such small, nonlinear models is to reduce the cellular complexity to only a

few components and reactions that still provide important insights into the main patterns of cellular behavior. This simplification is achieved by combining multiple essential enzymes and metabolic pathways into single catalytic units. These coarse-grained models are useful tools for identifying metabolic tradeoffs under different conditions without requiring much information about the organism (Burnap, 2015; Faizi et al., 2018; Jahn et al., 2018; Molenaar et al., 2009). The restriction to simplified cellular models was enforced by the difficulty of numerically solving nonlinear optimization problems for large-scale models (Wortel et al., 2018). However, recent studies have shown how the mathematical formulation of such nonlinear problems can be greatly simplified (Dourado et al., 2023; Dourado & Lercher, 2020). These studies not only indicate the feasibility of simulating larger nonlinear models, but also shed new light on the mathematical properties of such models.

Below, to study phototrophic growth in cyanobacteria, we reformulate and extend an existing model of the cyanobacterium *Synechocystis* sp. PCC 6803, hereafter simply referred to as "*Synechocystis*". Faizi et al. (2018) proposed two coarse-grained nonlinear models for this organism. The first basic model comprises four reactions: carbon transport (T), metabolism and carbon assimilation (M), ribosome and protein translation (R), and a "light reaction" catalyzed by a photosynthetic unit (PSU). The photosynthetic unit in this model is separated into two components: an "inactive" protein unit P^0 and an "active" unit P^* . In this model, photosynthesis involves the conversion between P^0 and P^* , providing "e" when absorbing light. Faizi et al. (2018) then propose a second model that also accounts for photoinhibition due to the photodamage of the active photosynthetic unit at high light intensities, leading to the degradation of the photosynthetic unit into its amino acid constituents. With both models, the authors simulated the optimal proteome allocation to the various reactions, considering different environmental conditions distinguished by the external carbon concentrations a_C (not distinguishing between CO_2 and HCO_3^-) and the light intensity a_L . The simulations require the solution of a system of differential equations and consider the maximization of growth rate under the constraints given by mass conservation, reaction kinetics, and the fixed total concentration of cellular components. The cellular components consist of internal carbon, an amino acid pool, and one protein acting as a catalyst for each reaction in the model. The growth rate is defined as the net mass influx of carbon divided by the fixed total mass concentration of cellular components.

Below, we present Growth Balance Analysis (GBA) as an alternative mathematical framework for modeling phototrophic growth in cyanobacteria. Compared to the methodology in (Faizi et al., 2018), GBA represents equivalent models with simpler equations that not only facilitate more efficient calculations, but also allow easier model extensions through additional reactions and metabolites. As a proof of concept, we first present a GBA model that corresponds to that of (Faizi et al., 2018) and uses identical parameter values. We then present a second, extended GBA model capable of predicting the optimal proteome allocation in more detail.

GBA has been introduced before as a general, simplified framework to study completely self-replicating kinetic models of cells (Dourado et al., 2023; Dourado & Lercher, 2020). The models

studied by GBA are completely defined by the triple (M, τ, ρ) . Here, M denotes a mass fraction matrix, *i.e.*, an internal stoichiometric matrix normalized by the molecular masses of the reactants and products, which includes a row “p” for the total protein concentration and a column “r” for a ribosome reaction that produces all proteins; $\tau = \tau(c, a)$ denotes a vector of reaction turnover times, which are functions of internal metabolite concentrations c and external concentrations a determined by – typically nonlinear – kinetic rate laws; and ρ denotes the fixed sum of mass concentrations across all cellular components, which are the metabolites and the proteins catalyzing the reactions. In the GBA framework, the optimal cell state is found by solving an optimization problem that is completely defined by the following two algebraic equations, formulated in terms of *flux fractions* $q := v/\mu\rho$, where v denotes the mass reaction fluxes and μ is the cellular growth rate (Dourado et al., 2023). The first equation defines the growth rate $\mu(q, a)$ as a function of flux fractions and external concentrations,

$$\mu(q, a) = \frac{M_r^p q_r}{q \cdot \tau(\rho M q, a)} , \quad (1)$$

where M_r^p is the entry in the matrix M corresponding to the ribosome column “r” and the total protein row “p”. Equation (1) emerges from a reformulation of the balanced growth problem in terms of dimensionless flux fractions. It expresses the growth rate as a ratio between the production of total protein mass by the ribosome, and the total proteome investment required to sustain all reaction fluxes. The second equation simply enforces the constraint on cellular density,

$$\sum_{i,j} M_j^i q_j = 1 \quad , \quad (2)$$

where we sum over all rows i and all columns j of M . Equation (1) already encodes all the constraints in Table 1 in terms of f (see Methods and (Dourado et al., 2023) for details), except for the density constraint captured by equation (2). The vector f determines all system properties listed in Table 1. This includes the concentrations c (metabolite concentrations c_m and total protein concentration c_p),

$$c = \rho M q \quad , \quad (3)$$

and each protein concentration p_j allocated to the catalyst or transporter of reaction j ,

$$p_j = \mu \rho q_j \tau_j(\rho M q, a). \quad (4)$$

The reaction turnover times τ are defined by rate laws that depend on metabolite mass concentrations c_m , which are in turn uniquely determined by equation (4); here, we use irreversible Michaelis-Menten rate laws with inhibition. The turnover times τ of transport reactions at the cell surface additionally depend on the external concentrations a that define the environmental condition of interest. Importantly, in GBA, biomass is not an assumed input but an emergent output determined from first principles: it corresponds to the sum of all intracellular

components, constrained by total density ρ . An overview of the parameters, properties, and equations used to define the base GBA model for *Synechocystis* is provided in **Table 1**.

Table 1. The parameters, properties, and equations defining the base GBA model for *Synechocystis*.

Model parameters	M : mass fraction matrix k_{cat}^j : turnover number of reaction j [h^{-1}] K_m^j : Michaelis constant of metabolite m in reaction j [g L^{-1}] I_n^j : Inhibition constant for external concentration n in reaction j [g L^{-1}] ρ : cellular density [g L^{-1}]
System properties	v_j : flux of reaction j [$\text{g L}^{-1} \text{h}^{-1}$] μ : growth rate [h^{-1}] c : reactant concentration vector (including concentrations c^m of metabolites and the total protein concentration c^p) [g L^{-1}] p_j : protein concentration of j [g L^{-1}] τ_j : turnover time of j [h]
Mass conservation constraint	$Mv = \mu c$, Where $M = \begin{matrix} & S & PSU & Met & R \\ \begin{matrix} C \\ e \\ AA \\ p \end{matrix} & \begin{bmatrix} 1 & 0 & -0.8 & 0 \\ -0.018 & 1 & -0.2 & -0.03 \\ 0 & 0 & 1 & -0.97 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix}$
Reaction kinetics constraint	$p_j = v_j \tau_j(c, a)$ where $\tau_j(c, a) = \frac{1}{k_{cat}^j} \prod_m \left(\frac{c_m}{K_m^j + c_m} \right)^{-1} \prod_n \left(\frac{a_n}{K_n^j + a_n} \right)^{-1} \left(\frac{I_n^j}{I_n^j + a_n} \right)^{-1}$
Total protein constraint	$c_p = \sum_j p_j$
Cellular density constraint	$\rho = c_p + \sum_m c^m$

3.2 Results

Base GBA model: a simple cyanobacterium model including photoinhibition

We first developed a base GBA model of *Synechocystis* based on the (Faizi et al., 2018) model, with few modifications that simplify its mathematical description while retaining its key biological properties (see **Figure 1** for a schematic representation of our base GBA model, with **Table 1** listing its details). For any given GBA model, a simple set of equations determining the balanced growth problem can be derived from first principles (Dourado et al., 2023). These equations are based solely on the mass balance of reactions within the matrix $\mathbf{M}$, as well as kinetic parameters and cell density data. The parameters used in our model are equivalent to those used in the *Synechocystis* model proposed by Faizi et al. (2018), except for those pertaining to photosynthesis and diffusion. In contrast to the Faizi model, the GBA framework does not account for proteins as substrates of reactions. We thus reformulate photosynthesis as a simple transport reaction “PSU” that imports energy “e” into the system, a mathematically equivalent approach that preserves the original model’s predictions. We also simplify the Faizi model by ignoring carbon passive diffusion through the cell membrane, only accounting for an active import of carbon by the protein transporter “S”. Due to the lack of corresponding experimental data, the physiological relevance of passive carbon uptake is currently unclear. While the Faizi model includes both passive and active carbon uptake, its optimal solution under high external carbon concentration relies exclusively on diffusion, resulting in zero investment into the carbon transporter. In contrast, our GBA model predicts sustained investment in active carbon transport by excluding passive diffusion.

We posit that the main effect of the photodamage of proteins is to place an additional burden of protein production on the ribosome, as damaged photosynthetic units are degraded into amino acids and need to be replaced. To model the corresponding photoinhibition of growth, we thus incorporate an inhibition term into the general kinetic rate law (Liebermeister & Klipp, 2006) describing the ribosome. Accordingly, the light inhibition in our GBA model is quantified by an inhibition constant I_L^r for the ribosome reaction r . The external light intensity x_L then modulates the inhibitory effect defined by this constant (I_L^r). Faizi et al. (2018) adjusted three photosynthesis parameters in their extended model to minimize the distance between experimental and simulated growth rates. We do not use these fitted values; instead, we simply set the inhibition constant I_L^r to a value that ensures that the optimal growth rate of the model coincides with the last point for the experimental growth rate at $1100 [\mu\text{E m}^{-2} \text{s}^{-1}]$ (100% light intensity).

In an environment $a = (a_C, a_L)$, defined by the external concentration of carbon a_C (not distinguishing between CO_2 and HCO_3^-) and the light intensity a_L , the optimal cellular state is the

set of flux fractions that maximizes the balanced growth rate μ under the constraints of mass conservation, reaction kinetics, and fixed cell density.

We note that the Faizi model employs molar units, while in GBA, all equations are mass-normalized, which simplifies calculations and promotes consistency. Mass normalization facilitates the comparison of theoretical calculations with experimental data by providing a common unit of measurement, reducing the need for complex unit conversions. Thus, we utilized molecular weights of metabolites to normalize the mass fraction matrix) as well as the Michaelis constants (K_m^j) and turnover numbers (k_{cat}^j) of the enzymes. Specifically, each K_m^j value in [mol L⁻¹] was multiplied by the molecular mass of the corresponding reactant in [g mol⁻¹], resulting in units of [g L⁻¹]. To convert turnover numbers, note that these are usually given in mole of product per mole of enzyme per unit time. Thus, the k_{cat}^j values in [h⁻¹] were multiplied by the molecular mass of the product and divided by the molecular mass of the catalyzing enzyme, resulting in a different numeric value in the same units, [h⁻¹].

To validate our *Synechocystis* base GBA model, we compared its growth rate predictions across different light intensities to the Faizi model on which it is based. **Figure 2** shows the predicted growth curves in light-limited (I), light-saturated (II), and light-inhibited (III) conditions for high (corresponds to 425 ppm of CO₂ in the air) and low concentrations (in relation to K_m of carbon transporter S) of external inorganic carbon. The predictions of the mathematically simpler GBA model (blue line) closely align with the Faizi model predictions (in red) across the entire range of growth rates. Additionally, the predictions showed robust agreement with the experimentally measured growth rates (black diamonds), which were used as the basis for fitting the parameters of the original Faizi model. Across the light-limited and the light-saturated regime, both the Faizi and GBA model similarly predict an increase in the growth rate of *Synechocystis*, reaching a maximum of 0.108 h⁻¹ (equivalent to a doubling time of 6.41 h). Subsequently, transitioning from the light-saturated to the light-inhibited growth phase, the growth rate begins to decline to 0.093 h⁻¹ (corresponding to a doubling time of 7.45 h) as excessive light inhibits the photosystem unit, resulting in a lower photosystem reaction rate that decreases the growth rate. Mechanistically, both models elucidate this phenomenon, highlighting the role of excessive light in inhibiting photosystem function and consequently impeding growth dynamics.

The two models differed in their predictions for the maximum growth rates at a low carbon concentration, defined as 5% of the high saturating concentration in relation to K_m of carbon transporter S (**Figure 2**). As indicated by the proteome allocation analysis (see below), this difference can be attributed to the presence of two carbon transporter mechanisms (active and passive) in the Faizi model, as opposed to the single active carbon transporter in the GBA model.

We compared the proteome allocation patterns predicted by the two models under different light intensities at high external carbon concentration, finding qualitative agreement (**Figure 3**). The base GBA model (blue) and Faizi model (red) both capture the overall empirical trends (black diamonds). Under conditions of light limitation, both models assign the majority of protein mass

to the photosynthetic unit (0.99 of the proteome fraction at the lowest light intensity), with a subsequent decrease in allocation as light intensity increases, thereby allowing the allocation of protein mass to other processes and hence increasing growth rate. Accordingly, the proteome fractions associated with metabolic enzymes and ribosomal reactions show an increase in line with the growth rate, a trend observed both in the experimental data and in both models.

Notably, while the Faizi model includes carbon translocation both through diffusion and through active transport, the optimal model solution at high external carbon concentration only uses diffusion. This results in a zero investment into carbon transport proteins (**Figure 3C**).

Unfortunately, experimental data is unavailable for the cellular investment into carbon transport, so it is unclear if this pattern is biologically realistic.

Overall, both models fail to capture absolute proteome fractions quantitatively, suggesting that some important processes or features were not considered adequately in the models' construction and/or parameterization. Refined models should integrate a richer set of parameters to reflect complex biological realities. Any such expansion of the Faizi model would necessitate a complete rewriting of its equations and the re-fitting of its parameters. In the GBA framework, in contrast, any model extension requires only the inclusion of new columns and rows in the matrix M , and a corresponding inclusion of kinetic parameters K_m , k_{cat} for the new reactions. We next exploit this inherent extensibility to develop a more detailed and comprehensive GBA model of *Synechocystis*.

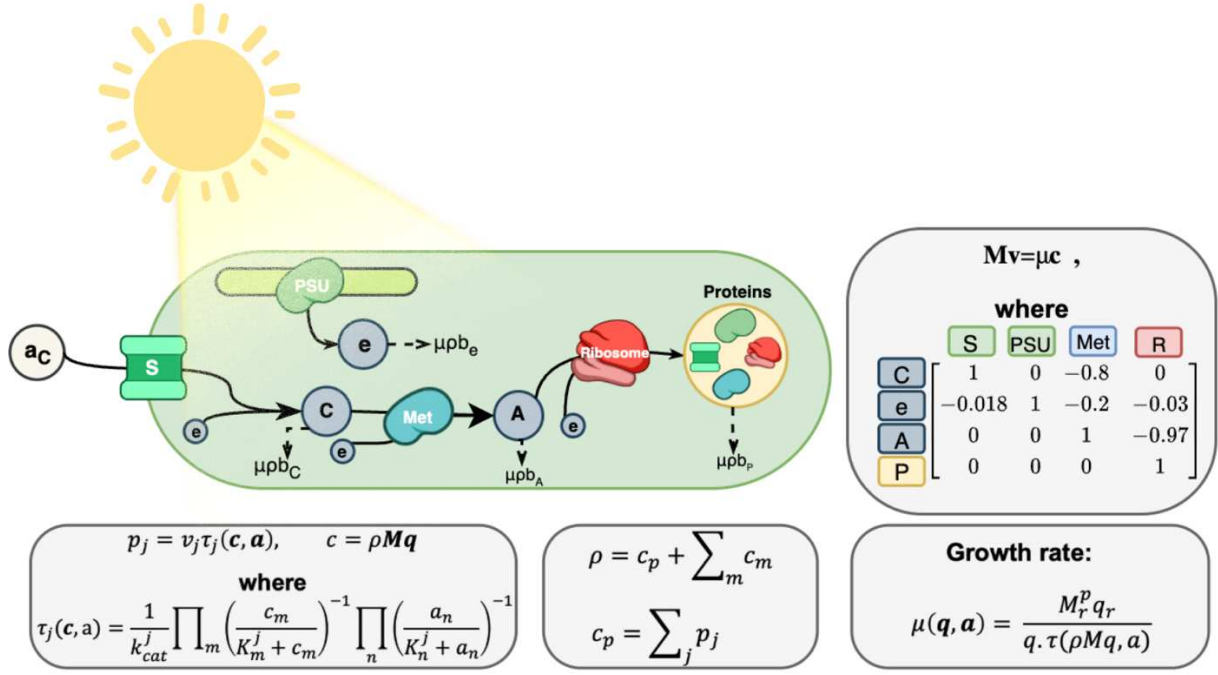


Figure 1. Overview of the base GBA model for *Synechocystis*. External “metabolites” (external inorganic carbon C and light e) are imported by the “transporters” S and PSU, respectively. A metabolic reaction “Met” converts the internal C, e into amino acids “AA”, which are used by the ribosome “R” to produce all protein “p” in the model. The protein “p” is assumed to be instantly distributed into the four reactions (PSU, S, Met, R) such that their protein catalyst maintains its concentration despite its dilution by balanced growth. Mass conservation of internal reactants (C, e, AA) at balanced growth is enforced by the equation $Mv = \mu c$, relating fluxes v , growth rate μ , and internal mass concentrations $c = (c_C, c_e, c_{AA}, c_p)$: the net production of each reactant, Mv , must balance its dilution by growth, μc . Each reaction j is catalyzed by a specific protein with concentration p_j and a turnover rate $\tau_j(c, a)$, which is determined by kinetic rate laws and depends on the internal concentrations c and external concentrations $a = (a_C, a_L)$ of reactants involved in the reaction. The fixed cellular density ρ constrains the sum of all metabolite concentrations c_m and total protein concentration c_p , which is itself defined as the sum of all individual protein concentrations p_j . The growth rate optimization problem in a given environment $a = (a_C, a_L)$ can be entirely formulated on flux fractions $q = v/\mu\rho$, greatly simplifying analytical and numerical studies (Dourado et al., 2023).

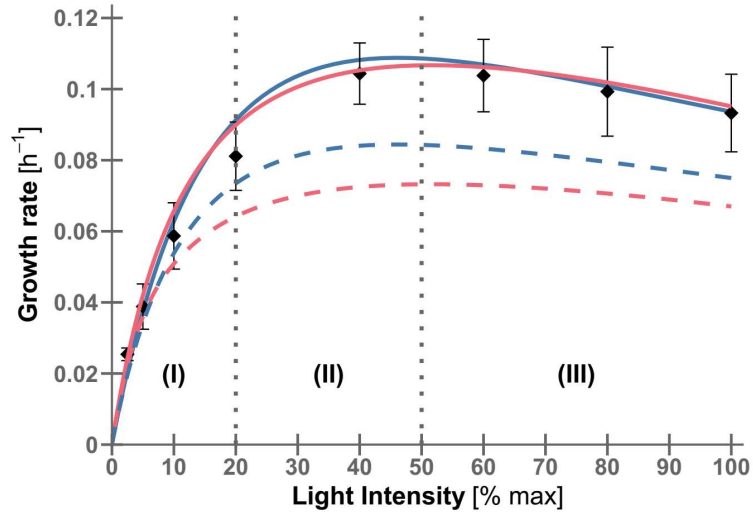


Figure 2. Simulated growth curve of *Synechocystis* with photoinhibition. The blue line shows the growth rate estimated for *Synechocystis* with the GBA model under different light intensities at two different levels of external inorganic carbon: high (solid line) or low (dashed line, 5% of high concentration). The red lines show the corresponding simulation results using the Faizi model. Three different growth regimes are delineated: light-limited (I); light-saturated (II); and light-inhibited (III). The experimentally observed steady-state growth rates for *Synechocystis* (black diamonds) (Faizi et al., 2018) are well explained by both models.

