

Modeling Complex Cellular Systems - from Differential Equations to Constraint-Based Models

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)



seit 1558

vorgelegt dem Rat der Biologisch-Pharmazeutischen Fakultät
der FRIEDRICH-SCHILLER-UNIVERSITÄT JENA

eingereicht von Dipl. Bioinf. Sascha Schäuble
geb. am 16.10.1982 in Frankfurt (Oder)

Examiners/Gutachter:

- Jun.-Prof. Christoph Kaleta, Friedrich-Schiller-Universität, Jena
- Prof. Ines Heiland, UiT The Arctic University of Norway, Tromsø
- Dr.-Ing. Steffen Klamt, Max-Planck-Institut, Magdeburg

Day of public defense/Tag der öffentlichen Verteidigung: 13.02.2014

Abstract

In the beginning of the 20th century, scientists realized the necessity of purifying enzymes to unravel their mechanistic nature. A century and tremendous progresses in the natural sciences later, molecular and particularly systems biology became fundamental pillars of modern biology. Next to descriptive studies, it enables the quantitative description of diverse biological processes. Along the advances in the field, natural scientists developed an increasing interest in theoretical models comprising the dynamics of biological systems. They allow not only a deeper understanding, but also a prediction of the behavior of the investigated process or species. In the first part of my thesis, I present my contribution to the field of studying the dynamics of biological phenomena. I present fundamental issues arising, when neglecting substrate inhibition in kinetic modeling. Furthermore, I describe a model that considers experimental data to simulate the transition of normal proliferating into cellular senescent cells. Cellular senescence is an irreversible cell cycle arrested state, which is believed to prevent uncontrolled cell growth.

In recent years, the size and complexity of published metabolic networks increased notably, however, given the advent of massive data generating high-throughput sequencing methods and the successful annotation of complete genomes. Since these large-scaled models are more comprehensive, they commonly prohibit a mechanistic modeling approach, due to missing kinetic data or computational infeasibility. In order to analyze such models, nevertheless, constraint-based methods proved to be suitable tools. In general, these methods are based on known physiological restrictions, stoichiometric information and network structure and do not require the specification of single kinetic parameter values. In the second part of my thesis, I contribute three studies to constraint-based modeling. I describe the established concept of elementary flux modes, which resemble non-decomposable and theoretically feasible pathways of metabolic networks. I demonstrate the procedure at a small example network of photosynthate metabolism. Subsequently, I present the analysis of the nitrogen metabolism network of *Chlamydomonas reinhardtii* with respect to circadian regulation, which gives rise to about three million elementary flux modes. Despite this large, yet still computable set, obtaining all elementary flux modes of larger, up to genome-scale metabolic models is commonly not feasible.

Flux balance analysis, a linear programming based method, circumvents this problem, by calculating single optimized fluxes. In the last study of this thesis, I present a comprehensive work on metabolic costs of amino acid and protein production in *Escherichia coli*. These costs were both manually calculated as well as based on a flux balance analysis of an *E. coli* genome-scale metabolic model.

I conclude with discussing the conducted studies in respect of the chosen analysis methods and the respective current knowledge base. Both approaches, either dynamic or constraint-based modeling, proved to be suitable strategies to describe biological processes at different levels. Whereas dynamic modeling allowed for a precise description of the temporal behavior of biological species, constraint-based modeling enabled studies, where the complexity of the investigated phenomena prohibits kinetic modeling. Finally, I give an outlook to further modeling possibilities, which were not considered in this thesis.

Zusammenfassung

Zu Beginn des 20. Jahrhunderts erkannten Wissenschaftler die Notwendigkeit, Enzyme aufzureinigen zu müssen, um ihrer mechanistischen Natur auf den Grund zu gehen. Ein Jahrhundert mit bedeutenden Fortschritten in den Biowissenschaften später, entwickelte sich sowohl die Molekular- als auch die Systembiologie zu bedeutenden Grundsäulen der modernen Biologie. Diese erlaubt neben rein deskriptiven Studien auch quantitative Beschreibungen von biologischen Prozessen. Darüber hinaus entwickelten die Wissenschaftler ein gesteigertes Interesse an theoretischen Modellen, welche die Dynamiken von biologischen Systemen beschreiben können. Solche Modelle erlauben nicht nur ein tieferes Verständnis, sondern auch eine präzisere Vorhersage des Verhaltens der untersuchten Prozesse oder Spezies. Im ersten Teil meiner Dissertation präsentiere ich meinen Beitrag zu Dynamikstudien von biologischen Phänomenen. Ich stelle eine Arbeit vor, die das Vernachlässigen von Substratkompetition in kinetischen Modellen und dessen Auswirkungen darlegt. Desweiteren beleuchte ich unter Berücksichtigung von experimentellen Daten die Überführung von proliferierenden zu seneszenten Zellen. Zelluläre Seneszenz ist ein unumkehrbarer zellzyklusarretierter Zustand und daher ein potenzieller Mechanismus zum Schutz gegen unkontrolliertes Zellwachstum.

Jedoch führten in jüngerer Zeit moderne Verfahren der Hochdurchsatzsequenzierung und die steigende Anzahl von annotierten Genomen zu einer erhöhten Komplexität publizierter metabolischer Netzwerke. Der bedeutend größere Umfang dieser Modelle verhindert in der Regel einen mechanistischen Modellierungsansatz, da kinetische Daten meist nicht komplett verfügbar sind oder der Rechenaufwand zur Simulation zu hoch ist. Um diese Modelle trotzdem vernünftig analysieren zu können, hat sich die Anwendung von auf Einschränkungen basierenden Methoden bewährt. Diese Methoden beruhen auf bekannten physiologischen Beschränkungen, stöchiometrischer Information und der Netzwerkstruktur an sich. Dabei benötigen sie keine Information über kinetische Parameter. Im zweiten Teil meiner Dissertation, stelle ich drei Arbeiten vor, die diesem Methodenbereich zuzuordnen sind. Ich beschreibe das etablierte Konzept der elementaren Flussmoden, welche nicht zerlegbare und theoretisch mögliche Flüsse in metabolischen Netzwerken repräsentieren. Ich demonstriere die Nützlichkeit der Methode anhand eines kleinen Beispielnetzwerkes

des photosynthetischen Metabolismus. Anschließend stelle ich eine Arbeit vor, die sich mit dem Stickstoffmetabolismus der Grünalge *Chlamydomonas reinhardtii* unter Beachtung von zirkadianer Regulation befasst. Diese Arbeit resultierte in der Berechnung und Analyse von ca. drei Millionen elementaren Flussmoden. Nichtsdestotrotz ist die Berechnung aller elementarer Flussmoden in einem noch größeren bzw. genomskaligen metabolischen Modell in der Regel nicht möglich. Die Flussbilanzanalyse, ein weiteres auf Einschränkungen basierendes Verfahren, umgeht dieses Problem, indem sie einzelne optimierte Flüsse berechnet. In der letzten Studie dieser Dissertation stelle ich eine umfassende Arbeit vor, die die metabolischen Kosten von Aminosäure- und Proteinproduktion in *Escherichia coli* darlegt. Neben einer manuellen Berechnung, basieren die ermittelten Kosten auf einer Flussbilanzanalyse des genomskaligen Modells von *E. coli*.

Ich schließe die Arbeit mit einer Diskussion der vorgestellten Studien im Kontext der gewählten Analysemethoden und des jeweiligen aktuellen Wissensstandes ab. Beide Ansätze, sowohl die dynamische Modellierung, als auch die Nutzung von auf Einschränkungen basierenden Methoden, erwiesen sich als günstige Strategien, um biologische Prozesse auf unterschiedlichen Ebenen beschreiben zu können. Während die dynamische Modellierung eine präzise Beschreibung von temporalem Verhalten biologischer Spezies erlaubte, ermöglichten auf Einschränkungen beruhende Methoden Studien, deren Komplexität eine kinetische Modellierung verhindern. Zuletzt gebe ich einen Ausblick auf weitere Modellierungsmöglichkeiten, welche im Rahmen dieser Arbeit keine Verwendung fanden.

Danksagung

Die Ergebnisse einer Dissertation sind selten auf eine einzige Person zurückzuführen. Daher möchte ich zuallererst allen Personen danken, die direkt oder indirekt zu den Resultaten der vorliegenden Arbeit beigetragen haben. Besonders danken möchte ich auch dem gesamten Lehrstuhl der Bioinformatik für eine jederzeit angenehme Arbeitsatmosphäre. Hervorheben möchte ich dabei meine Betreuer Prof. Stefan Schuster und Jun.-Prof. Christoph Kaleta, die mir nicht nur lediglich die Möglichkeit zur Promotion gegeben haben. Ferner unterstützten sie mich in jeglicher Hinsicht bei der Durchführung bis hin zur Veröffentlichung von wissenschaftlichen Projekten. Dies ist nicht selbstverständlich, wofür ich sehr dankbar bin. Ein besonderer Dank gilt auch Prof. Ines Heiland, von deren biochemischem Fachwissen ich u. a. in vielen fruchtbaren Diskussionen profitieren konnte.

Ich bedanke mich auch bei allen Freunden, die mir gerade auch in anstrengenden Zeiten zur Seite standen und mich daran erinnerten, dass es auch ein Leben neben einer Dissertation gibt. Zuletzt möchte ich meiner Familie und meiner Freundin Lisa danken, die einen ganz besonderen Anteil an meiner Arbeit haben. Sie gaben und geben mir die Kraft und die Sicherheit, die Ziele erreichen zu können, die ich mir setze.