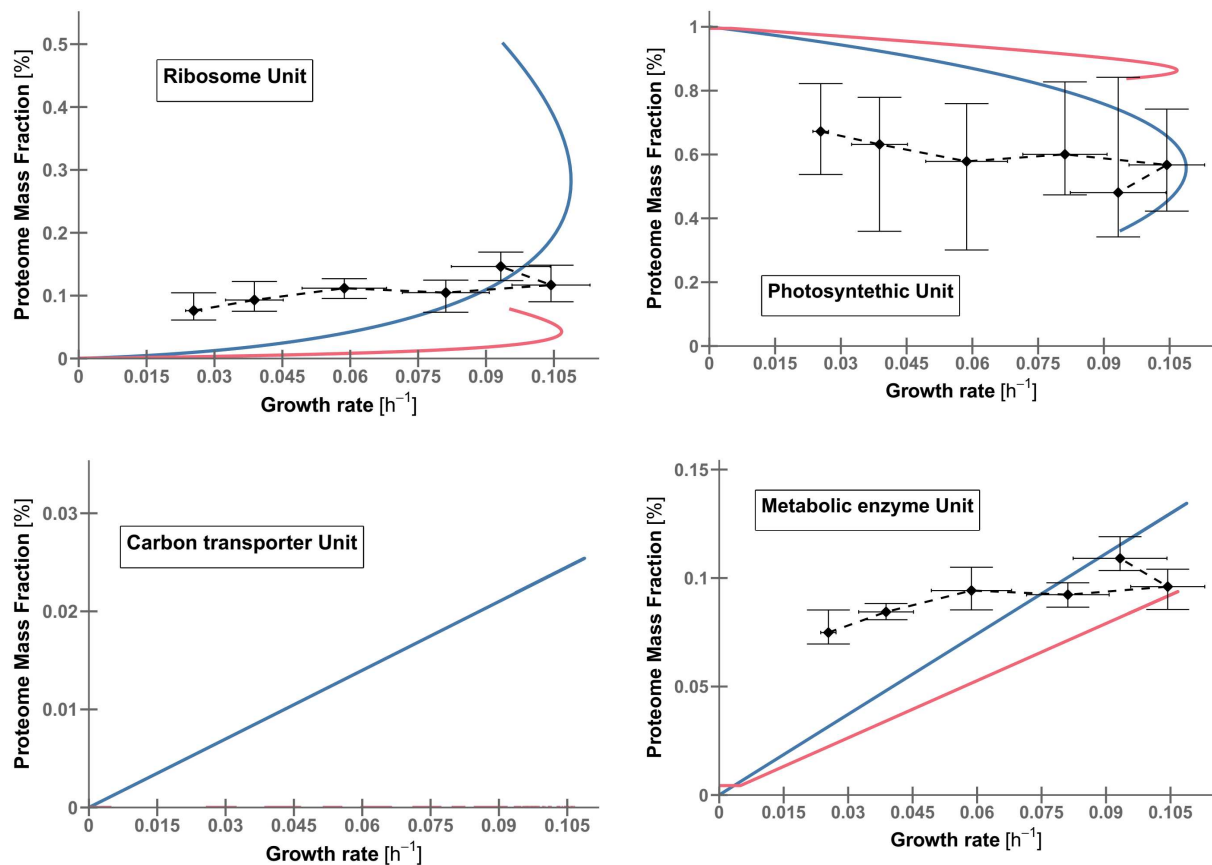


Figure 3. The predicted proteome allocation in the GBA model (blue) and in the Faizi model (red) at high external carbon concentration, including the effect of photoinhibition. The black diamonds are experimental proteomics data (Zavřel et al., 2019).

Development of an extended GBA model for examining photosynthesis components in cyanobacteria

Qualitatively, our base GBA model replicated major trends observed in the basic phototrophic growth behavior of cyanobacteria, including the main patterns of proteome allocation across growth rates. As light intensity increases, the GBA model predicts a decrease in the proteome fraction allocated to the photosynthesis unit. Photosynthesis is a fundamental biological process in cyanobacteria that has a profound impact on their overall physiology, and its complexity extends well beyond the boundaries of the mere light harvesting reactions represented in our base model and the Faizi model. In biological reality, photosynthesis interweaves various reactions and pathways to form a complex network of interrelated processes. It comprises not only the light-harvesting components of Photosystems I and II, but also the dynamic interplay of Cytochrome b6 and ATP Synthase. Any realistic model of cyanobacteria must thus represent at least four major light reactions of photosynthesis to convert light energy to chemical energy.

Additionally, a more realistic model must include at least a coarse-grained, quantitative description of carbon fixation, which should represent the conversion of atmospheric carbon dioxide into organic compounds and the synthesis of amino acids, ribosomal RNA, ribosomal proteins, and other cellular components essential for growth. **Figure 4** shows the mass fraction matrix M of an expanded GBA model that includes these additional subsystems.

	Carbon_T	PSII	PSI	Cytb6f	ATPsyn	C_fix	Met	Ribosome
CO ₂	1	0	0	0	0	-0.9	0	0
H ⁺	0	0.02	0	0.01	-1	0	0	0
PC	0	0	0.99	-0.7	0	0	0	0
PC ⁻	0	0	-0.9	0.69	0	0	0	0
PQ	0	-0.9	0	0.3	0	0	0	0
PQH ₂	0	0.98	0	-0.3	0	0	0	0
ATP	-0.01	0	0	0	1	-0.05	-0.05	-0.3
NADPH	0	0	0.01	0	0	-0.05	-0.05	0
C ₃	0	0	0	0	0	1	-0.9	0
A	0	0	0	0	0	0	1	-0.7
Protein	0	0	0	0	0	0	0	1

Figure 4. The mass fraction matrix for the extended GBA model, expanding the base model by including new reactions and new metabolites. Compared to the base model, the photosynthetic unit was extended to comprise Photosystem II (PSII), Photosystem I (PSI), Cytochrome b6f (Cytb6f), and ATP synthase (ATPsyn), and the metabolic reactions were extended to comprise carbon fixation and metabolism.

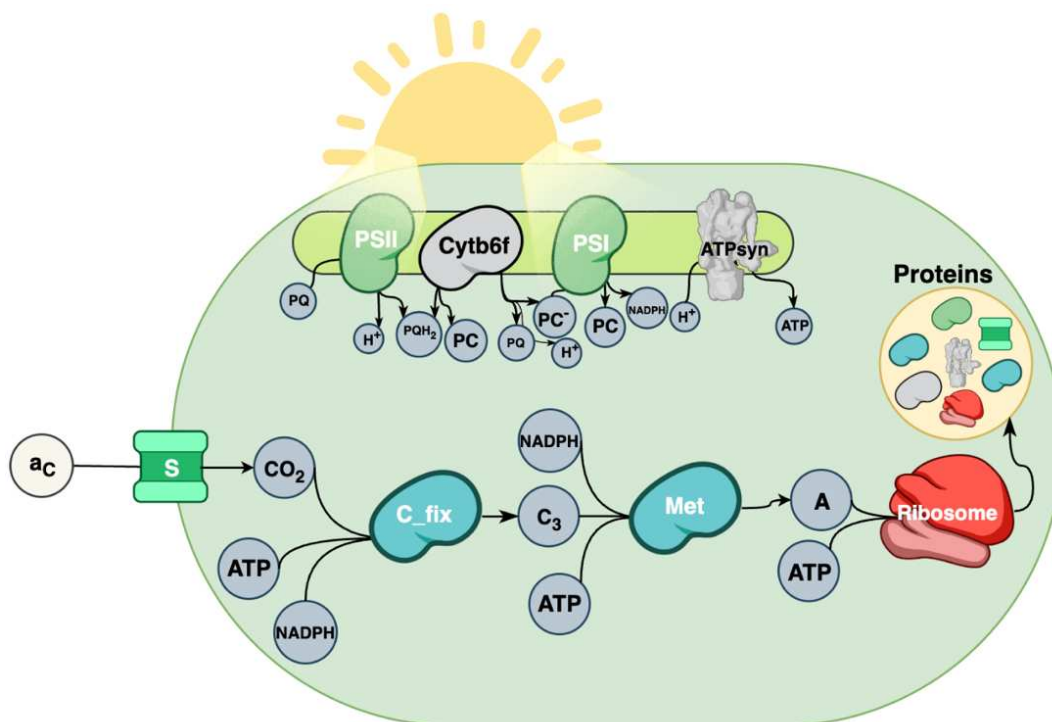


Figure 5. Overview of the extended GBA model for *Synechocystis*. The model encompasses 8 reactions and 11 metabolites. Photon absorption initiates in Photosystem II (PSII) and Photosystem I (PSI), triggering plastoquinone (PQ) reduction and plastocyanin (PC) oxidation. PQ and PC (in both reduced and oxidized forms) act as electron carriers, facilitating electron transfer among photosynthetic complexes, including Cytb6f, via reversible reactions. This process drives efficient energy flow and NADPH generation through PSI. ATP, a crucial energy carrier, is replenished by the ATPase complex. Inorganic carbon uptake and assimilation into organic compounds are unified, without distinguishing between CO_2 and HCO_3^- . The carbon fixation cycle is represented by a single reaction, yielding 3-phosphoglycerate (C3), a central metabolic precursor for subsequent biosynthesis reactions. C3 undergoes further conversion into amino acids and proteins via metabolism and ribosome reactions.

The model summarized in **Figure 5** consists of 8 reactions and 11 metabolites, drawing inspiration from a previously published, simplified metabolic network of *Synechocystis* (Rugen et al., 2015), with adjustments made to make the formulation consistent with the GBA framework. In this model, the absorption of photons occurs in the components associated with Photosystem II (PSII) and Photosystem I (PSI), leading to the reduction of plastoquinone (PQ) and oxidation of Plastocyanin (PC). Plastocyanin and Plastoquinone (in both reduced and oxidized states) serve as vital electron carriers, shuttling electrons between the various photosynthetic complexes, including Cytb6f through a reversible reaction, thereby facilitating the

efficient flow of energy and generation of NADPH via photosystem I (PSI). ATP, an essential energy carrier molecule, is regenerated by the ATPase complex. For simplicity, this model does not account for the oxidized states of the cofactors, specifically ADP and NADP^+ ; to avoid the apparent generation or vanishing of mass, we assume that the molecules representing the reduced state (ATP, NADPH) have masses corresponding to the difference in molecular weight between the oxidized and reduced states. Similar to our base model, inorganic carbon is taken up and assimilated into organic compounds, with no distinction made between CO_2 and HCO_3^- . The carbon fixation cycle is represented by a single reaction, yielding the 3-carbon molecule 3-phosphoglycerate (C3) (same as in Ref. (Rugen et al., 2015)). C3 serves as the central metabolic precursor for subsequent biosynthesis reactions, undergoing further conversion into amino acids and proteins via the Metabolism and Ribosome reactions.

The extended GBA model provides a detailed prediction of the proteome allocation into 8 different core processes at different light intensities, compared to only four processes represented in the base model. The kinetic parameter values (K_m and k_{cat} values) were obtained from the BRENDA database (Schomburg et al., 2013) when available, otherwise we directly sourced them from our base model for shared reactions (see Methods, “Parameters of GBA models for cyanobacteria”). Note that we did not introduce any more adjustable parameters; we still only adjusted the inhibition constant I_L^r to ensure that the model prediction for the highest light intensity matches the experimental value.

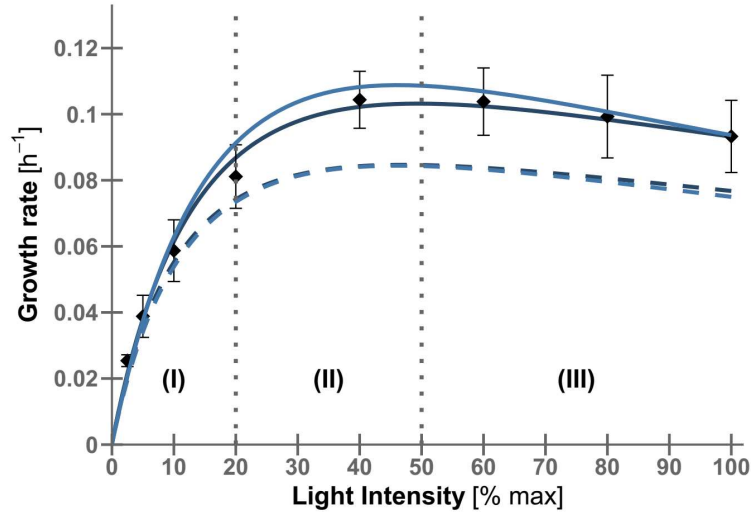
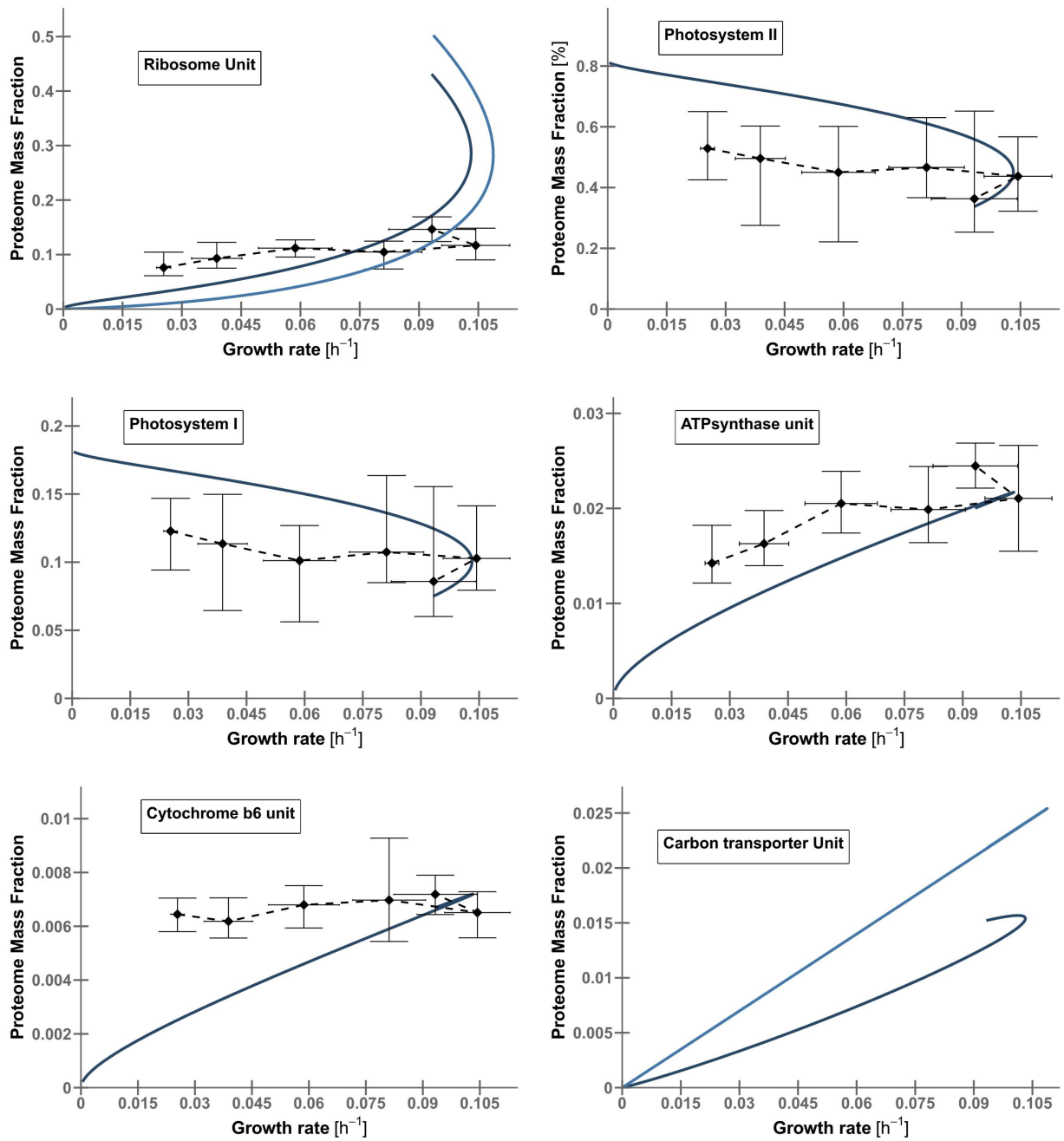


Figure 6. Growth curves of *Synechocystis* in the extended GBA model (dark blue) and base GBA model (light blue) at different light intensities at high (solid lines) or low (dashed lines, 5% of high concentration) external inorganic carbon concentrations. Three different growth regimes are depicted: light-limited (I), light-saturated (II), and light-inhibited (III), along with experimental steady-state growth rate for *Synechocystis* (black diamonds) (Faizi et al., 2018).

Across light intensities, the growth rates predicted with the extended model are highly consistent with the experimentally observed growth rates (**Figure 6**); the model extension slightly improves the agreement between predictions and observations compared to the base GBA model, despite no additional adjustable parameters. As light intensity increases from the light-limited to the saturated regime, the predicted cellular investment of two photosynthetic components (ATPSynthase and Cytochrome b6) behaves markedly different from the base model (**Figure 7**). The light harvesting components, represented by the PSI and PSII proteome sectors, decrease as growth rate increases until both reach the light-saturated level. Conversely, the protein allocation to the ATPSynthase and Cytochrome b6 units increases with increasing light intensity up to this level. The ribosome proteome allocation follows an almost identical pattern as for the base model. From the light-saturated to the light-inhibited level, a downward kink is observed in the light-harvesting sectors of PSI and PSII. The PSI and PSII proteome sectors continue to decrease as growth rate decreases. Meanwhile, ATPSynthase and Cytochrome b6, which increase with growth rate from the light-limited to light-saturated level, experience a reduction as they enter the light-inhibited level. Thus, in accordance with previous empirical observations in proteomics studies of *Synechocystis* under various growth conditions (Jahn et al., 2018; Zavřel et al., 2019), the cellular investment into the light harvesting sectors decreases with increasing light intensity, while investment into ATPSynthase and Cytochrome b6 increases (**Figure 7**). Furthermore, the model reproduced the experimentally observed dependence of total protein concentrations on the growth rate, demonstrating a reduction in overall protein content as the growth rate increased

(Figure 7); this reduction is likely due to photoinhibition within the cell, where damaged proteins are degraded into amino acids.



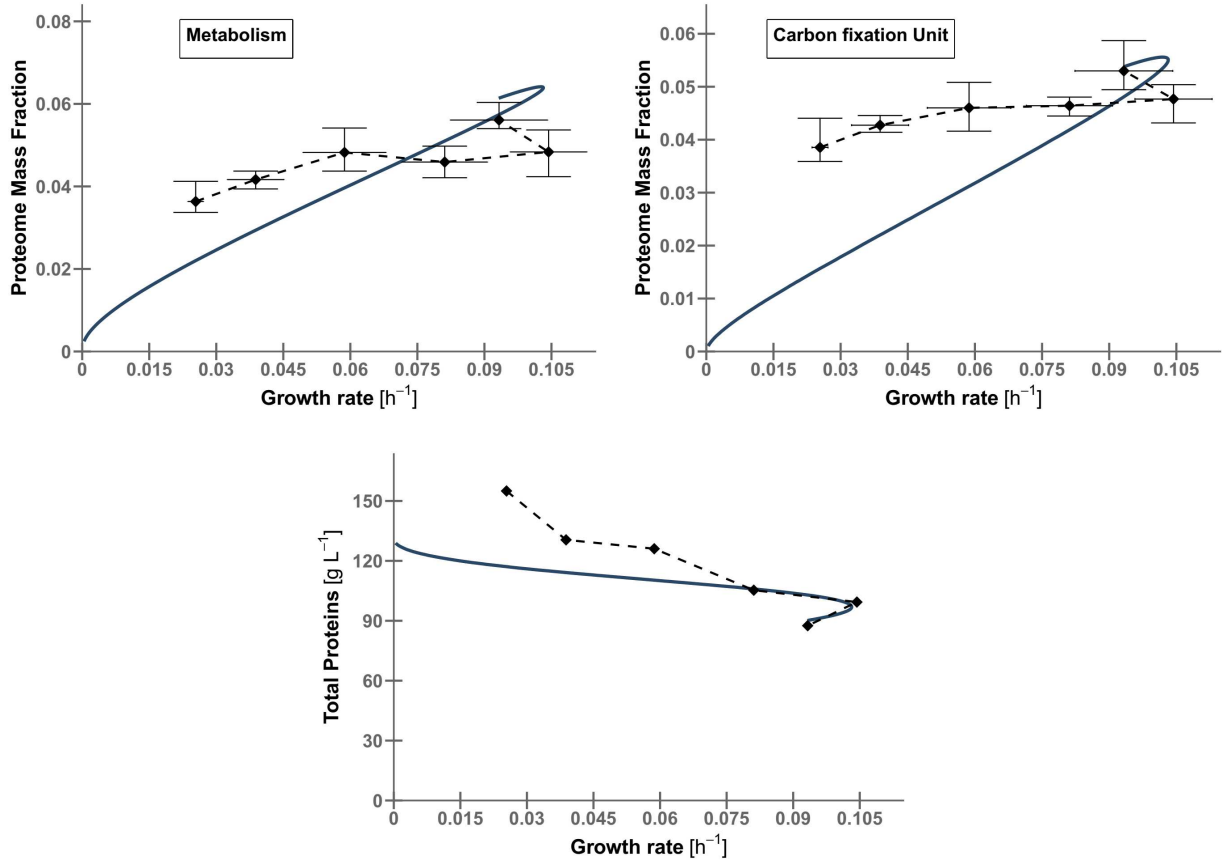


Figure 7. *Synechocystis* resource allocation predicted by the extended GBA model (dark blue line) compared to proteomics data at high external inorganic carbon availability. For comparison, the light blue lines show the simulation results from the base GBA model for the reactions common to both models. The data points are experimental proteomics data, obtained from (Zavřel et al., 2019) (black diamonds). Experimental data for the carbon transporter unit was not available.