Contents

1. Introduction	1
1.1. Dynamic modeling	4
1.2. Constraint-based methods	8
1.2.1. Elementary Flux Modes	10
1.2.2. Flux Balance Analysis	14
1.3. Biological background	16
1.3.1. Cellular senescence, the irreversible cell cycle arrest	17
1.3.2. The green algae <i>Chlamydomonas reinhardtii</i> as model organism for nitrogen assimilation	19
1.3.3. The microorganism <i>Escherichia coli</i> in a biotechnological context	20
2. Dynamic Modelling	22
Effect of substrate competition in kinetic models of metabolic networks. Schäuble S, Stavrum A K, Puntervoll P, Schuster S, Heiland I, <i>FEBS Lett</i> , 587(17): 2818–2824, 2013	23
Quantitative Model of Cell Cycle Arrest and Cellular Senescence in Primary Human Fibroblasts. Schäuble S, Klement K, Marthandan S, Munch S, Heiland I, Schuster S, Hemmerich P, Diekmann S, <i>PLoS ONE</i> , 7(8): e42150, 2012	30
3. Constraint-based modelling	44
Hands-on metabolism: analysis of complex biochemical networks using elementary flux modes. Schäuble S, Schuster S, Kaleta C, <i>Methods Enzymol</i> 500, 437–456, 2011	45
Predicting the Physiological Role of Circadian Metabolic Regulation in the Green Alga <i>Chlamydomonas reinhardtii</i> . Schäuble S, Heiland I, Voytsekh O, Mittag M, Schuster S, <i>PLoS ONE</i> 6(8): e23026, 2011 . .	65

Metabolic costs of amino acid and protein production in <i>Escherichia coli</i> . Kaleta C, Schäuble S, Rinas U, Schuster S, <i>Biotechnol J</i> , 2013, 8(9): 1105–1114, 2013	77
4. Discussion	87
4.1. ODE based modeling to study substrate competition and cellular senescence	87
4.1.1. Effect of substrate competition potentially influences simula- tion results	88
4.1.2. Modeling cellular senescence allows for a quantitative descrip- tion of growth curves of human fibroblasts and evaluation of cell state indicating biomarkers	91
4.2. Constraint-based methods enable the study of large-scale models . . .	94
4.2.1. Elementary Flux Modes provide a large application range and propose infeasibility of sucrose and starch production in plants at night	95
4.2.2. EFM analysis predicts a shut down of less effective nitrogen assimilation pathways in <i>C. reinhardtii</i> via the circadian con- trolled factor CHLAMY1 at night	98
4.2.3. Computing costs of metabolic compounds with FBA guides the search for suitable, efficient growth media for <i>E. coli</i>	102
5. Conclusion	105
Bibliography	107
A. Supplementary material	130
Effect of substrate competition in kinetic models of metabolic networks. Schäuble S, Stavrum A K, Puntervoll P, Schuster S, Heiland I, <i>FEBS Lett</i> , 587(17): 2818[Pleaseinsertintopreamble]2824, 2013.	130
Quantitative Model of Cell Cycle Arrest and Cellular Senescence in Pri- mary Human Fibroblasts. Schäuble S, Klement K, Marthandan S, Munch S, Heiland I, Schuster S, Hemmerich P, Diekmann S, <i>PLoS ONE</i> , 7(8): e42150, 2012.	147

Predicting the Physiological Role of Circadian Metabolic Regulation in the Green Alga <i>Chlamydomonas reinhardtii</i> . Schäuble S, Heiland I, Voytsekh O, Mittag M, Schuster S, <i>PLoS ONE</i> , 6(8): e23026, 2011a. . 155	
Metabolic costs of amino acid and protein production in <i>Escherichia coli</i> . Kaleta C, Schäuble S, Rinas U, Schuster S. <i>Biotechnol J</i> , 8(9): 1105– 1114, 2013. 167	
B. Beitragende Autoren	177
C. Selbständigkeitserklärung	178

1. Introduction

Truth is ever to be found in simplicity, and not in the multiplicity and confusion of things.

(Sir Isaac Newton)

Although nowadays the study of dynamics is essential and most promising in nearly all aspects of natural sciences, in the beginning it was a habitat for physicists only. In his most influential work *Philosophiæ Naturalis Principia Mathematica*, Isaac Newton presented his laws of motion in 1687 and applied them to the study of planet movements. Together with the concept of differential equations, scientists were soon able to study nature's temporal behavior in many facets.

Until the beginning of the 20th century, however, complex biochemical processes were extremely difficult to be realistically mapped onto mathematical models. Consequently, scientists began to purify enzymes to mathematically clarify their catalytic activity. In 1902, Victor Henri mathematically formulated the influence of substrate concentration on an intermediary enzyme-substrate complex in a catalytic reaction (Henri 1902, 1903). In 1913, Leonor Michaelis and Maud Menten rediscovered and elaborated on Henri's equation (Michaelis and Menten 1913), which is why its today's correct reference should be Henri-Michaelis-Menten equation. Together with the steady state modification, introduced by George Briggs and John Haldane (1925), virtually all kinetic velocity studies were based on these two enzyme descriptions until the 1950's. In 1956, Edward King and Carl Altman introduced a schematic method to derive a rate law under the steady state assumption, regardless of the number of enzyme-containing intermediate complexes (King and Altman 1956), allowing for the easy deduction of bi- or tri-reactant kinetics. Finally, writing sophisticated kinetic rate equations became conveniently easy with the proposed nomenclature by William Cleland (Cleland 1963).