3.3 Discussion

As seen in (Faizi et al., 2018), kinetic models are powerful tool to understand global patterns of cellular resource allocation in cyanobacteria. Above, we have combined insights from this work with theoretical advances in balanced growth modeling (Dourado et al., 2023) to construct a base GBA model of cyanobacterial physiology. This new model explains important aspects of cyanobacterial resource allocation directly from first principles.

To extend our understanding of the optimal proteome allocation in *Synechocystis*, we further used this base model as a platform to develop an extended GBA model that incorporates the four key proteins of the photosynthetic light reactions. Proteome allocation predictions from the

extended GBA model are highly consistent with experimental findings (Zavřel et al., 2019), capturing both the mean levels of proteome allocation and the growth rate-dependent trends (**Figure 7**). Specifically, as the growth rate transitions from a light-limited to a light-saturated regime, the extended GBA model predicts a decrease in the proteome sectors of PSI and PSII and an increase in ATP synthase and cytochrome b6 levels. This reflects the increased demand for ATP caused by photodamage in high-light conditions and the need for new protein synthesis, in line with the experimental results. Across the modeled cellular components, a comparison of observed and predicted proteome allocation patterns across growth rates generally shows good quantitative agreement at high growth rates. In contrast, with the exception of photosystems I and II, the model predicts a smaller proteome mass fraction than experimentally observed at low growth rates. It has been argued that some bacteria maintain a proteome reserve at low growth rates to enable the quick resumption of growth once conditions improve (Mori et al., 2017) such proteome reserves could plausibly also have evolved in cyanobacteria, potentially explaining the excess of observed over predicted proteome allocation into cellular components essential for growth.

Within the GBA modeling framework, all proteins are explicitly synthesized by the ribosome, including the ribosome itself. GBA can predict detailed, environment-dependent cell composition through constrained optimization of growth; it ensures the optimal utilization of all resources and optimal concentrations of all intracellular components, including proteins and metabolites (Dourado et al., 2023). This contrasts with linear modeling frameworks that do not consider the explicit synthesis of enzymes and the ribosome; instead, these alternative approaches either substitute these processes with a constant biomass reaction (Sánchez et al., 2017) or, in cases where explicit synthesis is considered, they replace concentration-dependent, non-linear enzyme kinetics with a constant catalytic rate (k_{cat}) and ignore the dilution of metabolites (Goelzer et al., 2011; O'Brien et al., 2013). The extended GBA model replicated the observed decrease in overall protein content with increasing growth rate, highlighting its ability to dynamically optimize resource allocation, in contrast to previous frameworks that relied on fixed total protein concentrations (Goelzer et al., 2011; Sánchez et al., 2017).

By providing a minimal and transparent framework that is biologically and biophysically meaningful, we can relate the resource allocation between major protein pools and growth rate under different nutrient availabilities. Our GBA model provides basically the same results as the differential-equation based approach pursued by (Faizi et al., 2018). Both approaches have in common that the steady-state of the system, given a particular proteome allocation, is calculated, and that this function (defining the growth rate as a function of the proteome) is used to find the proteome allocation which maximizes growth rate. However, whereas in (Faizi et al., 2018) the steady-state is calculated by solving a system of differential equations, the GBA approach is completely analytical, which allows formulating a single non-linear and non-convex optimization, solving which results in the optimal proteome allocation. The advantage of the concise mathematical formulation of GBA models is that it facilitates efficient modeling of

balanced cellular growth in general and thus makes extending the model by including more or more detailed processes conceptually straight-forward.

The framework developed above is general and derived from basic principles. While we have constructed and simulated our model around the data of Faizi for *Synechocystis*, the framework can be easily applied to other cyanobacteria, requiring only an adjustment of parameter values (such as kinetic parameters and cellular density) to account for alternative resource allocation strategies. Despite the simplicity of our assumptions, the GBA model not only exhibits comparable predictive capabilities to the coarse-grained model proposed by Faizi et al. (2018), but its streamlined mathematical formulation also offers substantial advantages for the construction and numerical solution of larger models of cyanobacteria. Recent studies suggested that *Synechocystis* as a model organism can introduce novel products in biotechnology and as a potential microbial cell factory (Blanc-Garin et al., 2022; Yu et al., 2013). Thus, GBA models of this organism can provide a new tool to study the direct conversion of CO₂ and light to value-added chemicals and fuels, contributing to the new field of blue bioeconomy.

3.4 Materials and methods

The growth balance analysis framework

In the growth balance analysis, the optimization problem is defined as finding the optimum growth rate (μ) subject to non-negativity constraints on metabolite and protein concentrations by varying the flux fraction $q := v/\mu\rho$. Moreover, the balanced growth model at steady-state is specified by the following constraints:

$$\begin{aligned} \sum_j M_j^i v^j &= \mu c^i && \text{(Mass conservation)} \\ v^j &= \frac{p^j}{\tau^j(c, a)} && \text{(Reaction kinetics)} \\ \sum_j p_j &= c^p && \text{(Protein density)} \\ \sum_i c^i &= \rho && \text{(Cellular density)} \\ c^m &\geq 0 && \text{(Non-negativity of metabolite concentration)} \\ p^j &\geq 0 && \text{(Non-negativity of protein concentration)} \end{aligned}$$

It is noteworthy to mention that the normalization of M results in the expression of protein concentrations (p_j) and reactants (c^i) in units of [g L⁻¹]. Accordingly, fluxes ([g L⁻¹ h⁻¹]) and kinetic parameters must be represented in mass units. For instance, Michaelis constants (K_m) are

expressed in $[g\ L^{-1}]$ and turnover numbers (k_{cat}) are represented as the amount of product per unit of protein per unit of time, resulting in units of $[h^{-1}]$. The cell density (ρ , $[g\ L^{-1}]$) is defined as the sum of all metabolite and protein concentrations, which is assumed to be constant. The comprehensive explanation of the growth balance analysis framework is provided in the original publication by (Dourado & Lercher, 2020), along with its supplementary material.

Construction of a base GBA model of *Synechocystis* sp. PCC 6803 based on the model by Faizi

The cyanobacterial cell model presented in this study builds upon Faizi's model in the growth balance analysis framework, but has been adapted and expanded to incorporate experimental proteomics data and better reflect the realistic characteristics of a cyanobacterial cell.

We take from the Faizi model the following set of parameters: I) stoichiometric coefficients, derived from the genome-scale reconstruction detailed in (Knoop et al., 2013); II) protein lengths, sourced from the Uniprot database; III) cell density, inferred from experimental data on cell dry weights and cell counts; IV) Michaelis constants K_m , estimated for energy and amino acids, while data for carbon transporter and metabolism reactions were collected from literature; V) turnover numbers k_{cat} of reactions, obtained from literature with exception of diffusion and photosynthesis; VI) photosynthesis turnover rate.

The normalized mass stoichiometry of the model is defined as the stoichiometric matrix S , containing rows for reactants, is multiplied by the respective molecular mass. Then, we normalized the column so that the sum of negative entries is equal to -1 and the sum of positive entries is equal to +1 preserving mass conservation in reactions (Dourado et al., 2023). To determine the mass of protein classes (PSU, S, Met, R) in our model, we used the reference proteome of *Synechocystis* from UniProt (Bateman et al., 2017). Each protein was mapped according to its protein class from (Faizi et al., 2018; Zavřel et al., 2019) to ensure consistency in data comparison. The kinetic data (K_m and k_{cat} values) were sourced from previously published model of (Faizi et al., 2018), as indicated in the parameters section of the methods, and then converted into $[g\ L^{-1}]$ for K_m and $[h^{-1}]$ for k_{cat} . It is noteworthy that unlike GBA, Faizi's model does not incorporate molecular weights in their formulation due to their molar $[M]$ units. Therefore in our model, molecular weights were estimated based on simple metabolites, participating in each reaction. Overall, the model encompasses 6 metabolites and 4 reactions, and parameters used in the base model can be found in **Table S1**.

Description of the extended model

The extended model is created following our implementation of cyanobacterial phototrophic growth. In simple terms, the model uses two inputs: a carbon source (with no distinction between CO_2 and HCO_3^-) and light, which serves as the energy source. The light is absorbed by the photosystem II and photosystem I light harvesting complexes, allowing the production of energy through ATP synthase and the electron transport chain Cytochrome b6 via a reversible reaction. The carbon source is taken in through a carbon transporter, and incorporates it into

organic molecules through the process of carbon fixation, which is performed by the Calvin-Benson cycle. These are then utilized in the metabolic and also ribosome enzymes during protein translation. The kinetic data (K_m and k_{cat} values) were obtained from the BRENDA (Schomburg et al., 2013) database and then converted into $[g\ L^{-1}]$ for K_m and $[h^{-1}]$ for k_{cat} . For each enzyme class, we queried the BRENDA database using the enzyme commission (EC) number of reactions in each class to find the value of the wild-type enzymes. Whenever possible, we preferred values from *Synechocystis* or other cyanobacterial cells, otherwise we directly sourced them from previously published model (Faizi et al., 2018), as indicated in the parameters section of the methods. Moreover, the Michaelis constant K_m of each protein category were calculated based on the availability of data, multiplied by its respective molecular weight and total number of reactions present within the proteome sector. Similarly, the turnover number k_{cat} for each enzyme was estimated by multiplying it by the total number of reactions catalyzed in each category. The model encompasses 11 metabolites, 8 reactions, and it is noteworthy that even upon expansion, the original formulation remains unchanged.

Model implementation

In this study, we used growth balance analysis (GBA) to simulate growth and resource allocation in cyanobacteria. R v4.1 programming language was used for implementation of balanced growth optimization problem using NLOpt library. The optimization problem is solved through AUGLAG, an augmented Lagrangian approach that utilizes method of moving asymptotes (MMA) or sequential quadratic programming (SLSQP) algorithms as the sub-solver, which is a free and open-source software for nonlinear optimization. The models are presented in the Excel format in **Supplementary File 1**, and the R script necessary for running the simulations can be obtained from the (Dourado et al., 2023).

Experimental proteomics data

The present study utilized experimental proteomics data originally obtained by (Zavřel et al., 2019), wherein the quantification of absolute protein abundance was carried out under varying light intensities. To determine the proteome fractions corresponding to each protein class within our model, we aggregated the protein quantities within each class and subsequently normalized them by the sum of all protein quantities at every given light intensity. This allowed us to integrate the obtained data with the relevant proteome sectors within our model.

Parameters of GBA models for cyanobacteria

Stoichiometric coefficients

We started from the stoichiometric matrix (S), described by Faizi to construct our normalized mass fraction (M) in GBA model. Moreover, for the construction of matrix M, we calculated the entries based on the stoichiometric matrix (S) of (Faizi et al., 2018), and normalized each entry based on its defined molecular weights, in such a way that in a reaction, the sum of all substrates with negative entries results in -1 and the sum of all products with positive entries results in +1.

Molecular mass of protein classes

In order to ascertain the mass distribution of protein classes within our model (S, PSU, Met, R, for the based model, and PSI, PSII, Cytb6f, ATPsynthase, carbon_transporter, carbon fixation, metabolism, and Ribosome, for the extended model), we employed the reference proteome of *Synechocystis* obtained from UniProt. Protein classes for the base model were obtained from the study conducted by (Zavřel et al., 2019), while protein classifications as described in (Faizi et al., 2018) along with corresponding Uniprot IDs, were utilized to expand our model to incorporate the relevant proteins. Further elaboration on the characteristics and attributes of each protein class can be found in **Supplementary Table 1**.

Molecular mass of metabolites

For the molecular masses of metabolites, we estimated 1 [g mol⁻¹] for “e” (corresponding to molecular mass of H⁺), 52.5 [g mol⁻¹] for “ci”(corresponding to mean molecular mass of HCO₃ and CO₂), 36 [g mol⁻¹] for “c3” (corresponding to three times of molecular mass of carbon), and 132 [g mol⁻¹] for “aa” (as the average molecular mass among all amino acids(Schmidt et al., 2016)).

Kinetic parameters

We sourced the kinetic parameters directly from (Faizi et al., 2018) and then performed the conversion of k_{cat} and K_m to mass units in the following way: i) All K_m were converted from [mol L⁻¹] to [g L⁻¹] by multiplying the original value with the corresponding molecular mass (see previous paragraph), ii) All k_{cat} were converted to mass units by multiplying the original value by product molecular mass, and then dividing by the mass of enzyme (see **Supplementary Table 1**).

Carbon transporter

The turnover number (k_{cat}) of carbon transporter in *Synechocystis*, which we obtained from Faizi model is 45360 [h⁻¹], which by mass normalization resulted in 8.434 [h⁻¹]. Besides, the Michaelis constant (K_m), 15 [μM] is also obtained from Faizi model, resulting in 0.0007875 [g L⁻¹] by mass normalization.

Ribosome

The turnover number of ribosome was adapted from Faizi model, 22 [s⁻¹], which is the highest elongation rate observed experimentally in *E. coli* (Bremer & Dennis, 2008), and the Michaelis constant for amino acids and energy is estimated as 100000 [molec cell⁻¹] in (Faizi et al., 2018), the ribosome molecular weight was measured 2,306,967 Da, which by mass normalization resulted, $k_{cat} = 22 \text{ AA s}^{-1} \times (132.60 \text{ Da AA}^{-1}) / (2,306,967 \text{ Da}) \times 3600 \text{ s h}^{-1} = 4.55 \text{ h}^{-1}$ and $K_m = 10000 \text{ molec cell}^{-1} \times (1 \text{ cell} / 2.24 \times 10^{-14} \text{ L}) \times (1 \text{ g} / 6.022 \times 10^{23} \text{ mol}) = 7.41 \times 10^{-7} [\text{g L}^{-1}]$ for “e” and similarly $9.78 \times 10^{-5} [\text{g L}^{-1}]$ for “aa”. We note that similar to (Faizi et al., 2018) $K_m = 7.41 \times 10^{-7} [\text{g L}^{-1}]$ was set for energy unit “e” in all reactions for simplicity.

Metabolism

The turnover number of metabolism for *Synechocystis* in Faizi model was reported 20 [s⁻¹] (Marcus et al., 2005), and the K_m value for internal carbon was 0.00018 [M], which by converting to mass units results in 0.794 [h⁻¹] and 0.00648 [g L⁻¹], respectively.

Photosynthesis

The turnover number of photosynthetic unit is estimated 200 [s⁻¹] as reported in (Milo & Phillips, 2015), which by converting to mass units results in 0.0847 [h⁻¹]. The K_m value for external light intensity was set to 440 [$\mu\text{E m}^{-2} \text{s}^{-1}$], corresponding to the half of intensity in light-saturated condition. The only parameter fitted was inhibition constant I_L^r with the value of 47 [$\mu\text{E m}^{-2} \text{s}^{-1}$]. The inhibition constant fitted in a way that at light intensity of 1100 [$\mu\text{E m}^{-2} \text{s}^{-1}$] (100% light intensity), the growth rate of the model fits to the last point in experimental growth rate.

Extended model

In this model, we followed the same procedure for calculating our model parameters. Here, we categorized the protein classes based on the new reactions in the model, and the parameters used in the extended model can be found in **Table S2**. Similarly, the detailed description of each protein class is provided in **Supplementary Table 2**.

Mass normalized fraction

We constructed the extended model of *Synechocystis*, based on our base model to incorporate major components of photosynthesis in our model and we chose simple decimal numbers in matrix M, reflecting the overall proportions of each column in S.

Molecular mass of metabolites

For the molecular masses of metabolites, we used 44 [g mol⁻¹] for “CO₂”, 1 [g mol⁻¹] for “H⁺”, 52.5 [g mol⁻¹] for “a_C”(corresponding to mean molecular mass of HCO₃ and CO₂), 10500 [g mol⁻¹] for “PC” and “PC⁻” (corresponding to Plastocyanin molecular weight), 749.2 [g mol⁻¹] for “PQ” (corresponding to plastoquinone molecular weight), 751.2 [g mol⁻¹] for “PQH₂” (corresponding to reduced plastoquinone molecular weight), 1 [g mol⁻¹] for “ATP” and “NADPH” (corresponding to the difference of molecular weight between their oxidized and reduced states to compensate the absence of each), 186 [g mol⁻¹] for “C3” (corresponding to molecular weight of 3-phosphoglycerate, similar to (Rugen et al., 2015)), and 100 [g mol⁻¹] for “aa” (estimated for the molecular mass of amino acids).

Kinetic parameters

We performed the same procedure for conversion of our kinetic parameters into mass units as was implemented on the base model; i) All K_m were converted from [mol L⁻¹] to [g L⁻¹] by multiplying the original value with the corresponding molecular mass (see previous section) and also by number of all reactions lumped in the corresponding enzyme (mentioned in each protein class, also in **Supplementary Table 2**), ii) All k_{cat} were converted to mass units by multiplying

the original value with the product molecular mass and the total number of catalyzing reactions, and then dividing by the mass of the enzyme (see **Supplementary Table 2**).

Carbon transporter

The turnover number of carbon transporter (k_{cat}) and Michaelis constant of carbon transporter is taken from the base model and multiplied by the number of 5 proteins in this category, which resulted in 51.54 [h⁻¹] and 0.00315 [g L⁻¹], respectively. The K_m value for ATP requirement to import carbon is set to 1.5×10^{-5} [M] (Omata et al., 2002), which by converting to mass units resulted in 0.0003 [g L⁻¹].

Photosystem II and photosystem I

The K_m values for light as the external concentration were assigned the same value as the base model for both PSI (14 reactions) and PSII (28 reactions). The K_m value for the enzymatic reaction catalyzed by PSI (1.18.1.2) and (1.10.3.9) for PSII, which are the only accessible reference in this context, and we found a reported value of 0.011 mM (1.617 [g L⁻¹] for “PQ”) from cyanobacteria of *Anabaena* and 0.06 mM (1.25 [g L⁻¹] for “PC”) for *anthus annuus*, respectively. The maximum turnover numbers for PSII and PSI enzymes were derived from a previously published model, with the values of 250 [s⁻¹] (Matuszyńska et al., 2019). For the light-harvesting components, we normalized their k_{cat} by multiplying it by the product mass of the protons, and then multiplied the result by the sum of the enzyme masses within their respective classes, which corresponded to 3.69 [h⁻¹] and 37.96 [h⁻¹], respectively. Additionally, the only parameter that was fitted was the inhibition constant I_L^r , which was estimated to have a value of 85 [μE m⁻² s⁻¹]. The fitting of the inhibition constant was carried out such that the model growth rate matched the experimental growth rate at a light intensity of 1100 [μE m⁻² s⁻¹] (100% light intensity).

Cytochrome b6

In the case of Cytb6f, the EC number 7.1.1.6, which corresponds to this enzyme, lacked any recorded measurements. However, for its partial EC number 7.1.1.2, there was an available value of 1.8 [s⁻¹], which was converted to mass units (15 reactions) resulting in 1404.89 [h⁻¹], with the estimation of its backward k_{cat} , equivalent to half of the forward k_{cat} , which amounts to 702 [h⁻¹]. The K_m value for both plastocyanin (PC) was determined as 9×10^{-6} [M] (EC number 7.1.1.6 for *Spinacia oleracea*), which were converted to mass units (15 reactions) resulting in 1.41 [g L⁻¹], and for reduced plastoquinone (PQH₂) the K_m value was estimated as 1 [g L⁻¹].

ATP synthase

The unique EC number for ATPsynthase unit was 7.1.2.2. The K_m value for proton (H⁺) in ATPsynthase was estimated as 1 [g L⁻¹] and for turnover number k_{cat} value, partial EC number of 7.1.1.2 upon availability was used 1.8 [s⁻¹], which were converted to mass units (one reaction of ATPsynthase) resulted in 4.9 [h⁻¹].

Metabolism & Carbon fixation

For the turnover number (k_{cat}), we utilized the EC. Number of 4.1.1.39 RuBisCo reaction with the value of $2.7 \text{ [s}^{-1}\text{]}$ as reported in (Horken & Tabita, 1999), which were converted to mass units (239 reactions) resulted in $3.45 \text{ [h}^{-1}\text{]}$. The Michaelis constant of metabolism unit was also determined the maximum reported value in EC number 4.1.1.39 for *Synechocystis*, 0.2680 [mM] , which were converted to mass units (239 reactions) resulted in $0.64 \text{ [g L}^{-1}\text{]}$, $0.65 \text{ [g L}^{-1}\text{]}$, $11.91 \text{ [g L}^{-1}\text{]}$ for ATP, NADPH and C₃, respectively. We incorporated these value into both the Metabolism and Carbon Fixation reactions of our extended model, recognizing their central role within the overarching framework of carbon metabolism.

Ribosome

The turnover number (k_{cat}) and Michaelis constant K_m of ribosome unit was identical to the base model. We also estimated the K_m value for ATP as $1 \text{ [g L}^{-1}\text{]}$.

Cell densities

Cell density (Rho) was determined through the analysis of experimental data provided in (Faizi et al., 2018). The average cell dry weight across various light intensities was calculated, yielding a value of $207.33 \text{ mg culture per liter}$. Additionally, the cell count was established at 2.4×10^{10} , and with a conversion factor accounting for the volume of each cell set at 2.24×10^{-14} , the resulting cell density was determined to be $385.67 \text{ [g L}^{-1}\text{]}$, and it is assumed to be constant for all the conditions for both models.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) [CRC 1310 and, under Germany's Excellence Strategy, EXC 2048/1–390686111] and by the European Union [ERC AdG “MechSys”–Project ID 101055141].

Conflict of interest

The authors declare no conflict of interest.

3.5 References

- Babele, P. K., Kumar, J., & Chaturvedi, V. (2019). Proteomic De-regulation in cyanobacteria in response to abiotic stresses. *Frontiers in Microbiology*, 10(JUN), 1–22. <https://doi.org/10.3389/fmicb.2019.01315>
- Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Bye-AJee, H., Cowley, A., Da Silva, A., De Giorgi, M., Dogan, T., Fazzini, F., Castro, L. G., Figueira, L., ... Zhang, J. (2017). UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, 45(D1), D158–D169. <https://doi.org/10.1093/nar/gkw1099>
- Blanc-Garin, V., Chenebault, C., Diaz-Santos, E., Vincent, M., Sassi, J. F., Cassier-Chauvat, C., & Chauvat, F. (2022). Exploring the potential of the model cyanobacterium *Synechocystis* PCC 6803 for the photosynthetic production of various high-value terpenes. *Biotechnology for Biofuels and Bioproducts*, 15(1), 1–11. <https://doi.org/10.1186/s13068-022-02211-0>
- Bremer, H., & Dennis, P. P. (2008). Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. *EcoSal Plus*, 3(1). <https://doi.org/10.1128/ecosal.5.2.3>
- Burnap, R. L. (2015). Systems and photosystems: Cellular limits of autotrophic productivity in cyanobacteria. *Frontiers in Bioengineering and Biotechnology*, 3(JAN), 1–13. <https://doi.org/10.3389/fbioe.2015.00001>
- Dourado, H., & Lercher, M. J. (2020). An analytical theory of balanced cellular growth. *Nature Communications*, 11(1). <https://doi.org/10.1038/s41467-020-14751-w>
- Dourado, H., Liebermeister, W., Ebenhöf, O., & Lercher, M. J. (2023). Mathematical properties of optimal fluxes in cellular reaction networks at balanced growth. *PLoS Computational Biology*, 19(6), 1–26. <https://doi.org/10.1371/journal.pcbi.1011156>
- Faizi, M., Zavřel, T., Loureiro, C., Červený, J., & Steuer, R. (2018). A model of optimal protein allocation during phototrophic growth. *BioSystems*, 166, 26–36. <https://doi.org/10.1016/j.biosystems.2018.02.004>
- Goelzer, A., Fromion, V., & Scorletti, G. (2011). Cell design in bacteria as a convex optimization problem. *Automatica*, 47(6), 1210–1218. <https://doi.org/10.1016/j.automatica.2011.02.038>
- Horken, K. M., & Tabita, F. R. (1999). Closely related form I ribulose biphosphate carboxylase/oxygenase molecules that possess different CO₂/O₂ substrate specificities. *Archives of Biochemistry and Biophysics*, 361(2), 183–194. <https://doi.org/10.1006/abbi.1998.0979>
- Jahn, M., Vialas, V., Karlsen, J., Maddalo, G., Edfors, F., Forsström, B., Uhlén, M., Käll, L., & Hudson, E. P. (2018). Growth of Cyanobacteria Is Constrained by the Abundance of Light and Carbon Assimilation Proteins. *Cell Reports*, 25(2), 478–486.e8. <https://doi.org/10.1016/j.celrep.2018.09.040>
- Knoop, H., Gründel, M., Zilliges, Y., Lehmann, R., Hoffmann, S., Lockau, W., & Steuer, R. (2013). Flux Balance Analysis of Cyanobacterial Metabolism: The Metabolic Network of *Synechocystis* sp. PCC 6803. *PLoS Computational Biology*, 9(6). <https://doi.org/10.1371/journal.pcbi.1003081>

- Liebermeister, W., & Klipp, E. (2006). Bringing metabolic networks to life: Convenience rate law and thermodynamic constraints. *Theoretical Biology and Medical Modelling*, 3. <https://doi.org/10.1186/1742-4682-3-41>
- Marcus, Y., Altman-Gueta, H., Finkler, A., & Gurevitz, M. (2005). Mutagenesis at two distinct phosphate-binding sites unravels their differential roles in regulation of Rubisco activation and catalysis. *Journal of Bacteriology*, 187(12), 4222–4228. <https://doi.org/10.1128/JB.187.12.4222-4228.2005>
- Matthias, K., Klähn, S., Ingeborg, S., Jasper, K. F. M., Wolfgang, R. H., & Björn, V. (2014). Comparative analysis of the primary transcriptome of *synechocystis* sp. PCC 6803. *DNA Research*, 21(5), 527–539. <https://doi.org/10.1093/dnares/dsu018>
- Matuszyńska, A., Saadat, N. P., & Ebenhöf, O. (2019). Balancing energy supply during photosynthesis – a theoretical perspective. *Physiologia Plantarum*, 166(1), 392–402. <https://doi.org/10.1111/ppl.12962>
- Milo, R., & Phillips, R. (2015). Cell Biology by the Numbers. *Cell Biology by the Numbers*. <https://doi.org/10.1201/9780429258770/CELL-BIOLOGY-NUMBERS-ROB-MILO-ROB-PHILLIPS>
- Molenaar, D., Van Berlo, R., De Ridder, D., & Teusink, B. (2009). Shifts in growth strategies reflect tradeoffs in cellular economics. *Molecular Systems Biology*, 5(323), 1–10. <https://doi.org/10.1038/msb.2009.82>
- Mori, M., Schink, S., Erickson, D. W., Gerland, U., & Hwa, T. (2017). Quantifying the benefit of a proteome reserve in fluctuating environments. *Nature Communications*, 8(1), 1225. <https://doi.org/10.1038/s41467-017-01242-8>
- O’Brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R., & Palsson, B. (2013). Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular Systems Biology*, 9(693). <https://doi.org/10.1038/msb.2013.52>
- Omata, T., Takahashi, Y., Yamaguchi, O., & Nishimura, T. (2002). Structure, function and regulation of the cyanobacterial high-affinity bicarbonate transporter, BCT1. *Functional Plant Biology*, 29(3), 151–159.
- Rugen, M., Bockmayr, A., & Steuer, R. (2015). Elucidating temporal resource allocation and diurnal dynamics in phototrophic metabolism using conditional FBA. *Scientific Reports*, 5(February), 1–16. <https://doi.org/10.1038/srep15247>
- Sánchez, B. J., Zhang, C., Nilsson, A., Lahtvee, P., Kerkhoven, E. J., & Nielsen, J. (2017). Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Molecular Systems Biology*, 13(8), 935. <https://doi.org/10.15252/msb.20167411>
- Santos-Merino, M., Yun, L., & Ducat, D. C. (2023). Cyanobacteria as cell factories for the photosynthetic production of sucrose. *Frontiers in Microbiology*, 14(February). <https://doi.org/10.3389/fmicb.2023.1126032>

- Schmidt, A., Kochanowski, K., Vedelaar, S., Ahrné, E., Volkmer, B., Callipo, L., Knoop, K., Bauer, M., Aebbersold, R., & Heinemann, M. (2016). The quantitative and condition-dependent *Escherichia coli* proteome. *Nature Biotechnology*, 34(1), 104–110. <https://doi.org/10.1038/nbt.3418>
- Schomburg, I., Chang, A., Placzek, S., Söhlgen, C., Rother, M., Lang, M., Munnaretto, C., Ulas, S., Stelzer, M., Grote, A., Scheer, M., & Schomburg, D. (2013). BRENDA in 2013: Integrated reactions, kinetic data, enzyme function data, improved disease classification: New options and contents in BRENDA. *Nucleic Acids Research*, 41(D1), 764–772. <https://doi.org/10.1093/nar/gks1049>
- Wortel, M. T., Noor, E., Ferris, M., Bruggeman, F. J., & Liebermeister, W. (2018). Metabolic enzyme cost explains variable trade-offs between microbial growth rate and yield. *PLoS Computational Biology*, 14(2), 1–21. <https://doi.org/10.1371/journal.pcbi.1006010>
- Yu, Y., You, L., Liu, D., Hollinshead, W., Tang, Y. J., & Zhang, F. (2013). Development of *synechocystis* sp. PCC 6803 as a phototrophic cell factory. *Marine Drugs*, 11(8), 2894–2916. <https://doi.org/10.3390/md11082894>
- Zavřel, T., Faizi, M., Loureiro, C., Poschmann, G., Stühler, K., Sinetova, M., Zorina, A., Steuer, R., & Červený, J. (2019). Quantitative insights into the cyanobacterial cell economy. *eLife*, 8, 1–29. <https://doi.org/10.7554/eLife.42508>

3.6 Supplementary text for Cyanobacterial resource allocation strategies explained by Growth Balance Analysis

Sajjad Ghaffarinasab¹, Oliver Ebenhöf², Martin J. Lercher¹, Hugo Dourado^{1*}

Table S1. Parameters of the base model

Parameters	definition	Value (unit)	Reference
K_m^e	Michaelis constant of e	0.0007875 [g L ⁻¹]	(Faizi et al., 2018)
$K_m^{a_C}$	Michaelis constant of a_C	0.0007875 [g L ⁻¹]	(Faizi et al., 2018)
K_m^{AA}	Michaelis constant of AA	0.000097812 [g L ⁻¹]	(Faizi et al., 2018)
K_m^c	Michaelis constant of c	0.00648 [g L ⁻¹]	(Faizi et al., 2018)
I_L^r	Inhibition constant	47 [μE m ⁻² s ⁻¹]	Fitted
k_{cat}^{PSU}	turnover number of Photosystem	0.0847 [h ⁻¹]	(Milo & Phillips, 2015)
k_{cat}^S	turnover number of Carbon transporter	8.434 [h ⁻¹]	(Faizi et al., 2018)
k_{cat}^M	turnover number of Metabolism	0.794 [h ⁻¹]	(Faizi et al., 2018)
k_{cat}^R	turnover number of Ribosome	4.55 [h ⁻¹]	(Faizi et al., 2018)
E_{mass}^{PSU}	Mass of the enzyme class of Photosystem	8,500,380 [Da]	Sourced from UniProt
E_{mass}^S	Mass of the enzyme class of Carbon Transporter	185,921 [Da]	Sourced from UniProt
E_{mass}^M	Mass of the enzyme class of Metabolism	11,968,859 [Da]	Sourced from UniProt
E_{mass}^R	Mass of the enzyme class of Ribosome	2,306,967 [Da]	Sourced from UniProt
ρ	Cell density	385.67 [g L ⁻¹]	(Faizi et al., 2018)

Table S2. Parameters of the extended model

Parameters	definition	Value (unit)	Reference
$K_m^{a_C}$	Michaelis constant of a_C	0.00315 [g L ⁻¹]	Sourced from the base model
K_m^{ATP}	Michaelis constant of ATP in Carbon_T	0.0003 [g L ⁻¹]	(Omata et al., 2002)
K_m^{PQ}	Michaelis constant of PQ in PSII	1.25 [g L ⁻¹]	Sourced from BRENDA
$K_m^{PC^-}$	Michaelis constant of PC^- in PSI	1.617 [g L ⁻¹]	Sourced from BRENDA
K_m^{PC}	Michaelis constant of PC in Cytb6f	1.41 [g L ⁻¹]	Sourced from BRENDA
$K_m^{PQH_2}$	Michaelis constant of PQH_2 in Cytb6f	1 [g L ⁻¹]	Estimated
$K_m^{H^+}$	Michaelis constant of H^+ in ATPsyn	1 [g L ⁻¹]	Estimated
$K_m^{CO_2}$	Michaelis constant of CO_2 in C_fix	11.91 [g L ⁻¹]	Sourced from BRENDA
K_m^{ATP}	Michaelis constant of ATP in C_fix & Met	0.64 [g L ⁻¹]	Sourced from BRENDA
$K_m^{C_3}$	Michaelis constant of C_3 in Met	11.91 [g L ⁻¹]	Sourced from BRENDA
K_m^{NADPH}	Michaelis constant of $NADPH$ in C_fix & Met	0.65 [g L ⁻¹]	Sourced from BRENDA
K_m^{ATP}	Michaelis constant of ATP in Ribosome	1 [g L ⁻¹]	Estimated
K_m^{AA}	Michaelis constant of AA in Ribosome	8.3 [g L ⁻¹]	Sourced from the base model
I_L^r	Inhibition constant	85 [μE m ⁻² s ⁻¹]	Fitted
$k_{cat}^{Carbon_T}$	turnover number of Carbon_T	51.54 [h ⁻¹]	Sourced from the base model
k_{cat}^{PSII}	turnover number of PSII	3.69 [h ⁻¹]	(Matuszyńska et al., 2019)

Parameters	definition	Value (unit)	Reference
k_{cat}^{PSI}	turnover number of PSI	37.96 [h ⁻¹]	(Matuszyńska et al., 2019)
k_{cat}^{Cytb6f}	turnover number of Cytb6f	1404.891 [h ⁻¹]	Sourced from BRENDA
k_{cat}^{ATPsyn}	turnover number of ATPsyn	4.9 [h ⁻¹]	Sourced from BRENDA
$k_{cat}^{C_fix}$	turnover number of C_fix	3.45 [h ⁻¹]	(Horken & Tabita, 1999)
k_{cat}^{Met}	turnover number of Met	3.45 [h ⁻¹]	(Horken & Tabita, 1999)
$k_{cat}^{Ribosome}$	turnover number of Ribosome	4.55 [h ⁻¹]	Sourced from the base model
$E_{mass}^{Carbon_T}$	Mass of the enzyme class of Carbon_T	185,921 [Da]	Sourced from UniProt
E_{mass}^{PSII}	Mass of the enzyme class of PSII	6,825,462 [Da]	Sourced from UniProt
E_{mass}^{PSI}	Mass of the enzyme class of PSI	331,876 [Da]	Sourced from UniProt
E_{mass}^{Cytb6f}	Mass of the enzyme class of Cytb6f	778,366 [Da]	Sourced from UniProt
E_{mass}^{ATPsyn}	Mass of the enzyme class of ATPsyn	564,676 [Da]	Sourced from UniProt
E_{mass}^{Met}	Mass of the enzyme class of Met	67,261,272 [Da]	Sourced from UniProt
$E_{mass}^{Ribosome}$	Mass of the enzyme class of Ribosome	2,306,967 [Da]	Sourced from UniProt
ρ	Cell density	385.67 [g L ⁻¹]	(Faizi et al., 2018)

Chapter 4

Cell Growth Simulator: a user-friendly web server for growth balance analysis

S.G. and H.D. conceived the GBA Solver platform and wrote the manuscript. S.G. developed and implemented the functionalities and user interface. H.D. supervised the project, provided conceptual guidance, contributed to the user experience improvements.

Manuscript status: Under review at the journal of Bioinformatics

Cell Growth Simulator: a user-friendly web server for growth balance analysis

Sajjad Ghaffarinasab^{1¶}, Hugo Dourado^{1¶*}

¹Institute for Computer Science and Department of Biology, Heinrich Heine University, Universitätsstraße 1, D-40225 Düsseldorf, Germany

*Corresponding author

Email: hugo.dourado@hhu.de

¶ The authors contributed equally to this work

Abstract

Computational models are essential for understanding the complex biochemical processes that govern cellular growth and metabolism. Growth Balance Analysis (GBA) provides a powerful framework for modeling cellular self-replication by incorporating metabolite concentrations and their non-linear influence on reaction kinetics. However, the complexity of implementing and solving such models has made them inaccessible to researchers without programming expertise. Here we present *Cell Growth Simulator*, a web-based application that enables efficient construction and analysis of GBA models through an intuitive interface and interactive visualizations. *Cell Growth Simulator* uses an intuitive spreadsheet interface, eliminating the need for coding, and integrates data from the BRENDA enzyme database to facilitate the incorporation of kinetic parameters. *Cell Growth Simulator* provides interactive visualizations, including customizable plots and dynamic metabolic pathway maps for interpreting optimization results. The platform makes nonlinear modeling of resource allocation in coarse-grained cellular systems accessible to a broader scientific audience, fostering interdisciplinary collaboration and advancing our understanding of cellular metabolism and growth. *Cell Growth Simulator* is freely available at: <https://cellgrowthsim.com/>.

Keywords:

Growth balance analysis, Optimal Resource allocation, Web application, Nonlinear Cellular Model, Systems Biology

Author summary

We created *Cell Growth Simulator* to make modeling cellular self-replication and growth more accessible. Our goal was to create a web-based tool that helps researchers, regardless of their programming background, investigate how cells allocate their resources under realistic, nonlinear conditions. By focusing on a simplified framework that captures essential features of metabolism, we enable scientists to explore core principles of cellular growth, such as how proteins and metabolites interact to drive balanced growth.

Cell Growth Simulator provides a simple spreadsheet-like interface where users can build or import their models, eliminating the need for coding. Our application connects to a widely used enzyme database to help researchers select realistic parameter values, and provides interactive graphs and pathway maps to visualize and interpret results. By streamlining the complex mathematical steps involved in modeling and analysis, *Cell Growth Simulator* lowers technical barriers and encourages collaboration between biologists, computational scientists, and anyone interested in understanding how cells work. We hope it will serve as a valuable resource for studying metabolism and advancing our knowledge of cell growth.

4.1 Introduction

Metabolic modeling has emerged as a crucial approach for understanding the complex biochemical processes that govern cellular growth and metabolism. Computational models enable researchers to predict cellular behaviors, optimize metabolic pathways, and explore resource allocation strategies from simple principles. Linear computational approaches have been widely used to study resource allocation during balanced growth, a steady-state condition where concentrations of cellular components are constant in time. These methods typically approximate biochemical reaction kinetics by assuming linear relationships between fluxes and catalyst concentrations, neglecting the nonlinear effects of metabolite concentrations on reaction kinetics (Goelzer & Fromion, 2011; O’Brien et al., 2013; Sánchez et al., 2017). More sophisticated nonlinear models can explicitly account for these metabolite-dependent effects through kinetic rate laws (Molenaar et al., 2009). Recently, Growth Balance Analysis (GBA) was introduced as a general framework to model and analyze such nonlinear systems, investigating how cells grow and allocate their resources under fundamental physical constraints, including mass conservation, nonlinear reaction kinetics, and a limit on cellular density, accounting for the dilution of all cellular components by growth (Dourado et al., 2023; Dourado & Lercher, 2020).

While the GBA framework simplifies and facilitates the mathematical modeling and analysis of cellular balanced growth under nonlinear constraints, its computational implementation requires mathematical and programming skills. To increase the accessibility of GBA to a wider audience, we developed *Cell Growth Simulator* a user-friendly web-based platform built using R/Shiny. R is a widely used programming language and software environment for statistical computing and graphics, while Shiny is a web application framework for R that facilitates the creation of interactive and user-friendly web-based interfaces.