A decade before, Watson and Crick discovered the structure of DNA (Watson and

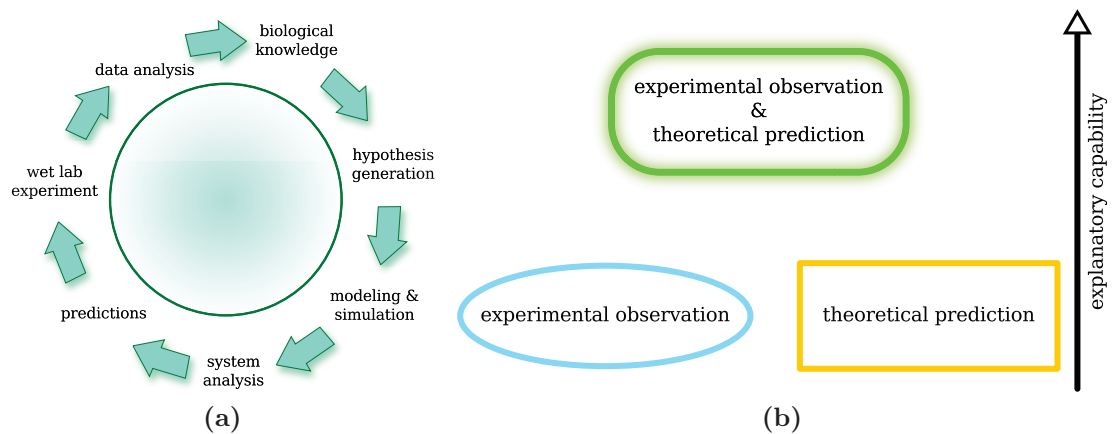


Figure 1.1.: Scheme of the explanatory capacity. By iteratively refining and testing model parameters (Fig. 1.1a, adapted from Kitano (2002)), experimental and theoretical efforts can be combined resulting in a model, which potentially allows for an increasing comprehension of the studied phenomena (Fig. 1.1b).

Crick 1953). This groundbreaking discovery, the increasing interest in metabolic fluxes and by a great extent the development of metabolic control theory (allowing for the determination of flux control, Kacser and Burns (1973); Heinrich and Rapoport (1974)) eventually led to the rise of systems biology as a new field in the life sciences.

In general, systems biology endeavors to combine discovery and hypothesis-driven science (Ideker et al. 2001; Kitano 2002). Whereas the former led to among others, the complete sequencing of the human genome (Lander et al. 2001; Consortium 2004), the latter is dedicated to create testable hypotheses (Ideker et al. 2001). Its overall aim centers around the intricate connection of experimental efforts and computational methods leading to an iterative process of experiment and model reconciliation (cf. Fig. 1.1a). At the end of such an endeavor, the studied phenomena can be comprehended by both, experimental observations and theoretical predictions. In general, this leads to a more comprehensive picture than if studied only by experiment or modeling (see also scheme in Fig. 1.1b).

Along the rise of sophisticated metabolic models up to genome-scale, however, determining realistic and tissue specific accompanying kinetic parameters became even more difficult than before. This is primarily due to a common lack of comprehensive kinetic data, especially in the context of specific tissues. Together with the advent of next-generation high-throughput sequencing techniques (Margulies et al.

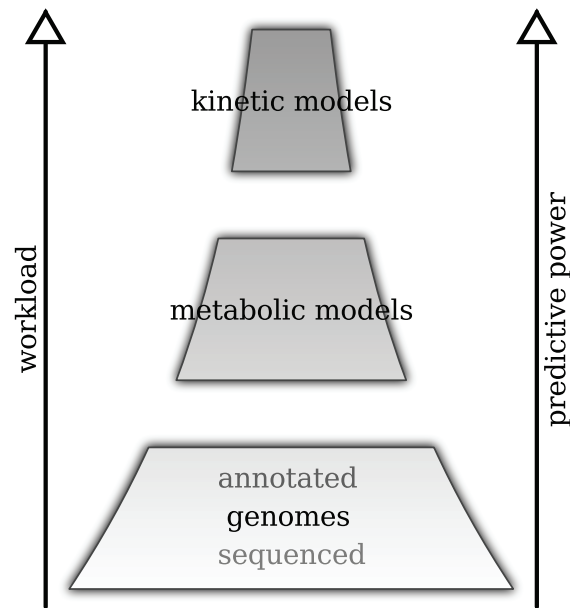


Figure 1.2.: Modell availability scheme. The narrowing width of subsequent trapezoids indicates the bottleneck in deriving models of increasing predictive power based on the layers below.

2005; Soon et al. 2013), the disbalance between data and model availability increased even more. Naturally, shortly after the first reconstructed metabolic model had been analyzed and published (Edwards and Palsson 1999), the number of further published models steadily increased. This led to metabolic reconstructions of more than sixty organisms (Kim et al. 2012), including genome-scale models of *Escherichia coli* (Feist et al. 2007) and *Homo sapiens* (Duarte et al. 2007; Thiele et al. 2013), among others. The speed of upcoming raw data from next-generation sequencing projects (Weckwerth 2011; Scholz et al. 2012) surpassed the model generation rate, nevertheless, due to its highly automated processing. Thus, the bottleneck of scientific discovery gradually shifted from data generation to model derivation and evaluation. In fact, this led to a whole cascade of bottlenecks ranging from the generation of genome data to annotated genomes towards realistic kinetic modeling (see scheme at Fig. 1.2).