A rich ecosystem of web-based platforms already supports metabolic modeling. For linear, constraint-based workflows, such as flux balance analysis (FBA) (Orth et al., 2010) and its extensions, tools like CNApy (Thiele et al., 2022), Escher (King et al., 2015), ModelExplore (Martyushenko & Almaas, 2019), CAVE (Mao et al., 2023), and Fluxer (Hari & Lobo, 2020), provide convenient model construction and visualization. For nonlinear kinetics, mature software applications including COPASI (Hoops et al., 2006), Tellurium (Choi et al., 2018), Virtual Cell (Schaff et al., 1997), and AMICI (Fröhlich et al., 2021) support arbitrary rate laws and time-course analyses, and web-based environments such as JWS Online (Olivier & Snoep, 2004) and runBioSimulations (Shaikh et al., 2021) have long enabled web execution of nonlinear cellular models. To contribute to the community and extend this ecosystem, we developed the *Cell Growth Simulator*, a web-based implementation of the GBA formalism for nonlinear self-replicator models, where resource allocation strategies emerge from the optimization of the

cellular growth rate, incorporating metabolite-dependent (nonlinear) kinetics and the dilution of all cellular components by growth. By focusing on coarse-grained systems, *Cell Growth Simulator* enables efficient exploration of cellular resource allocation and proteome efficiency (Doan et al., 2022; Ghaffarinasab et al., 2023; Hui et al., 2015; Molenaar et al., 2009; Scott et al., 2010a; Weiße et al., 2015).

Coarse-grained nonlinear models have proven to be especially useful for revealing fundamental principles of cellular physiology. By simplifying complex networks into a few effective reactions and catalytic sectors, these models can capture important trade-offs, such as growth laws (ribosome allocation) (Erickson et al., 2017; Scott et al., 2010b), shifts in metabolic strategies (Molenaar et al., 2009), and proteome partitioning, without requiring exhaustive mechanistic detail (Doan et al., 2022; Ghaffarinasab et al., 2023; Hui et al., 2015; Weiße et al., 2015). This simplification renders nonlinear optimization problems tractable while providing insight into global growth laws and resource allocation strategies (Erickson et al., 2017; Molenaar et al., 2009; Scott et al., 2010b; Weiße et al., 2015). In this spirit, the *Cell Growth Simulator* is designed for coarse-grained self-replicator models, enabling the rapid, interpretable exploration of how metabolite concentrations and growth-driven dilution shape cellular behavior.

The GBA framework estimates the cell state at balanced growth by solving a nonlinear optimization problem that maximizes the growth rate (μ), subject to mass conservation, nonlinear kinetic rate laws of biochemical reactions, constant cell density, and non-negative concentrations. This optimization is greatly simplified by its formulation on the flux fraction vector defined as $f := v/\mu\rho$, where v denotes the vector of mass fluxes of reactions in units [$\text{g L}^{-1} \text{h}^{-1}$] and ρ is the cell mass density in [g L^{-1}] (Dourado et al., 2023). Fig 1 presents a general view of the framework. A detailed mathematical description of the GBA framework can be found in the original publications (Dourado et al., 2023; Dourado & Lercher, 2020).

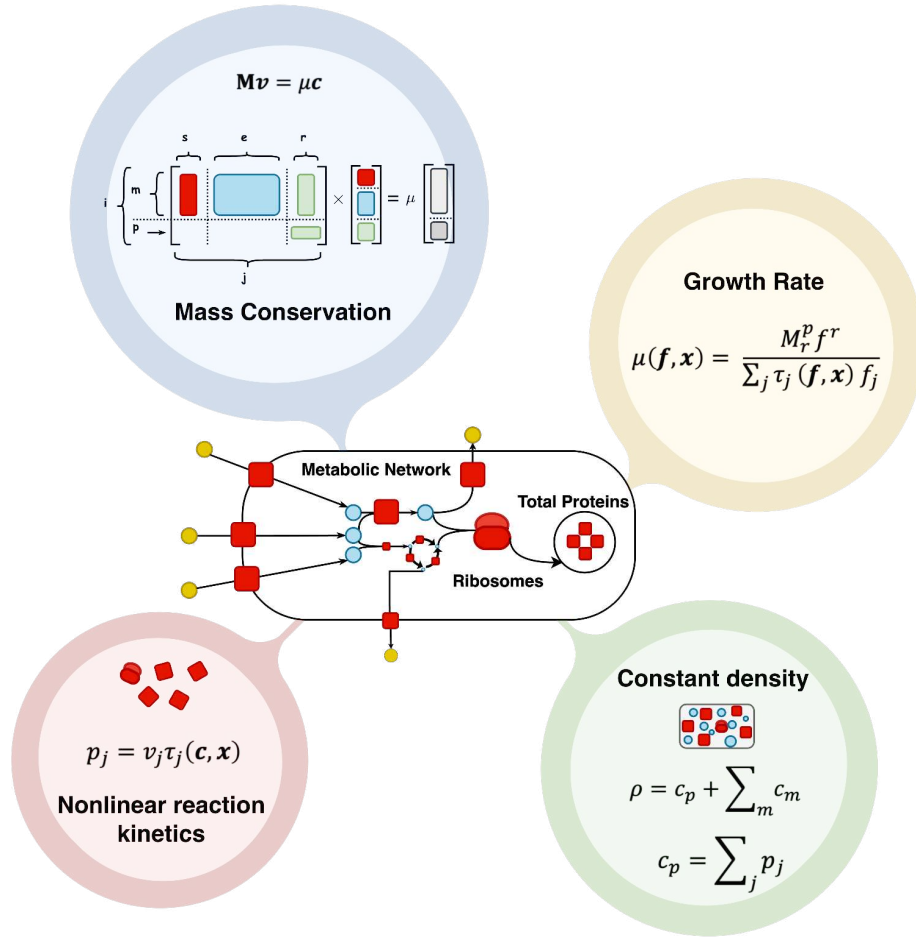


Fig 1. Schematic representation of the Growth Balance Analysis (GBA) framework as implemented in the *Cell Growth Simulator*. The diagram illustrates the key components and constraints governing the optimal balanced growth of a self-replicating cell model defined by the triple (M, τ, ρ) , where M is the mass fraction matrix encoding the reaction network structure, τ are kinetic rate laws, and ρ is the cell density. The mass conservation constraint (blue) relates mass fluxes of reactions v with the growth rate μ and concentrations c of reactants in the model, including concentrations of metabolites c_m and the total protein c_p (equal to the sum of individual protein concentrations p). The reaction kinetics (red) express the relationship between reaction fluxes v , protein concentrations p , and reaction turnover times τ that depend on internal concentrations c and external concentrations x . The density constraint (green) defines the total cellular density ρ as the sum of metabolite and protein concentrations. All these constraints can be encoded into a single equation (yellow) for the constrained growth rate μ in terms of the flux fractions (Dourado et al., 2023).

4.2 Design and implementation

User interface and front-end development

Cell Growth Simulator is built using R (version 4.4.1) and the Shiny framework, creating an interactive web application that combines robust backend functionality with an intuitive user interface. The frontend architecture integrates HTML for content structure, CSS for visual styling, and JavaScript for dynamic interactions and enhanced client-side performance. Building upon the responsive Mazer dashboard template (<https://github.com/zuramai/mazer>), we extensively customized the interface through additional CSS and JavaScript implementations to create a specialized environment optimized for cellular growth modeling and analysis. This layered architecture ensures both technical functionality and user accessibility while maintaining high performance across different devices and screen sizes.

Data sources and management

Cell Growth Simulator integrates comprehensive enzyme kinetic data from the BRENDA database (Chang et al., 2021), providing users access to curated turnover numbers (k_{cat}) and Michaelis constants (K_m). The data processing pipeline filters the JSON-formatted database to include only wild-type enzyme parameters, which are then systematically organized by Enzyme Commission (EC) numbers. Through an interactive table powered by the "reactable" package (Lin, 2019), users can efficiently search and filter enzyme parameters based on multiple criteria including organism, EC number classification, and substrate specificity. This streamlined interface simplifies the often challenging task of identifying appropriate kinetic parameters for metabolic models.

Input data structure and model configuration

Kinetic models such as GBA describe reaction dynamics through reaction fluxes v_j that depend on protein concentrations p_j catalyzing the reaction (transport, enzymatic, or “ribosome” reaction producing proteins) and turnover times τ_j according to general kinetic rate laws:

$$v_j = \frac{p_j}{\tau_j(c, x)} \quad (1)$$

The turnover times $\tau_j(c, x)$ depend on internal (c) and external (x) metabolite concentrations through several kinetic parameters: turnover numbers (k_{cat}) for both forward and backward directions, Michaelis constants (K_m), and where applicable, activation (KA) and inhibition constants (KI).

In *Cell Growth Simulator*, we implement convenience kinetics (Liebermeister & Klipp, 2006), a general framework that can describe reversible reactions and regulatory effects from activators

and inhibitors. For an irreversible reaction j (where the backward turnover number $k_{catb}^j = 0$), the turnover time τ_j is given by:

$$\tau_j(c, x) = \frac{1}{k_{catf}^j} \Pi_m \left(\frac{c_m}{K_m^j + c_m} \right)^{-1} \left(\frac{c_m}{KA_m^j + c_m} \right)^{-1} \left(\frac{KI_m^j}{KI_m^j + c_m} \right)^{-1} \Pi_n \left(\frac{x_n}{K_n^j + x_n} \right)^{-1} . \quad (2)$$

The parameters in this equation represent:

1. Forward turnover number (k_{catf}^j): the maximum rate at which protein j (transporter, enzyme, or ribosome) can convert substrates to products.
2. Michaelis constants (K_m^j): the substrate concentration where the reaction rate reaches half its maximum value. Lower values indicate stronger protein-substrate binding. For metabolites not participating in reaction j , we set $K_m^j = 0$.
3. Regulatory constants:
 - a. Activation constants (KA_m^j): quantify activator effects. Set to 0 for non-activating metabolites.
 - b. Inhibition constants (KI_m^j): quantify inhibitor effects. Set to ∞ for non-inhibiting metabolites.

This formulation captures how reaction fluxes are modulated by both enzyme concentrations and metabolite-dependent effects, with the turnover time τ_j representing the time required to process one unit of substrate under the given conditions.

Cell Growth Simulator accepts GBA models in open-source spreadsheet format (ODS), consisting of the following matrices and vectors:

- **Mass fraction matrix M :** quantifies the mass fraction of each reactant going through each reaction, with negative entries representing reactant consumption and positive entries indicating product formation. Rows correspond to reactants, columns to reactions. Due to mass conservation, the sum of positive entries in each column equals 1, and the sum of negative entries equals -1. External reactants are denoted by the prefix "x_". By default, the last row represents the total protein concentration in mass units, while the last column corresponds to the ribosomal reaction that produces all cellular proteins.
- **Michaelis constant matrix K :** Contains K_m^j values (in $[g \cdot L^{-1}]$) representing the Michaelis constant for metabolite m in reaction j , organized in the same order as in M . When unspecified, a low default value of $0.1 \text{ g} \cdot \text{L}^{-1}$ is applied.
- **Matrices of regulatory constants:** inhibition constants (KI) and activation constants (KA) for each metabolite m in each reaction j , expressed in $[g \cdot L^{-1}]$ and arranged as in matrix M . Zero values for individual entries indicate that no inhibition or activation effects are considered.

- **Turnover number matrix (k_{cat}):** Comprises two rows: the first (k_{catf}) contains forward-direction turnover numbers, while the second (k_{catb}) contains backward-direction values. Column labels match those in matrix M . Values represent molecules of product produced per catalyst protein per unit time (in $[\text{h}^{-1}]$).
- **Condition matrix:** Defines simulation parameters for each growth condition. The first row ("rho") specifies cell density in $[\text{g}\cdot\text{L}^{-1}]$. Subsequent rows contain external reactant concentrations ($\text{g}\cdot\text{L}^{-1}$), with the number of rows corresponding to the number of external reactants (with "x_" prefix) in M . Each column represents a distinct growth condition for the simulations.

Cell Growth Simulator provides a downloadable GBA model template as a reference for users developing their own models. Alternatively, users can build models directly within the "Create Model" section of *Cell Growth Simulator*. The application employs the "shinyMatrix" R package to generate and display matrices across model tabs, with custom JavaScript modifications that enhance usability through features like arrow key navigation and full-cell editing capabilities (Andreas Neudecker, 2019).

The model creation process begins by specifying the number of reactants and reactions, followed by entering their corresponding labels. The system automatically designates "Ribosome" for the final column and "Protein" for the final row. Users must input kinetic parameters (K_{m} and k_{cat}) on dedicated tabs. To facilitate this process, these tabs display collapsible cards containing relevant kinetic parameters sourced from the BRENDA database (Chang et al., 2021), with an additional option to query and filter parameters based on organism, EC numbers, and enzyme substrates.

In the "Condition" tab, users input the number of external reactants and "rho" for cell density, and specify the total number of growth conditions for optimization. The application then dynamically adjusts the "condition" matrix, enabling users to configure different external concentrations for each optimization scenario.

After uploading or creating a GBA model, the application presents a comprehensive preview of the input data. This allows users to review their model thoroughly and make direct modifications within the interface. Once finalized, users can export their refined model in ODS format for future use, supporting an iterative development approach that progressively enhances model performance and accuracy.

Model validation and numerical optimization process in *Cell Growth Simulator*

After importing their GBA model, users can validate its compliance with framework requirements through the "Check Model" function. This comprehensive validation process ensures proper parameter import by performing several critical checks:

1. **Data integrity:** verifies that no missing (NA) or non-numeric values exist throughout the model;
2. **Dimensional consistency:** confirms that all matrices maintain consistent dimensions, with matching numbers of reactions and reactants;
3. **Non-negative value verification:** validates that all external concentrations are non-negative and that cell density is positive;
4. **Michaelis constant validation:** identifies instances where substrates (negative entries in the matrix M) have Michaelis constants incorrectly set to zero ($K_m=0$).

If $K_m=0$ errors are detected, the system issues a warning and automatically applies the default low value of $0.1 \text{ [g}\cdot\text{L}^{-1}]$ to these parameters. Should any other validation criteria fail, the application displays a specific error message directing users to reset their session and revise the problematic aspects of their model.

Upon successful validation, the system notifies users that their model meets all requirements, enabling them to proceed confidently to subsequent analysis steps. This rigorous validation workflow ensures that only properly formatted models enter the computational pipeline, reducing the likelihood of simulation errors or biologically implausible results.

Users have the option to bypass the model validation step and directly initiate growth balance analysis by selecting the "Run" button. The optimization process utilizes the "nloptr" R package, which implements non-linear optimization through the AUGLAG (augmented Lagrangian) method (Ypma & Johnson, 2011). This framework allows users to choose from several specialized local solvers:

- **SLSQP** (Sequential Least Squares Quadratic Programming): the default solver, well-suited for smaller models requiring precise constraint management.
- **LBFGS** (Low-storage BFGS): Selected for its efficient memory usage and robust performance;
- **MMA** (Method of Moving Asymptotes): particularly effective for highly non-linear models with complex constraints;

While SLSQP serves as the default, users can select alternative solvers through the 'Advanced option' panel when facing performance challenges. Due to the nature of numerical optimization, it is not always possible to know beforehand which solver will be most appropriate for a given model. If the default solver exhibits slow convergence or fails to reach a solution that satisfies the optimization criteria, users are encouraged to evaluate alternative solvers to enhance performance outcomes.

The selection of solver can significantly impact both convergence rate and solution accuracy, with the optimal choice depending on model complexity and problem characteristics. During

optimization, a dynamic progress bar tracks completion for each growth condition, providing users with real-time estimates of the total processing time. For reference, the included case study model (comprising 4 reactions and 4 internal reactants) typically completes optimization in under 1 second per condition on standard personal computing hardware.

Interactive metabolic pathway and data visualization

Cell Growth Simulator uses the "apexcharter" package (Perrier & Meyer, 2019) to generate comprehensive, interactive visualizations of numerical solutions across all simulation conditions. These visualizations include growth rates, protein concentrations, protein fractions, reaction fluxes, and metabolite concentrations. Each interactive plot supports zooming and panning to enable researchers to investigate specific trends. Users can export these visualizations in both SVG and JPEG formats for publication or presentation purposes.

Beyond conventional plotting capabilities, *Cell Growth Simulator* integrates d3flux (<https://github.com/pstjohn/d3flux>) to automatically generate interactive metabolic pathway diagrams directly within the web interface. These pathway visualizations dynamically represent the GBA analysis results from the first growth condition, with line thicknesses and node sizes proportionally reflecting flux magnitudes and concentration values, respectively. This feature enhances biological understanding, helping researchers to quickly identify key metabolic features or potential pathway errors. Users can interact with the pathway map by zooming, dragging nodes to refine the layout, and viewing detailed information on metabolites and reactions. For collaboration and documentation purposes, users can save pathway configurations as JSON files for future editing or export them as publication-ready SVG files.

In addition to the visual outputs, users can download the numerical optimization results as a CSV file, supporting rigorous statistical analysis and custom visualization.

4.3 Results

Case study: example of a GBA model using *Cell Growth Simulator*

We next provide an example of a GBA model to showcase the practical application of *Cell Growth Simulator*, using a streamlined GBA model previously reported in the literature (Dourado et al., 2023). This model consists of 4 reactions and 4 internal reactants, with two redundant reactions, only one of which is active under optimal growth conditions.

The model assumes irreversible Michaelis-Menten kinetics, where the turnover times (τ) are determined by a matrix (K) of Michaelis constants and a vector (k_{cat}) of forward turnover numbers for each reaction. Fig 2a-c illustrates the schematic representation of the model, along with its parameters (M , K , k_{cat} , and ρ).

While most model parameters are arbitrary for demonstration purposes, several key values are based on empirical data: the ribosome reaction parameters ($k_{cat} = 4.55 \text{ h}^{-1}$ and $K_m = 8.3 \text{ g L}^{-1}$ for its primary substrate) reflect *E. coli* estimates (Dourado & Lercher, 2020), and the cell density ($\rho = 340 \text{ g L}^{-1}$) reflects the measured *E. coli* dry mass density (Zimmerman & Trach, 1991). This example deliberately omits regulatory mechanisms, with activation (KA) and inhibition (KI) matrices set to zero.

Users can examine the complete model in the "Model Preview" section (Fig 2d), where they can also modify parameters and save customized versions. The "Check Model" function validates parameter integrity before optimization, while the "Run" button initiates numerical optimization using the SLSQP solver. Upon completion, *Cell Growth Simulator* generates a comprehensive overview of key results, such as the number of simulated growth conditions, the number of converged optimizations, the maximum achieved growth rate, and the average total protein concentration (Fig 2e). In the "Interactive Plots" panel, *Cell Growth Simulator* displays customizable plots where the x-axis can represent external reactant concentrations or growth rate (μ), while the y-axis can display various metrics including growth rates, protein fractions, protein concentrations, reaction fluxes, and metabolite concentrations.

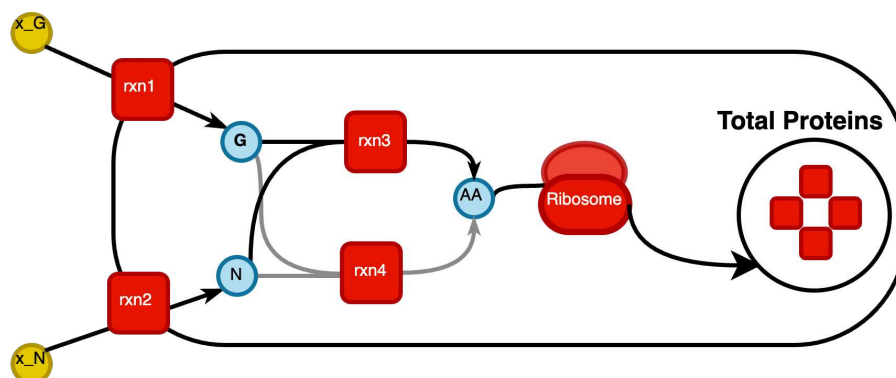
Analysis of the example model shows that growth rates at different external concentrations are consistent with the Monod equation (Fig 2g). The results show an approximately linear relationship between ribosomal mass fraction ϕ_r and growth rate μ . The protein fraction allocated to the transporter G decreases with the increasing external concentration x_G , corroborating previous theoretical findings regarding the efficiency of carbon source utilization (Burnap, 2015; Faizi et al., 2018; Molenaar et al., 2009; Weiße et al., 2015). These relationships reflect fundamental "bacterial growth laws," supporting the notion that optimal resource allocation drives maximal growth across diverse environmental conditions (Scott et al., 2010a; Scott & Hwa, 2011; You et al., 2013).

Analysis of reaction activities reveals a dynamic metabolic shift between rxn_3 and rxn_4: rxn_3 is active at lower growth rates (up to 1.58 h^{-1}), while rxn_4 is utilized at higher growth rates (Fig 2f). Both reactions share a similar structural role in the matrix M (linearly dependent columns when considering only the 4 internal reactants) but have different kinetic parameters, resulting in different resource allocation costs; this favors the sole use of rxn_3 at lower growth rates and the sole use of rxn_4 at higher growth rates.