Consequently, to assess the quality of genome-scaled models, tools are required, which are appropriate to cope with large-scale metabolic networks. These tools or methods should potentially deliver valuable solutions in affordable time given simplifying, yet reasonable assumptions. Among these methods elementary flux mode analysis and flux balance analysis emerged as most promising tools. In short, they

are based on the steady state assumption and tackle questions arising in connection with metabolic networks, such as: What is the theoretical maximum yield of a certain metabolite based on the availability of another metabolite? or What is the optimal energy or biomass production rate with respect to certain physiological constraints? Thus, these methods are qualified to answer some of the questions regarding the dynamics of a system, particularly concerning its steady state phenotype.

In summary, the field of systems biology provides more raw data at decreasing costs than ever before and challenges bioinformatics to select the most promising analysis or model construction tools for a given study project. In fact, metabolic models up to genome-scale require efficient analyses and prohibit a comprehensive kinetic simulation due to a lack of data and computational power. Specific biological processes can be studied very well with kinetic modeling, nevertheless, if the problem is moderately scaled and public databases allow for the deduction of kinetic parameters.

In my thesis, I present several techniques and tools, in order to study complex biological phenomena. I will start with introducing the concept of dynamic modeling via differential equations. Next, I will explain the concepts of elementary flux modes and flux balance analysis that were used in order to investigate models, which are inaccessible to classic kinetic simulations.

1.1. Dynamic modeling

A promising method to approach the dynamics of biological processes or mechanisms is the application of differential equations. Both, Leibniz and Newton independently developed the fundamental principles of differential equations in the second half of the 17th century¹. They are since then of widespread use in the study of the dynamics of various observations ranging from the natural sciences to technology. A differential equation is an equation of a function of one or more variables including not only the function itself, but also its derivatives of any orders. Differential equations can be furthermore categorized into ordinary and partial differential equations. Whereas the former involves only derivatives of one independent variable, the latter contains more than one independent variable. The conducted studies presented in chapter 2

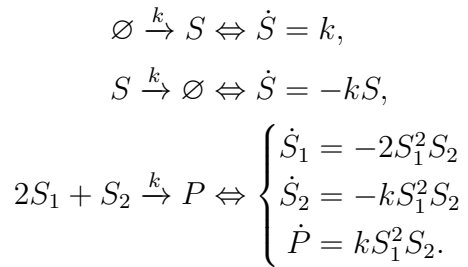
¹In fact, Newton mistrusted Leibniz and accused him of having replicated his ideas, which led to one of the most famous controversies in science history (cf. Hall (1980)).

comprise investigations of temporal behaviors and therefore could be modeled by ordinary differential equations (ODEs). Next, ODEs with time t as the independent variable will be introduced in more detail. Let $\dot{x}_i \equiv dx_i/dt$, ($i = 1, \dots, n$), then a system of ordinary differential equations can be described by:

$$\begin{aligned} \dot{x}_1 &= f_1(x_1, \dots, x_n) \\ &\vdots \\ \dot{x}_n &= f_n(x_1, \dots, x_n). \end{aligned} \tag{1}$$

In biology, the variables x_i typically reflect the concentrations of molecular compounds or of populations in a given (eco-)system. The functions f_i are determined in the context of the studied problem. An ODE system is said to be linear, if the right hand side of \dot{x}_i contains only x_i to the first power, otherwise nonlinear. In general, nonlinear systems are more complex and usually difficult to solve analytically. The functions f of the ODE system (1) are termed autonomous, since they are not explicitly dependent on time as e.g. given by $\dot{x} = f(x, t)$. Such systems are usually more complicated to solve, since more information is required (x and t) to predict the systems behavior in the future. To circumvent this issue, typically, an additional equation is added to an n -dimensional given system, whereby the time-dependency is described by setting $x_{n+1} = t$ and $\dot{x}_{n+1} = 1$. Thus, any time-dependent or nonautonomous n -dimensional system can be rewritten as an $(n+1)$ -dimensional autonomous system.

Since in molecular biology, most systems commonly contain numerous reaction species, these reactions can be described by the deterministic form of ODEs. If the number of involved species of the reactant system is sufficiently high, mass action kinetics provide a satisfying approximation of the real reaction processes. Based on mass-action kinetics, elementary chemical reactions can be represented by an ODE (cf. Heinrich and Schuster (1996); Strogatz (2000)), as e.g.:



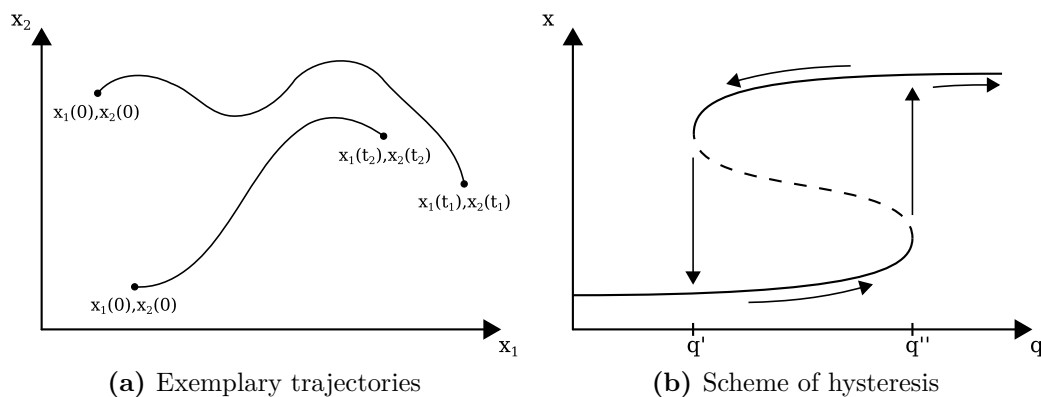


Figure 1.3.: Phase plane and hysteresis concept. (a) Depicted are exemplary trajectory courses starting at two different initial value pairs $(x_1(0), x_2(0))$. The curve or trajectory starting at given initial values represents the function solutions $x_1(t)$ and $x_2(t)$ in the n -dimensional solution space (here, $n=2$) or phase plane. (b) Scheme of hysteresis curve in parameter space. By increasing q the lower stable fixed point vanishes beyond q'' and the system variables “jump” into the remaining stable fixed point (upper branch). Once, in the large-amplitude branch, the system stays at the fixed point, even if q is reduced again, but not beyond q' . This “robustness” against parameter variation is termed hysteresis.

Although simple, such rate equations can be challenging to be analyzed analytically. The situation is even more problematic with nonlinear rate equations, which can be derived based on the assumption of elementary mass action reactions. Typical representatives of nonlinear rate equations are Henry-Michaelis-Menten kinetics, Hill kinetics or kinetics describing bi-molecular reaction mechanisms. Thus, an analytical analysis is often infeasible, which is why such systems are commonly solved by numerical algorithms allowing e. g. for time-course studies. Examples of solving algorithms may be given by, but are not limited to, the Runge-Kutta or Hooke & Jeeves procedures (Runge 1895; Kutta 1901; Hooke and Jeeves 1961).

Notably, even without actually solving the system, it is still possible to derive meaningful properties by e. g. geometric inspecting the solution of the system given by trajectories in the phase space. A trajectory describes the progress of the system or model variables x_1, \dots, x_n in the n -dimensional phase space (see also Fig. 1.3a).

Furthermore, either by a graphical analysis of the vector field (spanned e. g. by \dot{x} over x) or by an analytical analysis the properties of ODE systems can be further characterized. At points where $\dot{x} = 0$, there is no change in concentration, population number or any other described quantity. These points are called fixed

points and are distinguishable as unstable (repellers), stable (attractors) or half-stable fixed points. The latter is a mixed form of the first two, as it is repelling system quantities from one and attracting from the opposite direction. A fourth possibility is given by neutrally stable fixed points, where perturbations are neither declining nor accumulating (a nearby trajectory is neither attracted nor repelled). This fixed point type is typical for mechanical systems in the absence of friction, such as in harmonic oscillators.

Whenever the stability of a fixed point is changed due to a change in parameter values, a bifurcation occurs, wherefore such points are also called bifurcation points. Bifurcations occur in various forms, e.g. saddle-node or transcritical bifurcations. A saddle-node bifurcation is characterized as a parameter variation that lets two fixed points move towards, collide and destroy each other. In other situations, however, a fixed point must exist for all possible parameter values, but may change its stability as parameters are varied. Such a mechanism is typically represented by transcritical bifurcations. Going into more detail of bifurcation types is beyond the scope of this thesis, but might be obtained by studying e.g. Strogatz (2000) or Jetschke (2009). One important property in bifurcation analysis arises when multiple stable states exist. In such cases, the variation of parameters can invoke jumps or hysteresis of the modeled quantities as exemplified in Fig. 1.3b. In general, hysteresis characterizes a system, whose current stable state is not only dependent on its current parameter values, but also from the past states of the system. Upon a sufficient change in parameter values the current stable state vanishes, causing the system to rapidly change into another possible stable state. Notably, the system cannot immediately turn back by solely inverting the source of that change. Instead, in order to switch back to the original state the parameters need to be reversely changed beyond the range, where multiple stable states exist, if possible. Thus, the system “remembers” its past history. The study Schäuble et al. (2012), presented and described in more detail in chapter 2, makes use of this behavior in order to model different cell fates upon different amount of stress. Given high stress doses, our model switches from proliferation and recovery of healthy or mildly damaged cells to transforming cells into cellular senescence. Among other things, we show that the hysteresis property is critical to our model in order to explain the sudden change of cell fates upon altered stress amounts.