The "Pathway Visualization" panel offers an intuitive graphical representation of the reaction network (Fig 2h), with visual elements scaled proportionally to simulation results from the first growth condition (representing maximal external concentrations in this example). Node sizes of metabolites reflect metabolite concentrations, while line thicknesses indicate protein concentrations. The redundant pathway, rxn_3 in this case, is highlighted using a dashed line.

Users can export the complete optimization results as a ".csv" file through the "Results" button in the main panel.

a)



b)

	rxn1	rxn2	rxn3	rxn4	Rib
x_G	-1	0	0	0	0
x_N	0	-1	0	0	0
G	1	0	-0.7	-0.8	0
N	0	1	-0.3	-0.2	-0.2
AA	0	0	1	1	-0.8
Prot	0	0	0	0	1

c)

K _m	rxn1	rxn2	rxn3	rxn4	Rib
x_G	10	0	0	0	0
x_N	0	10	0	0	0
G	0	0	10	5	0
N	0	0	10	5	10
AA	0	0	0	0	8.3
Prot	0	0	0	0	0

k _{cat}	rxn1	rxn2	rxn3	rxn4	Rib
k _{cat_f}	30	20	20	18	4.55
k _{cat_b}	0	0	0	0	0

d)

Your GBA Model

Mass fraction Km KI KA kcat Condition

Your Mass Fraction (M) matrix

	rxn1	rxn2	rxn3	rxn4	Ribosome
x_G	-1	0	0	0	0
x_N	0	-1	0	0	0
G	1	0	-0.7	-0.8	0
N	0	1	-0.3	-0.2	-0.2
AA	0	0	1	1	-0.8
Protein	0	0	0	0	1

e)

Number of Conditions 15	Maximum Growth rate 2.2488 h⁻¹
Average of Total Protein Concentration 203.28 [g L⁻¹]	Number of Converged Optimization 15

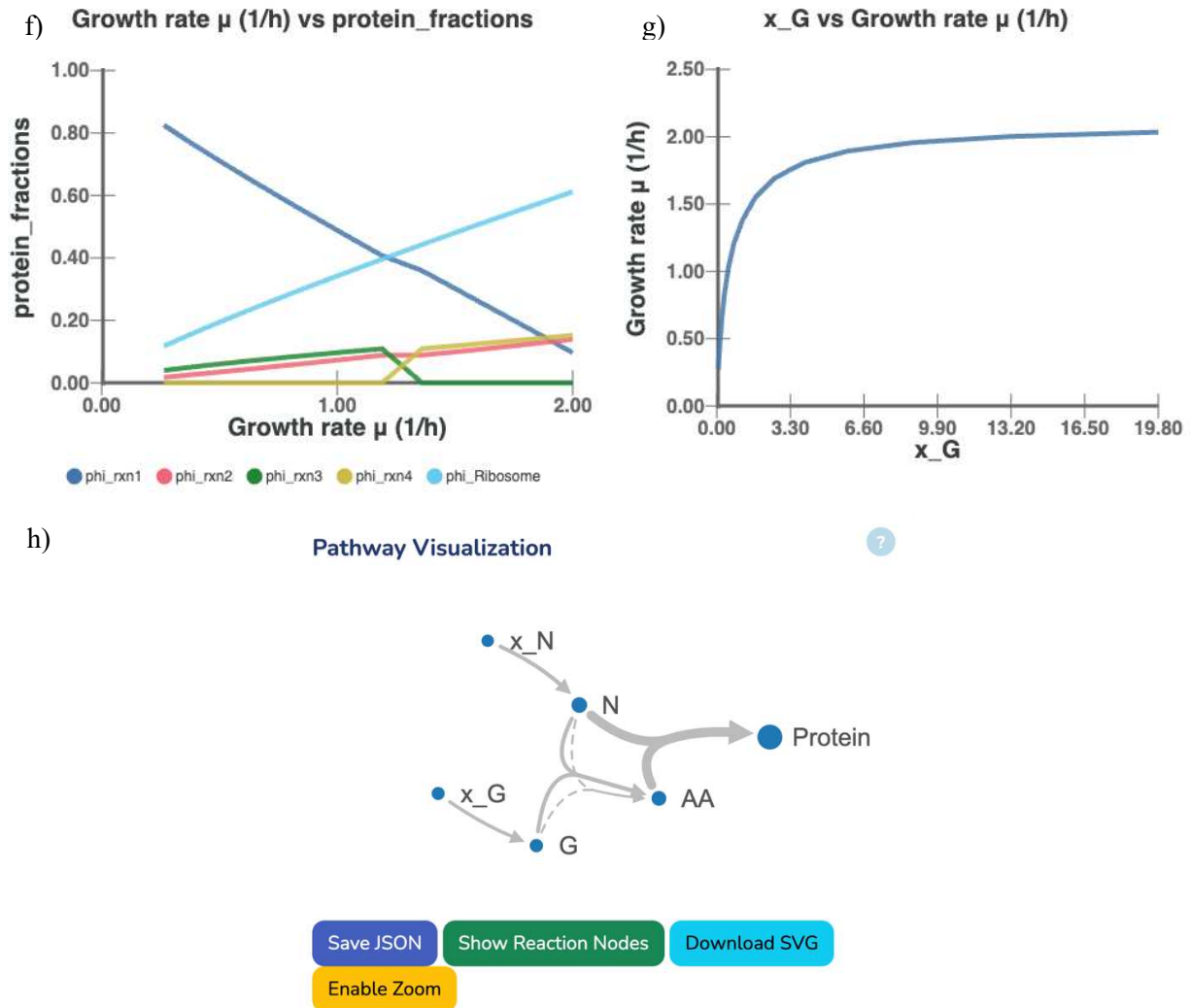


Fig 2. Analysis of an example GBA model using *Cell Growth Simulator*. (a) Schematic representation of the simplified GBA model, illustrating reaction pathways, metabolite interactions, and redundant routes. (b) Mass fraction matrix (M), detailing the stoichiometric relationships between reactants (rows) and reactions (columns), with negative values indicating consumption and positive values indicating production. Note that in panel b-c, “Ribosome” is shortened to “Rib” and “Protein” to “Prot” for clarity. (c) Kinetic parameter set comprising the Michaelis constant matrix (K) and turnover number vector (k_{cat}). (d) *Cell Growth Simulator* model preview interface displaying imported parameters and allowing for real-time modification before analysis. (e) Results summary dashboard presenting key optimization outcomes: total growth conditions analyzed, number of successfully converged optimizations, maximum achieved growth rate, and average total protein concentration. (f) Optimal protein allocation across varying growth rates. (g) Monod-like growth curve in response to external carbon source availability. (h) Interactive metabolic pathway visualization with node sizes proportional to metabolite concentrations and edge thicknesses reflecting protein allocation, with dashed lines indicating redundant pathways.

4.4 Discussion

We introduce *Cell Growth Simulator*, a user-friendly web application for Growth Balance Analysis (GBA) that enhances accessibility for researchers interested in exploring cellular growth and resource allocation under nonlinear kinetic rate laws. *Cell Growth Simulator* provides an intuitive interface built with R/Shiny, supplemented by HTML, CSS, and JavaScript to enhance the user experience. The application allows users to construct and modify models through interactive matrices that resemble spreadsheet environments (here, .ods files), a familiar open-source format for many researchers. This design choice simplifies the process of inputting and managing model parameters – such as the mass fraction matrix (M), kinetic parameters (K_m and k_{cat}), cell density, and external growth conditions – without requiring direct coding. *Cell Growth Simulator* significantly streamlines the incorporation of kinetic parameters into models by integrating data tables retrieved from the BRENDA enzyme database (Chang et al., 2021).

Cell Growth Simulator has an emphasis on customizable and interactive visualization to aid in the interpretation of optimization results. By integrating "d3flux", a package specifically designed for metabolic pathway visualization, and the "Apexcharter" package for generating interactive plots, the application provides a dynamic and informative representation of the metabolic network and the simulation results.

The GBA formalism itself imposes no intrinsic limit on model size. In principle, it can be applied to networks ranging from minimal self-replicators to genome-scale models. However, we recommend using the *Cell Growth Simulator* with coarse-grained models that have fewer than ~20 reactions and internal species. At this scale, models remain interpretable, optimization is feasible with standard solvers, and interactive visualization is responsive on modest hardware. While larger models can be constructed within the same framework, they are better explored through programmatic workflows (e.g., R/Python notebooks) coupled with more powerful numerical solvers. Future development will focus on improving computational efficiency and providing programmatic interfaces to facilitate the analysis of larger models.

Cell Growth Simulator addresses key challenges in analyzing self-replicating cell models with nonlinear kinetics, making such fundamental analyses accessible to a wide range of researchers and fostering interdisciplinary collaboration by lowering technical barriers. Furthermore, *Cell Growth Simulator's* approach aligns with the growing emphasis on open science and reproducibility. By offering a platform that is both user-friendly and transparent, it supports the sharing of models and results, facilitating peer review and collaborative contributions to advancing our understanding of cellular metabolism and growth.

Availability and Future Directions

Cell Growth Simulator is freely available as a web server at <https://cellgrowthsim.com/>, where researchers can analyze and share their models without additional registration or login requirements. Looking ahead, we plan to expand *Cell Growth Simulator*'s capabilities in two key areas. First, although the underlying GBA formalism does not restrict model size, we will improve computational efficiency by using better solvers and parallelization strategies. This will make it possible to study larger, more detailed models than the coarse-grained scale recommended here. Second, we intend to incorporate dynamical simulations so that users can study time-dependent behavior and transient responses in metabolic and proteomic networks. By coupling these dynamic features with coarse-grained approaches for cellular resource allocation, *Cell Growth Simulator* will provide both steady-state and temporal insights into how cells adapt and grow under varying conditions. We welcome community-driven enhancements and collaborative projects to further improve *Cell Growth Simulator*'s functionality, usability, and applications in diverse biological contexts.

Data availability

Cell Growth Simulator is a web server freely accessible without login requirement at <https://cellgrowthsim.com/>.

The source code for *Cell Growth Simulator* is available at: <https://github.com/Sijr73/CellGrowthSimulator>.

Acknowledgement

We thank Martin J. Lercher for many valuable discussions and suggestions. We are also grateful to Thomas Spitzlei for his support in deploying the application online.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) [CRC 1310 and, under Germany's Excellence Strategy, EXC 2048/1–390686111] and by the European Union [ERC AdG “MechSys”–Project ID 101055141].

Conflict of interest

The authors declare no conflict of interest.

4.5 References

- Andreas Neudecker. (2019). *shinyMatrix: Shiny Matrix Input Field* (p. 0.8.0) [Dataset].
<https://doi.org/10.32614/CRAN.package.shinyMatrix>
- Burnap, R. L. (2015). Systems and Photosystems: Cellular Limits of Autotrophic Productivity in Cyanobacteria. *Frontiers in Bioengineering and Biotechnology*, 3.
<https://doi.org/10.3389/fbioe.2015.00001>
- Chang, A., Jeske, L., Ulbrich, S., Hofmann, J., Koblitz, J., Schomburg, I., Neumann-Schaal, M., Jahn, D., & Schomburg, D. (2021). BRENDA, the ELIXIR core data resource in 2021: New developments and updates. *Nucleic Acids Research*, 49(D1), D498–D508.
<https://doi.org/10.1093/nar/gkaa1025>
- Choi, K., Medley, J. K., König, M., Stocking, K., Smith, L., Gu, S., & Sauro, H. M. (2018). Tellurium: An extensible python-based modeling environment for systems and synthetic biology. *Biosystems*, 171, 74–79.
<https://doi.org/10.1016/j.biosystems.2018.07.006>
- Doan, D. T., Hoang, M. D., Heins, A.-L., & Kremling, A. (2022). Applications of Coarse-Grained Models in Metabolic Engineering. *Frontiers in Molecular Biosciences*, 9, 806213. <https://doi.org/10.3389/fmolb.2022.806213>
- Dourado, H., & Lercher, M. J. (2020). An analytical theory of balanced cellular growth. *Nature Communications*, 11(1), 1226. <https://doi.org/10.1038/s41467-020-14751-w>
- Dourado, H., Liebermeister, W., Ebenhöf, O., & Lercher, M. J. (2023). Mathematical properties of optimal fluxes in cellular reaction networks at balanced growth. *PLOS Computational Biology*, 19(6), e1011156.
<https://doi.org/10.1371/journal.pcbi.1011156>
- Erickson, D. W., Schink, S. J., Patsalo, V., Williamson, J. R., Gerland, U., & Hwa, T. (2017). A global resource allocation strategy governs growth transition kinetics of *Escherichia coli*. *Nature*, 551(7678), 119–123. <https://doi.org/10.1038/nature24299>
- Faizi, M., Zavřel, T., Loureiro, C., Červený, J., & Steuer, R. (2018). A model of optimal protein allocation during phototrophic growth. *Biosystems*, 166, 26–36.
<https://doi.org/10.1016/j.biosystems.2018.02.004>
- Fröhlich, F., Weindl, D., Schälte, Y., Pathirana, D., Paszkowski, Ł., Lines, G. T., Stapor, P., & Hasenauer, J. (2021). AMICI: High-performance sensitivity analysis for large ordinary differential equation models. *Bioinformatics*, 37(20), 3676–3677.
<https://doi.org/10.1093/bioinformatics/btab227>
- Ghaffarinasab, S., Lercher, M. J., & Dourado, H. (2023). *Growth balance analysis models of cyanobacteria for understanding resource allocation strategies*.
<https://doi.org/10.1101/2023.04.18.537335>

- Goelzer, A., & Fromion, V. (2011). Bacterial growth rate reflects a bottleneck in resource allocation. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1810(10), 978–988. <https://doi.org/10.1016/j.bbagen.2011.05.014>
- Hari, A., & Lobo, D. (2020). Fluxer: A web application to compute, analyze and visualize genome-scale metabolic flux networks. *Nucleic Acids Research*, 48(W1), W427–W435. <https://doi.org/10.1093/nar/gkaa409>
- Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., & Kummer, U. (2006). COPASI—a COmplex PAthway SIMulator. *Bioinformatics*, 22(24), 3067–3074. <https://doi.org/10.1093/bioinformatics/btl485>
- Hui, S., Silverman, J. M., Chen, S. S., Erickson, D. W., Basan, M., Wang, J., Hwa, T., & Williamson, J. R. (2015). Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Molecular Systems Biology*, 11(2), 784. <https://doi.org/10.15252/msb.20145697>
- King, Z. A., Dräger, A., Ebrahim, A., Sonnenschein, N., Lewis, N. E., & Palsson, B. O. (2015). Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. *PLOS Computational Biology*, 11(8), e1004321. <https://doi.org/10.1371/journal.pcbi.1004321>
- Liebermeister, W., & Klipp, E. (2006). Bringing metabolic networks to life: Convenience rate law and thermodynamic constraints. *Theoretical Biology and Medical Modelling*, 3(1), 41. <https://doi.org/10.1186/1742-4682-3-41>
- Lin, G. (2019). *reactable: Interactive Data Tables for R* (p. 0.4.4) [Dataset]. <https://doi.org/10.32614/CRAN.package.reactable>
- Mao, Z., Yuan, Q., Li, H., Zhang, Y., Huang, Y., Yang, C., Wang, R., Yang, Y., Wu, Y., Yang, S., Liao, X., & Ma, H. (2023). CAVE: A cloud-based platform for analysis and visualization of metabolic pathways. *Nucleic Acids Research*, 51(W1), W70–W77. <https://doi.org/10.1093/nar/gkad360>
- Martyushenko, N., & Almaas, E. (2019). ModelExplorer—Software for visual inspection and inconsistency correction of genome-scale metabolic reconstructions. *BMC Bioinformatics*, 20(1), 56. <https://doi.org/10.1186/s12859-019-2615-x>
- Molenaar, D., Van Berlo, R., De Ridder, D., & Teusink, B. (2009). Shifts in growth strategies reflect tradeoffs in cellular economics. *Molecular Systems Biology*, 5(1), 323. <https://doi.org/10.1038/msb.2009.82>
- O’Brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R., & Palsson, B. Ø. (2013). Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular Systems Biology*, 9(1), 693. <https://doi.org/10.1038/msb.2013.52>
- Olivier, B. G., & Snoep, J. L. (2004). Web-based kinetic modelling using JWS Online. *Bioinformatics*, 20(13), 2143–2144. <https://doi.org/10.1093/bioinformatics/bth200>

- Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis? *Nature Biotechnology*, 28(3), 245–248. <https://doi.org/10.1038/nbt.1614>
- Perrier, V., & Meyer, F. (2019). *apexcharter: Create Interactive Chart with the JavaScript “ApexCharts” Library* (p. 0.4.4) [Dataset]. <https://doi.org/10.32614/CRAN.package.apexcharter>
- Sánchez, B. J., Zhang, C., Nilsson, A., Lahtvee, P., Kerkhoven, E. J., & Nielsen, J. (2017). Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Molecular Systems Biology*, 13(8), 935. <https://doi.org/10.15252/msb.20167411>
- Schaff, J., Fink, C. C., Slepchenko, B., Carson, J. H., & Loew, L. M. (1997). A general computational framework for modeling cellular structure and function. *Biophysical Journal*, 73(3), 1135–1146. [https://doi.org/10.1016/S0006-3495\(97\)78146-3](https://doi.org/10.1016/S0006-3495(97)78146-3)
- Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z., & Hwa, T. (2010a). Interdependence of Cell Growth and Gene Expression: Origins and Consequences. *Science*, 330(6007), 1099–1102. <https://doi.org/10.1126/science.1192588>
- Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z., & Hwa, T. (2010b). Interdependence of Cell Growth and Gene Expression: Origins and Consequences. *Science*, 330(6007), 1099–1102. <https://doi.org/10.1126/science.1192588>
- Scott, M., & Hwa, T. (2011). Bacterial growth laws and their applications. *Current Opinion in Biotechnology*, 22(4), 559–565. <https://doi.org/10.1016/j.copbio.2011.04.014>
- Shaikh, B., Marupilla, G., Wilson, M., Blinov, M. L., Moraru, I. I., & Karr, J. R. (2021). RunBioSimulations: An extensible web application that simulates a wide range of computational modeling frameworks, algorithms, and formats. *Nucleic Acids Research*, 49(W1), W597–W602. <https://doi.org/10.1093/nar/gkab411>
- Thiele, S., Von Kamp, A., Bekiaris, P. S., Schneider, P., & Klamt, S. (2022). CNApy: A CellNetAnalyzer GUI in Python for analyzing and designing metabolic networks. *Bioinformatics*, 38(5), 1467–1469. <https://doi.org/10.1093/bioinformatics/btab828>
- Weiß, A. Y., Oyarzún, D. A., Danos, V., & Swain, P. S. (2015). Mechanistic links between cellular trade-offs, gene expression, and growth. *Proceedings of the National Academy of Sciences*, 112(9). <https://doi.org/10.1073/pnas.1416533112>
- You, C., Okano, H., Hui, S., Zhang, Z., Kim, M., Gunderson, C. W., Wang, Y.-P., Lenz, P., Yan, D., & Hwa, T. (2013). Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature*, 500(7462), 301–306. <https://doi.org/10.1038/nature12446>
- Ypma, J., & Johnson, S. G. (2011). *nloptr: R Interface to NLOpt* (p. 2.1.1) [Dataset]. <https://doi.org/10.32614/CRAN.package.nloptr>
- Zimmerman, S. B., & Trach, S. O. (1991). Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *Journal of Molecular Biology*, 222(3), 599–620. [https://doi.org/10.1016/0022-2836\(91\)90499-V](https://doi.org/10.1016/0022-2836(91)90499-V)

Chapter 5

Discussion and Outlook

5.1 Discussion

This work had three main objectives related to the study of cyanobacterial metabolism and cellular resource allocation: first, to demonstrate how the extensive metabolic networks of cyanobacteria confer a remarkable ability to adapt to diverse environments, highlighting their ecological importance and industrial potential; second, to deepen our understanding of how these organisms allocate resources as a function of environmental conditions, by employing the growth balance analysis (GBA) framework that explicitly captures the intracellular processes underlying photoautotrophic growth; and third, to develop *Cell Growth Simulator*, a user-friendly framework that facilitates the modeling and analysis of resource allocation within such systems.