1.2. Constraint-based methods

When studying dynamic systems in biology and particularly metabolic pathways, two main problems arise regularly: The intrinsic parameters of a model are often not well determined or the simulation of a model may be computationally infeasible. The reasons for the first issue are manifold and arise from the context of the model itself. In genome-scale models it is often simply not possible to derive all kinetic parameters associated to the model reactions, since it requires a vast amount of experimental lab work and time, which is commonly not available. In other cases, kinetic parameters may not be identifiable. Here, existing experimental tools deliver measurements of limited resolution or identify parameters values, which differ notably from an *in vivo* situation (Teusink et al. 2000). Further sophisticated biological mechanisms might require even more advanced sensitive measurements. An example is given by the attempt to measure the concentration ratio of the redox pair of nicotinamide adenine dinucleotide, NAD^+ and NADH : In a basic environment, NAD^+ is extremely unstable, whereas NADH is unstable in an acidic milieu. Thus, both metabolites cannot be extracted in parallel, which makes it difficult to determine the exact ratio of these molecules by extraction methods alone. Beside these measurement issues, a model structure itself might simply not allow for identifying a parameter. A more complete and detailed discussion on identifiability analysis and its potential is provided by Raue et al. (2010).

The second problem arises primarily from the emergence of high-throughput sequence analysis tools and the maturation of early models into comprehensive up to genome-scale metabolic models. To analyze these models, fitting or simulating hundreds of species with an even higher number of variable parameters is required. In consequence, the computational demand for parameter estimation increases exponentially with model size and thus, becomes computationally infeasible.

Both problems, lack of knowledge and computational infeasibility, asked for the development or application of alternative approaches, which circumvent these obstacles. A common possibility is to assume reasonable simplifications that allow for an efficient computation, but still provide results. In the 1980's Seressiotis and Bailey pioneered by proposing an algorithm capable of computing a metabolic pathway, based on the concept of artificial intelligence (Seressiotis and Bailey 1986). Later, Mavrovouniotis et al. (1990) presented an improved method that explored all the possibilities of a given network. Both algorithms had in common that the computed

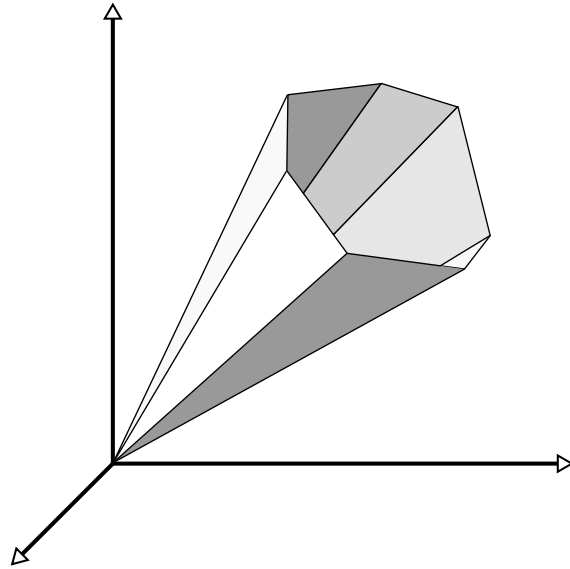


Figure 1.4.: Scheme of feasible flux cone for FBA and EFM analysis. The cone is spanned by the stoichiometric balance of intermediate metabolites as well as thermodynamic admissible reaction directions. The solution of an FBA optimization is restricted to be one of the corners of the flux cone, which corresponds to the optimized function value. In contrast, EFM analysis allows every valid linear combination of fluxes within the convex cone.

solution is in agreement to the stoichiometric constraints of the studied reaction system, although not adequately supported by an underlying theory. The study of network invariants (Reder 1988) and extreme currents (Clarke 1988; Schuster and Schuster 1993) led soon to the concept of extreme pathways and T-invariants, among others. Finally, the concepts of **elementary flux modes** (EFMs, Schuster (1994)) and **flux balance analysis** (FBA, Varma and Palsson (1994)) emerged as approaches, which are categorized as constraint-based methods. Both methods have fundamental assumptions in common: All metabolite species contained in the network model are assumed to operate at steady state. Moreover, irreversible reactions can only be used in the thermodynamical feasible direction. More formally, these assumptions read:

$$SV = 0, \tag{2}$$

$$V^{irr} \geq 0 \tag{3}$$

with S being the stoichiometric matrix representation of all metabolites (rows) taking part in the model comprised reactions (columns). The vector $V = (V^{rev} \ V^{irr})^T$

is the sorted vector of reversible and irreversible feasible fluxes through the associated reactions in S . A nonzero coefficient of S at entry (i, j) resembles either consumption (negative coefficient) or production (positive coefficient) of the i^{th} metabolite taking part in the j^{th} reaction. If the coefficient is zero, the j^{th} reaction is neither consuming nor producing the i^{th} metabolite. Eq. (2) ensures that at steady state, all internal metabolite concentrations in a given metabolic system are balanced and hence, are produced and consumed in the same amount by the system intrinsic reactions. Thus, they do not accumulate or deplete over the investigated time frame. The inequality (3) ensures that irreversible reactions can only be used in the thermodynamical feasible direction.

Both constraints, stoichiometric balancing and feasible flux directions, restrict an allowed flux distribution to lie within a convex cone of the solution space (Fig. 1.4). Since the studies presented in chapter 3 of this thesis are based either on EFM or FBA analysis, both methods will be introduced in more detail further on.