Together, these studies shed light on different aspects of cyanobacterial adaptation. At the genome/network evolution level, the versatile metabolic repertoire of cyanobacteria provides them with adaptive flexibility: with relative ease, they can acquire the biochemical pathways needed for new environments. At the physiological level, they demonstrate adaptive flexibility by reallocating protein resources to optimize growth under specific conditions. When shifting from light-limited to light-saturated conditions, cyanobacteria exhibit a decrease in the proteome fractions of photosystem I (PSI) and photosystem II (PSII) along with an increase in ribosome protein fraction, mirroring experimental observations (Faizi et al., 2018; Zavřel et al., 2019). Importantly, the evolutionary and physiological scales likely reinforce one another. A larger metabolic network not only allows growth on varied substrates (long-term adaptation), but also offers redundancy or alternative pathways that cells can differentially regulate (short-term acclimation). Thus, the overarching theme is that resource allocation and adaptive evolution are interlinked: cyanobacteria that maintain a broad enzymatic repertoire adapt faster over evolutionary timescales and cope better with fluctuating conditions through nimble resource redistribution.

Our finding that metabolic complexity accelerates evolvability challenges long-standing assumptions in evolutionary theory. According to Fisher's model and subsequent analyses (Fisher, 1930; Orr, 2005), mutations in more complex organisms are less likely to be beneficial due to widespread pleiotropic side effects. However, by focusing on metabolic networks, we found the opposite trend: more complex networks are more adaptable. This apparent contradiction can be explained by considering the biological context. Empirical evidence from protein-protein interaction (PPI) networks supports Fisher's view that genes that code for highly connected proteins are rarely gained through horizontal transfer. This is presumably because successful integration would require the concurrent co-evolution of multiple binding sites, suggesting that complexity is costly (Cohen et al., 2011; Jain et al., 1999). In metabolic networks, however, enzymes largely function independently, so adding a new enzyme can plug into pathways without requiring co-evolved binding sites. The adaptive value of an enzyme lies

primarily in its catalytic activity, which can often be harnessed if the regulatory mechanisms adjust accordingly (Lenski, 2017; Lozada-Chávez et al., 2006). Indeed, prior empirical work on *E. coli* revealed that nearly all metabolic innovations in its lineage emerged through the acquisition of single DNA segments, frequently leveraging pre-existing "promiscuous" enzymes or substrates (Szappanos et al., 2016).

Our pan-genome analysis generalized this concept across diverse bacteria, including cyanobacteria, and highlighted the role of exaptations and collateral adaptation. Generalists often gained the ability to grow in additional environments through one adaptation, which is a form of collateral benefit that has also been observed in network simulations by Barve and Wagner (2013). Specialists, on the other hand, showed little collateral adaptation. However, when forced to adapt, they tended to reuse previously gained functions in a stepwise fashion, resulting in higher exaptation indices. This finding aligns with the idea that rare expansions from a specialist niche likely occur through the sequential building on prior gains. An unexpected outcome was the magnitude of the difference. For example, the endosymbiont *Buchnera aphidicola* required an average of 52 new reactions to survive in a new medium, whereas *Synechocystis* sp. PCC 6803 and *Synechococcus elongatus* PCC 7942, despite being obligate photoautotrophs, required only three to four additions. Thus, even phototrophic cyanobacteria, which are often viewed as niche specialists, exhibit considerable latent metabolic potential. This underscores how their two-billion-year evolutionary history in fluctuating environments may have preserved metabolic complexity as a hedge against change (Cao et al., 2020). Our results provide a mechanistic basis for anecdotal observations of the versatility of cyanobacteria (e.g., their ability to tolerate extreme habitats) by revealing a robust metabolic network that defies the typical specialization trade-off.

The GBA modeling results align with and extend prior studies on cyanobacterial physiology. Molenaar et al. (2009) and subsequent studies have proposed that microbes allocate proteins in a way that optimizes growth-limiting processes (Burnap, 2015; Faizi et al., 2018; Jahn et al., 2018; Zavřel et al., 2019). For *Synechocystis*, the trade-off between light harvesting and carbon assimilation is an example of this: under low light conditions, *Synechocystis* allocates most of the resources to photosynthetic components, whereas under carbon-limited or high-light conditions, investment shifts to metabolism and biosynthesis. The smooth reallocation exhibited by our model is consistent with continuous cultures of cyanobacteria grown under different light intensities (Jahn et al., 2018; Zavřel et al., 2019). Our extended model demonstrated the ability to capture the reduction in total cellular protein content with faster growth, a trend that was observed in experiments but not explained by earlier models that assumed a fixed proteome size (Goelzer & Fromion, 2011; O'Brien et al., 2013; Sánchez et al., 2017). In our framework, this behavior emerges naturally. At higher light intensities, despite higher growth rates, less pigment and antenna protein are needed, which frees up mass for other cell components. Additionally, photodamage at very high light levels necessitates the allocation of some resources to ribosomal

protein fractions instead to net growth. Compared to a previous coarse-grained nonlinear model of *Synechocystis* (Faizi et al., 2018), the GBA formulation proved to be mathematically simpler yet equally effective in reproducing observed growth trends and proteome allocation under different light intensities. This streamlined approach is particularly advantageous for exploring more sophisticated phototrophic behaviors and extending the model to include additional cellular processes, such as major photosynthetic complexes or carbon fixation.

A useful aspect of this study is its integrated perspective on adaptation. We linked evolutionary innovation and physiological acclimation under the unifying theme of resource allocation. Using computational models at both scales allowed us to compare our findings with a broad spectrum of previous studies, ranging from theoretical models of evolvability to laboratory measurements of proteomes, and demonstrate that a consistent picture emerges. We showed that from a minimal set of governing principles (mass balance, density, and kinetic constraints), an explanation of diverse phenomena emerges, such as the long-term diversification of bacterial lineages and the short-term acclimation of a cyanobacterial culture. Furthermore, we introduced *Cell Growth Simulator*, a practical innovation and user-friendly web server that provides an accessible interface for these complex nonlinear models. It lowers the barrier for other researchers to apply growth balance analysis to their systems. Although GBA offers a robust approach for modeling self-replicating cells – explicitly incorporating metabolite concentrations and their effects on reaction fluxes – its implementation previously demanded significant programming expertise and computational resources. *Cell Growth Simulator* uses an intuitive spreadsheet interface, integrated kinetic parameter retrieval from the BRENDA enzyme database (Chang et al., 2021), and interactive visualization tools. This platform not only makes nonlinear modeling of resource allocation in coarse-grained cellular systems accessible to researchers with limited programming skills, but it also serves as a valuable tool for fostering interdisciplinary collaboration and enriching our understanding of cellular metabolism and growth. In summary, the combination of conceptual advances (pan-genomic analysis of adaptation), mechanistic modeling (growth balance analysis of phototrophic growth), and methodological tools (*Cell Growth Simulator*) constitutes a solid contribution to understanding cyanobacterial systems.

However, these contributions come with certain limitations that must be acknowledged. First, the pan-genome metabolic modeling in our adaptation study relied on available genome-scale models of the BiGG database and simulated “virtual HGT” events (Schellenberger et al., 2010). While this approach is effective for scanning thousands of scenarios, its accuracy depends on the quality of the underlying metabolic reconstructions. Some cyanobacterial models may lack pathways, especially regulatory or stress response pathways, that in nature would affect viability. For instance, our definition of a “viable environment” focused solely on the metabolic ability to synthesize biomass precursors or energy generation, and did not consider other growth-limiting factors, such as light, temperature, and pH levels. In nature, cyanobacteria face multifaceted

challenges, and adaptability involves more than just metabolism (*e.g.*, regulatory plasticity and physiological tolerance).

One shortcoming of our adaptability metric is that it is narrow. A strain may appear metabolically adaptable according to our criteria, yet still fail to colonize a niche due to ecological constraints, *e.g.*, unfavorable light spectra or high turbidity/UV exposure, and salinity or osmotic shocks. A key limitation in resource allocation modeling is the coarse-grained approach of GBA models. To keep the models mathematically simple, complex processes (such as the Calvin cycle or diverse metabolic pathways) are lumped into single reactions. Consequently, the models cannot capture certain details, such as the regulation of individual enzymes, the spatial organization of thylakoids, and the diurnal cycling of metabolism. Although the model accurately reproduced steady-state trends, it cannot yet describe transient dynamics or regulatory phenomena during acclimation. Finally, the *Cell Growth Simulator* itself, while user-friendly, is limited to relatively small models (≤ 20 reactions) and steady-state conditions. Scaling up to genome-scale kinetic models remains computationally challenging, and users must supply many kinetic parameters; this is a non-trivial task given the limited experimental data available for many enzymes, although machine learning models for predicting kinetic parameters from amino acid sequences and chemical reaction representations can fill this gap (Kroll et al., 2021, 2023).

Despite these limitations, our holistic approach has revealed a coherent narrative of cyanobacterial adaptation. Organisms that were once considered photosynthetic specialists now seem to be well-positioned to expand their ecological and biochemical roles. Their genomes encode a variety of metabolic functions, making them generalists in potential. It may then seem surprising that not more heterotrophs have evolved from cyanobacterial ancestors; a potential reason is the ecological superiority of at least facultative photosynthesis, though more research may be needed to settle this question. The metabolic adaptability of cyanobacteria helps explain how these organisms have managed to colonize environments ranging from hot springs to polar lakes over geologic time (Rybak et al., 2024; Wejnerowski et al., 2023). We also learned that, much like other unicellular organisms, cyanobacteria allocate their cellular resources in a highly optimized manner constrained by their phototrophic lifestyle (*e.g.*, they have to balance light harvesting components and carbon fixation). Our study notably showed that the frequently discussed trade-off between light-harvesting and growth machinery in phototrophs can be quantified and predicted. The strong correspondence between the predicted proteome fractions of our model and the empirical measurements builds confidence in our understanding of the allocative priorities of these cells.

These mechanistic insights are not only academically interesting but also have practical implications. For instance, when bioengineering cyanobacteria to produce biofuels or chemicals, one could use these findings to ensure that the engineered pathways do not disrupt the cell's

balanced allocation. Alternatively, one could use synthetic biology to re-engineer the allocation to favor production over growth.

5.2 Future directions

This work opens several avenues for further investigation. One clear direction is to bridge the evolutionary and physiological scales more directly. While we studied them separately, an integrated model could simulate how a cyanobacterium's growth advantage in a new environment (as predicted by GBA) translates into a selection advantage that drives genome evolution. For instance, one could incorporate our resource allocation model into an evolutionary simulation to test which new metabolic gene, when acquired, results in the maximum growth under specific conditions and whether that aligns with the historical sequence of gene acquisitions in cyanobacterial lineages. This would require expanding GBA models to include alternative metabolic routes that are initially "absent" and then "added" to simulate horizontal gene transfer (HGT) – an approach that our pan-genomic analysis already employed in this study. Combining these approaches would result in a multi-scale evolutionary simulation in which the genotype (network) and phenotype (allocation) co-evolve.

Therefore, one next logical step is to develop genome-scale GBA models that capture the full metabolic complexity of cyanobacteria. Current GBA models (including the *Synechocystis* model developed here) focus on relatively small, coarse-grained networks. In contrast, genome-scale metabolic reconstructions already exist for cyanobacteria, providing a comprehensive definition of the stoichiometry of hundreds of reactions (Höper et al., 2024; Knoop et al., 2013; Montagud et al., 2011; Nogales et al., 2012). In addition, a genome-scale GBA model would require extensive kinetic data, such as turnover numbers, Michaelis constants, and activation/inhibition constants, for potentially thousands of enzymes. Acquiring the data of this scale is a formidable challenge because experimentally measured enzyme parameters cover only a small fraction of reactions, even in well-studied bacteria. Recent advances in machine learning predictions of the properties of enzymes and transporters indicate that this significant challenge is becoming addressable (Heckmann et al., 2018; Kroll et al., 2021). Using these predictions to parameterize proteome-constrained models significantly improved their accuracy (Li et al., 2022). These breakthroughs suggest that creating a genome-scale GBA model for cyanobacteria is becoming feasible. Such a model would expand the scope of GBA to include every metabolic pathway and require new computational strategies to handle the high dimensionality and nonlinearity of a whole-cell kinetic model. New numerical optimization strategies, including dedicated high-performance solvers, are emerging to keep genome-scale kinetic models solvable. A particularly promising development is GBACpp (<https://github.com/charlesrocabert/gbacpp>), a high-performance C++ solver that combines gradient-ascent optimization with parallelization and optional global-search heuristics developed at the Computational Cell Biology group at HHU. This approach delivers speed increases of orders of magnitude for GBA calculations.

Once achieved, a genome-scale GBA model could effectively unite the system-level insights of FBA (*e.g.*, optimal flux distributions) with the detailed resource allocation mechanics of GBA, providing a more holistic platform for exploring cyanobacterial physiology.

Second, modeling frameworks such as FBA, RBA, and GBA at steady-state conditions may not fully capture the dynamics of organisms whose metabolism varies over time. This is exemplified by the pronounced diurnal (day/night) cycles in cyanobacteria. As cyanobacterial growth and metabolism are tied to the light/dark cycle, assuming a constant steady state can overlook important regulatory and metabolic shifts. Therefore, extending GBA to include time-dependent phenomena beyond steady-state growth is a promising direction. Initial efforts in this area show great potential. For example, Reimers et al. (2017) developed a constraint-based framework that uses a genome-scale model of a *Synechococcus* to simulate optimal resource allocation over a full diurnal cycle. Their time-resolved model could predict cyclic patterns, such as glycogen storage, across day/night phases. These predictions were in qualitative agreement with experimental observations. Future GBA models could build on such approaches by incorporating dynamic regulation to better reflect how cells reallocate resources under cyclic and fluctuating conditions. These dynamic, GBA models would provide a more complete picture of how cyanobacteria balance metabolic demands over time and offer deeper insight into processes such as the circadian regulation of metabolism. This knowledge could inform more refined strategies for metabolic engineering and biotechnology. For instance, it could help identify the optimal time or conditions for inducing the production of a desired bioproduct in cyanobacterial cultures.

Finally, as the GBA framework scales up to the genome level, maintaining accessibility and user-friendliness will be a key challenge. Genome-scale kinetic models are computationally intensive and complex. Therefore, future versions of our *Cell Growth Simulator* web-server should be improved to efficiently handle larger networks while remaining easy to use. Using more efficient optimization algorithms, such as better nonlinear solvers or decomposition methods (such as Augmented-Lagrangian (Birgin & Martínez, 2008; Conn et al., 1991)), and leveraging parallel computing could significantly reduce run times for large GBA models. Additionally, enhanced features such as automated parameter estimation to fit or refine kinetic parameters from data, expanded sensitivity analysis tools, and direct interfaces to genome-scale or enzyme databases to fetch reaction kinetics would streamline the modeling workflow. Crucially, these upgrades should preserve and build on the current design philosophy of accessibility in the *Cell Growth Simulator*. By lowering computational and technical barriers even further, an improved *Cell Growth Simulator* could support iterative model building and refinement at the genome scale while remaining accessible to experts and new users alike. We envision that, in the future, *Cell Growth Simulator* will become a central hub for collaborative, iterative, genome-scale modeling of cyanobacterial metabolism, enabling researchers to easily test hypotheses, integrate new data, and advance our understanding of these ecologically and biotechnologically important microbes.

References

- Angermayr, S. A., Hellingwerf, K. J., Lindblad, P., & Teixeira de Mattos, M. J. (2009). Energy biotechnology with cyanobacteria. *Current Opinion in Biotechnology*, 20(3), 257–263. <https://doi.org/10.1016/j.copbio.2009.05.011>
- Babele, P. K., Kumar, J., & Chaturvedi, V. (2019). Proteomic De-regulation in cyanobacteria in response to abiotic stresses. *Frontiers in Microbiology*, 10(JUN), 1–22. <https://doi.org/10.3389/fmicb.2019.01315>
- Barve, A., & Wagner, A. (2013). A latent capacity for evolutionary innovation through exaptation in metabolic systems. *Nature*, 500(7461), 203–206. <https://doi.org/10.1038/nature12301>
- Bernstein, D. B., Sulheim, S., Almaas, E., & Segrè, D. (2021). Addressing uncertainty in genome-scale metabolic model reconstruction and analysis. *Genome Biology*, 22(1), 64. <https://doi.org/10.1186/s13059-021-02289-z>
- Birgin, E. G., & Martínez, J. M. (2008). Improving ultimate convergence of an augmented Lagrangian method. *Optimization Methods Software*, 23(2), 177–195. <https://doi.org/10.1080/10556780701577730>
- Bothe, H., Schmitz, O., Yates, M. G., & Newton, W. E. (2010). Nitrogen Fixation and Hydrogen Metabolism in Cyanobacteria. *Microbiology and Molecular Biology Reviews : MMBR*, 74(4), 529–551. <https://doi.org/10.1128/MMBR.00033-10>
- Bremer, H., & Dennis, P. P. (2008). Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. *EcoSal Plus*, 3(1). <https://doi.org/10.1128/ecosal.5.2.3>
- Burnap, R. L. (2015). Systems and Photosystems: Cellular Limits of Autotrophic Productivity in Cyanobacteria. *Frontiers in Bioengineering and Biotechnology*, 3. <https://doi.org/10.3389/fbioe.2015.00001>
- Cao, H., Shimura, Y., Steffen, M. M., Yang, Z., Lu, J., Joel, A., Jenkins, L., Kawachi, M., Yin, Y., & Garcia-Pichel, F. (2020). The Trait Repertoire Enabling Cyanobacteria to Bloom Assessed through Comparative Genomic Complexity and Metatranscriptomics. *mBio*, 11(3), e01155-20. <https://doi.org/10.1128/mBio.01155-20>
- Chang, A., Jeske, L., Ulbrich, S., Hofmann, J., Koblitz, J., Schomburg, I., Neumann-Schaal, M., Jahn, D., & Schomburg, D. (2021). BRENDA, the ELIXIR core data resource in 2021:

- New developments and updates. *Nucleic Acids Research*, 49(D1), D498–D508.
<https://doi.org/10.1093/nar/gkaa1025>
- Cohen, O., Gophna, U., & Pupko, T. (2011). The Complexity Hypothesis Revisited: Connectivity Rather Than Function Constitutes a Barrier to Horizontal Gene Transfer. *Molecular Biology and Evolution*, 28(4), 1481–1489.
<https://doi.org/10.1093/molbev/msq333>
- Conn, A. R., Gould, N. I. M., & Toint, P. (1991). A Globally Convergent Augmented Lagrangian Algorithm for Optimization with General Constraints and Simple Bounds. *SIAM Journal on Numerical Analysis*, 28(2), 545–572. <https://doi.org/10.1137/0728030>
- Doan, D. T., Hoang, M. D., Heins, A.-L., & Kremling, A. (2022). Applications of Coarse-Grained Models in Metabolic Engineering. *Frontiers in Molecular Biosciences*, 9, 806213. <https://doi.org/10.3389/fmolb.2022.806213>
- Dodds, W. K. (2002). 16—Nutrient Use and Remineralization. In W. K. Dodds (Ed.), *Freshwater Ecology* (pp. 312–335). Academic Press. <https://doi.org/10.1016/B978-012219135-0/50017-9>
- Dourado, H., & Lercher, M. J. (2020a). An analytical theory of balanced cellular growth. *Nature Communications*, 11(1). <https://doi.org/10.1038/s41467-020-14751-w>
- Dourado, H., & Lercher, M. J. (2020b). An analytical theory of balanced cellular growth. *Nature Communications*, 11(1), 1226. <https://doi.org/10.1038/s41467-020-14751-w>
- Dourado, H., Liebermeister, W., Ebenhöf, O., & Lercher, M. J. (2023a). Mathematical properties of optimal fluxes in cellular reaction networks at balanced growth. *PLOS Computational Biology*, 19(6), e1011156. <https://doi.org/10.1371/journal.pcbi.1011156>
- Dourado, H., Liebermeister, W., Ebenhöf, O., & Lercher, M. J. (2023b). Mathematical properties of optimal fluxes in cellular reaction networks at balanced growth. *PLoS Computational Biology*, 19(6), 1–26. <https://doi.org/10.1371/journal.pcbi.1011156>
- Elseman, I. E., Rodriguez Prado, A., Grigaitis, P., Garcia Albornoz, M., Harman, V., Holman, S. W., van Heerden, J., Bruggeman, F. J., Bisschops, M. M. M., Sonnenschein, N., Hubbard, S., Beynon, R., Daran-Lapujade, P., Nielsen, J., & Teusink, B. (2022). Whole-cell modeling in yeast predicts compartment-specific proteome constraints that drive metabolic strategies. *Nature Communications*, 13(1), 801.
<https://doi.org/10.1038/s41467-022-28467-6>
- Faizi, M., Zavřel, T., Loureiro, C., Červený, J., & Steuer, R. (2018). A model of optimal protein allocation during phototrophic growth. *Biosystems*, 166, 26–36.
<https://doi.org/10.1016/j.biosystems.2018.02.004>

- Feist, A. M., & Palsson, B. O. (2010). The biomass objective function. *Current Opinion in Microbiology*, 13(3), 344–349. <https://doi.org/10.1016/j.mib.2010.03.003>
- Fisher, R. A. (1930). *The genetical theory of natural selection* (pp. xiv, 272). Clarendon Press. <https://doi.org/10.5962/bhl.title.27468>
- Ghaffarinasab, S., Lercher, M. J., & Dourado, H. (2023). *Growth balance analysis models of cyanobacteria for understanding resource allocation strategies*. <https://doi.org/10.1101/2023.04.18.537335>
- Goelzer, A., & Fromion, V. (2011). Bacterial growth rate reflects a bottleneck in resource allocation. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1810(10), 978–988. <https://doi.org/10.1016/j.bbagen.2011.05.014>
- Goelzer, A., Fromion, V., & Scorletti, G. (2011a). Cell design in bacteria as a convex optimization problem. *Automatica*, 47(6), 1210–1218. <https://doi.org/10.1016/j.automatica.2011.02.038>
- Goelzer, A., Fromion, V., & Scorletti, G. (2011b). Cell design in bacteria as a convex optimization problem. *Automatica*, 47(6), 1210–1218. <https://doi.org/10.1016/j.automatica.2011.02.038>
- Goelzer, A., Muntel, J., Chubukov, V., Jules, M., Prestel, E., Nölker, R., Mariadassou, M., Aymerich, S., Hecker, M., Noirot, P., Becher, D., & Fromion, V. (2015). Quantitative prediction of genome-wide resource allocation in bacteria. *Metabolic Engineering*, 32, 232–243. <https://doi.org/10.1016/j.ymben.2015.10.003>
- Grigorieva, G., & Shestakov, S. (1982). Transformation in the cyanobacterium *Synechocystis* sp. 6803. *FEMS Microbiology Letters*, 13(4), 367–370.
- Gu, C., Kim, G. B., Kim, W. J., Kim, H. U., & Lee, S. Y. (2019). Current status and applications of genome-scale metabolic models. *Genome Biology*, 20(1), 121. <https://doi.org/10.1186/s13059-019-1730-3>
- Hari, A., & Lobo, D. (2020). Fluxer: A web application to compute, analyze and visualize genome-scale metabolic flux networks. *Nucleic Acids Research*, 48(W1), W427–W435. <https://doi.org/10.1093/nar/gkaa409>
- Heckmann, D., Lloyd, C. J., Mih, N., Ha, Y., Zielinski, D. C., Haiman, Z. B., Desouki, A. A., Lercher, M. J., & Palsson, B. O. (2018). Machine learning applied to enzyme turnover numbers reveals protein structural correlates and improves metabolic models. *Nature Communications*, 9(1), 5252. <https://doi.org/10.1038/s41467-018-07652-6>
- Hohmann-Marriott, M. F., & Blankenship, R. E. (2011). Evolution of Photosynthesis. *Annual Review of Plant Biology*, 62(Volume 62, 2011), 515–548. <https://doi.org/10.1146/annurev-arplant-042110-103811>

- Holzhütter, H.-G. (2004). The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *European Journal of Biochemistry*, 271(14), 2905–2922. <https://doi.org/10.1111/j.1432-1033.2004.04213.x>
- Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., & Kummer, U. (2006). COPASI—a COmplex Pathway SIMulator. *Bioinformatics*, 22(24), 3067–3074. <https://doi.org/10.1093/bioinformatics/btl485>
- Höper, R., Komkova, D., Zavřel, T., & Steuer, R. (2024). A quantitative description of light-limited cyanobacterial growth using flux balance analysis. *PLOS Computational Biology*, 20(8), e1012280. <https://doi.org/10.1371/journal.pcbi.1012280>
- Hui, S., Silverman, J. M., Chen, S. S., Erickson, D. W., Basan, M., Wang, J., Hwa, T., & Williamson, J. R. (2015). Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Molecular Systems Biology*, 11(2), 784. <https://doi.org/10.15252/msb.20145697>
- Jahn, M., Vialas, V., Karlsen, J., Maddalo, G., Edfors, F., Forsström, B., Uhlén, M., Käll, L., & Hudson, E. P. (2018). Growth of Cyanobacteria Is Constrained by the Abundance of Light and Carbon Assimilation Proteins. *Cell Reports*, 25(2), 478–486.e8. <https://doi.org/10.1016/j.celrep.2018.09.040>
- Jain, R., Rivera, M. C., & Lake, J. A. (1999). Horizontal gene transfer among genomes: The complexity hypothesis. *Proceedings of the National Academy of Sciences*, 96(7), 3801–3806. <https://doi.org/10.1073/pnas.96.7.3801>
- Jeong, Y., Hong, S.-J., Cho, S.-H., Yoon, S., Lee, H., Choi, H.-K., Kim, D.-M., Lee, C.-G., Cho, S., & Cho, B.-K. (2021). Multi-Omic Analyses Reveal Habitat Adaptation of Marine Cyanobacterium *Synechocystis* sp. PCC 7338. *Frontiers in Microbiology*, 12. <https://doi.org/10.3389/fmicb.2021.667450>
- Jojima, T., Omumasaba, C. A., Inui, M., & Yukawa, H. (2010). Sugar transporters in efficient utilization of mixed sugar substrates: Current knowledge and outlook. *Applied Microbiology and Biotechnology*, 85(3), 471–480. <https://doi.org/10.1007/s00253-009-2292-1>
- Kaneko, T., Nakamura, Y., Sasamoto, S., Watanabe, A., Kohara, M., Matsumoto, M., Shimpo, S., Yamada, M., & Tabata, S. (2003). Structural analysis of four large plasmids harboring in a unicellular cyanobacterium, *Synechocystis* sp. PCC 6803. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, 10(5), 221–228. <https://doi.org/10.1093/dnares/10.5.221>
- Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A.,

- Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpō, S., Takeuchi, C., ... Tabata, S. (1996a). Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. Strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Research*, 3(3), 109–136. Scopus. <https://doi.org/10.1093/dnares/3.3.109>
- Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpō, S., Takeuchi, C., ... Tabata, S. (1996b). Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. Strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, 3(3), 109–136. <https://doi.org/10.1093/dnares/3.3.109>
- King, Z. A., Dräger, A., Ebrahim, A., Sonnenschein, N., Lewis, N. E., & Palsson, B. O. (2015). Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. *PLOS Computational Biology*, 11(8), e1004321. <https://doi.org/10.1371/journal.pcbi.1004321>
- Kitano, H. (2002). Systems biology: A brief overview. *Science*, 295(5560), 1662–1664. Scopus. <https://doi.org/10.1126/science.1069492>
- Knoop, H., Gründel, M., Zilliges, Y., Lehmann, R., Hoffmann, S., Lockau, W., & Steuer, R. (2013). Flux Balance Analysis of Cyanobacterial Metabolism: The Metabolic Network of *Synechocystis* sp. PCC 6803. *PLOS Computational Biology*, 9(6), e1003081. <https://doi.org/10.1371/journal.pcbi.1003081>
- Knoot, C. J., Ungerer, J., Wangikar, P. P., & Pakrasi, H. B. (2018). Cyanobacteria: Promising biocatalysts for sustainable chemical production. *Journal of Biological Chemistry*, 293(14), 5044–5052. <https://doi.org/10.1074/jbc.R117.815886>
- Kroll, A., Engqvist, M. K. M., Heckmann, D., & Lercher, M. J. (2021). Deep learning allows genome-scale prediction of Michaelis constants from structural features. *PLOS Biology*, 19(10), e3001402. <https://doi.org/10.1371/journal.pbio.3001402>
- Kroll, A., Rousset, Y., Hu, X.-P., Liebrand, N. A., & Lercher, M. J. (2023). Turnover number predictions for kinetically uncharacterized enzymes using machine and deep learning. *Nature Communications*, 14(1), 4139. <https://doi.org/10.1038/s41467-023-39840-4>
- Lenski, R. E. (2017). Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. *The ISME Journal*, 11(10), 2181–2194. <https://doi.org/10.1038/ismej.2017.69>

- Lerman, J. A., Hyduke, D. R., Latif, H., Portnoy, V. A., Lewis, N. E., Orth, J. D., Schrimpe-Rutledge, A. C., Smith, R. D., Adkins, J. N., Zengler, K., & Palsson, B. O. (2012). In silico method for modelling metabolism and gene product expression at genome scale. *Nature Communications*, 3(1), 929. <https://doi.org/10.1038/ncomms1928>
- Li, F., Yuan, L., Lu, H., Li, G., Chen, Y., Engqvist, M. K. M., Kerkhoven, E. J., & Nielsen, J. (2022). Deep learning-based kcat prediction enables improved enzyme-constrained model reconstruction. *Nature Catalysis*, 5(8), 662–672. <https://doi.org/10.1038/s41929-022-00798-z>
- Lozada-Chávez, I., Janga, S. C., & Collado-Vides, J. (2006). Bacterial regulatory networks are extremely flexible in evolution. *Nucleic Acids Research*, 34(12), 3434–3445. <https://doi.org/10.1093/nar/gkl423>
- Luan, G., & Lu, X. (2018). Tailoring cyanobacterial cell factory for improved industrial properties. *Biotechnology Advances*, 36(2), 430–442. <https://doi.org/10.1016/j.biotechadv.2018.01.005>
- Mao, Z., Yuan, Q., Li, H., Zhang, Y., Huang, Y., Yang, C., Wang, R., Yang, Y., Wu, Y., Yang, S., Liao, X., & Ma, H. (2023). CAVE: A cloud-based platform for analysis and visualization of metabolic pathways. *Nucleic Acids Research*, 51(W1), W70–W77. <https://doi.org/10.1093/nar/gkad360>
- Martyushenko, N., & Almaas, E. (2019). ModelExplorer—Software for visual inspection and inconsistency correction of genome-scale metabolic reconstructions. *BMC Bioinformatics*, 20(1), 56. <https://doi.org/10.1186/s12859-019-2615-x>
- Maslov, S., Krishna, S., Pang, T. Y., & Sneppen, K. (2009). Toolbox model of evolution of prokaryotic metabolic networks and their regulation. *Proceedings of the National Academy of Sciences*, 106(24), 9743–9748. <https://doi.org/10.1073/pnas.0903206106>
- Matthias, K., Klähn, S., Ingeborg, S., Jasper, K. F. M., Wolfgang, R. H., & Björn, V. (2014). Comparative analysis of the primary transcriptome of synechocystis sp. PCC 6803. *DNA Research*, 21(5), 527–539. <https://doi.org/10.1093/dnares/dsu018>
- Molenaar, D., Van Berlo, R., De Ridder, D., & Teusink, B. (2009). Shifts in growth strategies reflect tradeoffs in cellular economics. *Molecular Systems Biology*, 5(1), 323. <https://doi.org/10.1038/msb.2009.82>
- Montagud, A., Zelezniak, A., Navarro, E., de Córdoba, P. F., Urchueguía, J. F., & Patil, K. R. (2011). Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium *Synechocystis* sp. PCC6803. *Biotechnology Journal*, 6(3), 330–342. <https://doi.org/10.1002/biot.201000109>

- Mori, M., Hwa, T., Martin, O. C., Martino, A. D., & Marinari, E. (2016). Constrained Allocation Flux Balance Analysis. *PLOS Computational Biology*, 12(6), e1004913. <https://doi.org/10.1371/journal.pcbi.1004913>
- Mukherjee, B., Madhu, S., & Wangikar, P. P. (2020). The role of systems biology in developing non-model cyanobacteria as hosts for chemical production. *Current Opinion in Biotechnology*, 64, 62–69. <https://doi.org/10.1016/j.copbio.2019.10.003>
- Nogales, J., Gudmundsson, S., Knight, E. M., Palsson, B. O., & Thiele, I. (2012). Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. *Proceedings of the National Academy of Sciences*, 109(7), 2678–2683. <https://doi.org/10.1073/pnas.1117907109>
- O'Brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R., & Palsson, B. (2013). Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular Systems Biology*, 9(693). <https://doi.org/10.1038/msb.2013.52>
- Oren, A. (2015). Cyanobacteria in hypersaline environments: Biodiversity and physiological properties. *Biodiversity and Conservation*, 24(4), 781–798. <https://doi.org/10.1007/s10531-015-0882-z>
- Orr, H. A. (2005). The genetic theory of adaptation: A brief history. *Nature Reviews Genetics*, 6(2), 119–127. <https://doi.org/10.1038/nrg1523>
- Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010a). What is flux balance analysis? *Nature Biotechnology*, 28(3), 245–248. <https://doi.org/10.1038/nbt.1614>
- Orth, J. D., Thiele, I., & Palsson, B. O. (2010b). What is flux balance analysis? *Nature Biotechnology*, 28(3), 245–248. Scopus. <https://doi.org/10.1038/nbt.1614>
- Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010c). What is flux balance analysis? *Nature Biotechnology*, 28(3), 245–248. <https://doi.org/10.1038/nbt.1614>
- Reimers, A. M., Knoop, H., Bockmayr, A., & Steuer, R. (2017). Cellular trade-offs and optimal resource allocation during cyanobacterial diurnal growth. *Proceedings of the National Academy of Sciences of the United States of America*, 114(31), E6457–E6465. <https://doi.org/10.1073/pnas.1617508114>
- Rybak, A. S., Dziuba, M., Pelechata, A., Rybak, M., Akter, S., Czerepska, A., Dulić, T., Gąbka, M., Hindáková, A., Jurczak, T., Kendir, A., Mankiewicz-Boczek, J., Meriluoto, J., & Wejnerowski, Ł. (2024). Characterization of cyanobacterial mats from an artificial hot spring in Uniejów (Poland) and the potential use of their biomass. *Algal Research*, 82, 103646. <https://doi.org/10.1016/j.algal.2024.103646>
- Sahu, A., Blätke, M.-A., Szymański, J. J., & Töpfer, N. (2021). Advances in flux balance analysis by integrating machine learning and mechanism-based models. *Computational*

- and Structural Biotechnology Journal*, 19, 4626–4640.
<https://doi.org/10.1016/j.csbj.2021.08.004>
- Sánchez, B. J., Zhang, C., Nilsson, A., Lahtvee, P., Kerkhoven, E. J., & Nielsen, J. (2017). Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Molecular Systems Biology*, 13(8), 935.
<https://doi.org/10.15252/msb.20167411>
- Santos-Merino, M., Yun, L., & Ducat, D. C. (2023). Cyanobacteria as cell factories for the photosynthetic production of sucrose. *Frontiers in Microbiology*, 14(February).
<https://doi.org/10.3389/fmicb.2023.1126032>
- Schellenberger, J., Park, J. O., Conrad, T. M., & Palsson, B. Ø. (2010). BiGG: A Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, 11(1), 213. <https://doi.org/10.1186/1471-2105-11-213>
- Schuster, S., Pfeiffer, T., & Fell, D. A. (2008). Is maximization of molar yield in metabolic networks favoured by evolution? *Journal of Theoretical Biology*, 252(3), 497–504.
<https://doi.org/10.1016/j.jtbi.2007.12.008>
- Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z., & Hwa, T. (2010). Interdependence of Cell Growth and Gene Expression: Origins and Consequences. *Science*, 330(6007), 1099–1102. <https://doi.org/10.1126/science.1192588>
- Shih, P. M., Wu, D., Latifi, A., Axen, S. D., Fewer, D. P., Talla, E., Calteau, A., Cai, F., Tandeau de Marsac, N., Rippka, R., Herdman, M., Sivonen, K., Coursin, T., Laurent, T., Goodwin, L., Nolan, M., Davenport, K. W., Han, C. S., Rubin, E. M., ... Kerfeld, C. A. (2013). Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proceedings of the National Academy of Sciences*, 110(3), 1053–1058.
<https://doi.org/10.1073/pnas.1217107110>
- Singh, J. S., Kumar, A., Rai, A. N., & Singh, D. P. (2016). Cyanobacteria: A Precious Bio-resource in Agriculture, Ecosystem, and Environmental Sustainability. *Frontiers in Microbiology*, 7. <https://doi.org/10.3389/fmicb.2016.00529>
- Stanier, R. Y., Kunisawa, R., Mandel, M., & Cohen-Bazire, G. (1971). Purification and properties of unicellular blue-green algae (order Chroococcales). *Bacteriological Reviews*, 35(2), 171–205.
- Szappanos, B., Fritzemeier, J., Csörgő, B., Lázár, V., Lu, X., Fekete, G., Bálint, B., Herczeg, R., Nagy, I., Notebaart, R. A., Lercher, M. J., Pál, C., & Papp, B. (2016). Adaptive evolution of complex innovations through stepwise metabolic niche expansion. *Nature Communications*, 7(1), 11607. <https://doi.org/10.1038/ncomms11607>

- Thiele, S., Von Kamp, A., Bekiaris, P. S., Schneider, P., & Klamt, S. (2022). CNAPy: A CellNetAnalyzer GUI in Python for analyzing and designing metabolic networks. *Bioinformatics*, 38(5), 1467–1469. <https://doi.org/10.1093/bioinformatics/btab828>
- Töpfer, N., Braam, T., Shameer, S., Ratcliffe, R. G., & Sweetlove, L. J. (2020). Alternative crassulacean acid metabolism modes provide environment-specific water-saving benefits in a leaf metabolic model. *Plant Cell*, 32(12), 3689–3705. Scopus. <https://doi.org/10.1105/tpc.20.00132>
- Vazquez, A., Beg, Q. K., deMenezes, M. A., Ernst, J., Bar-Joseph, Z., Barabási, A.-L., Boros, L. G., & Oltvai, Z. N. (2008). Impact of the solvent capacity constraint on *E. coli* metabolism. *BMC Systems Biology*, 2(1), 7. <https://doi.org/10.1186/1752-0509-2-7>
- Wang, L., York, S. W., Ingram, L. O., & Shanmugam, K. T. (2019). Simultaneous fermentation of biomass-derived sugars to ethanol by a co-culture of an engineered *Escherichia coli* and *Saccharomyces cerevisiae*. *Bioresource Technology*, 273, 269–276. <https://doi.org/10.1016/j.biortech.2018.11.016>
- Weiß, A. Y., Oyarzún, D. A., Danos, V., & Swain, P. S. (2015). Mechanistic links between cellular trade-offs, gene expression, and growth. *Proceedings of the National Academy of Sciences*, 112(9). <https://doi.org/10.1073/pnas.1416533112>
- Wejnerowski, Ł., Poniecka, E., Buda, J., Klimaszyk, P., Piasecka, A., Dziuba, M. K., Mugnai, G., Takeuchi, N., & Zawierucha, K. (2023). Empirical testing of cryoconite granulation: Role of cyanobacteria in the formation of key biogenic structure darkening glaciers in polar regions. *Journal of Phycology*, 59(5), 939–949. <https://doi.org/10.1111/jpy.13372>
- Wolf, Y. I., & Koonin, E. V. (2013). Genome reduction as the dominant mode of evolution. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 35(9), 829–837. <https://doi.org/10.1002/bies.201300037>
- Wortel, M. T., Noor, E., Ferris, M., Bruggeman, F. J., & Liebermeister, W. (2018). Metabolic enzyme cost explains variable trade-offs between microbial growth rate and yield. *PLoS Computational Biology*, 14(2), 1–21. <https://doi.org/10.1371/journal.pcbi.1006010>
- Zavřel, T., Faizi, M., Loureiro, C., Poschmann, G., Stühler, K., Sinetova, M., Zorina, A., Steuer, R., & Červený, J. (2019). Quantitative insights into the cyanobacterial cell economy. *eLife*, 8, 1–29. <https://doi.org/10.7554/eLife.42508>
- Zavřel, T., Očenášová, P., & Červený, J. (2017). Phenotypic characterization of *Synechocystis* sp. PCC 6803 substrains reveals differences in sensitivity to abiotic stress. *PLOS ONE*, 12(12), e0189130. <https://doi.org/10.1371/journal.pone.0189130>
- Zavřel, T., Sinetova, M. A., Búzová, D., Literáková, P., & Červený, J. (2015). Characterization of a model cyanobacterium *Synechocystis* sp. PCC 6803 autotrophic growth in a flat-

panel photobioreactor. *Engineering in Life Sciences*, 15(1), 122–132.
<https://doi.org/10.1002/elsc.201300165>