1.2.1. Elementary Flux Modes

In principal, EFMs serve to decompose a given metabolic network in well defined and distinct minimal set of reactions that can operate at steady state (Schuster 1994; Schuster et al. 1999, 2000).

Essential to the concept of EFMs is the classification of metabolites as internal or external metabolites. Internal metabolites comprise intermediary components of metabolic systems and have to be in agreement with the steady state assumption (2). On the contrary, external metabolites can take part in reactions not included in the model and are allowed to vary in their concentrations. Hence, they form the source and the sink of the model. The categorization of internal or external metabolites has a noteworthy impact on the number of computed EFMs and thus, on the computation time (see also Fig. 1.5).

To allow for the computation of EFMs it is assumed that the metabolites of a given system are homogeneously distributed throughout the model. The time-course of the metabolites X_1, \dots, X_n can then be described by:

$$\dot{X} = SV(X, p) \tag{4}$$

with S being the stoichiometric matrix and V being a vector of net reaction rates,

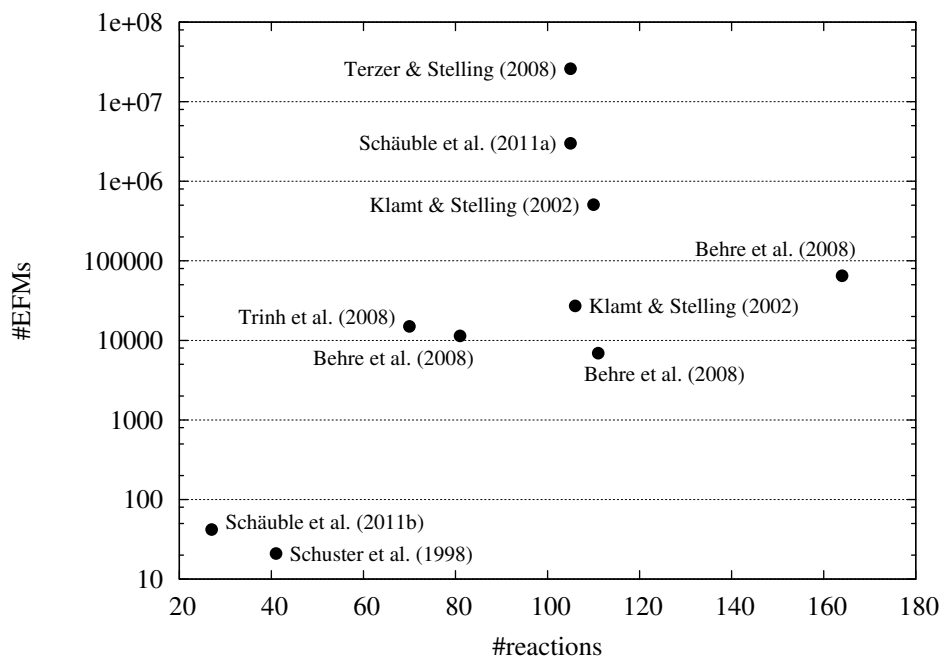


Figure 1.5.: Exemplified relationship between the number of reactions and the number of valid EFMs. The number of EFMs is given in logarithmic scale. The different magnitudes of EFM sets for models with approximately an equal number of reactions arise from different interconnectivity and number of internal metabolites of the respective models.

dependent on the metabolite X and parameters p , respectively. Since S comprises only internal metabolites, Eq. (4) can be simplified according to the steady state condition, as represented by Eq. (2) (Schuster 1994):

$$SV(X, p) = 0. \quad (5)$$

In fact, dynamic processes are naturally never completely constant over time. This assumption is still largely true in the approximate context of network models, as long as no intermediate concentration varies significantly in the studied time frame.

The basis for EFM computation is the determination of the nullspace of the stoichiometric matrix S . The nullspace, also known as kernel, describes the Euclidian subspace of all vectors V satisfying Eq. (5). The calculation of basis vectors corresponding to the nullspace already infers potential system pathways. However, a set of basis vectors does not account for irreversible reactions and is not unique by definition. This implies that desired pathways may not be found by a basis vector

analysis alone. Note that a reaction is also determined as being irreversible, if at any time point one reaction direction has a surpassing flux compared to the opposite direction. Furthermore, modeling gene knockouts or enzyme deficiencies could infer a loss of basis vectors, although still feasible fluxes might exist.

EFMs are designed to circumvent the limitations given by a set of basis vectors. First, the subvector V^{irr} of V needs to satisfy the inequality (3). The relations (3) and (5) then form a linear (in-)equality system, where the solutions to V are freely scalable. Thus, a flux mode is defined as follows:

$$F = \{V \in \mathbb{R}^r \mid V = \lambda V^*, \lambda > 0\}, \quad (6)$$

where F is a set of flux modes, r is the number of reactions and \mathbb{R}^r is the r -dimensional Euclidean solution space. The constant λ is an arbitrary scaling factor, which is commonly chosen as small as possible, for the sake of simplicity. The nonzero vector V^* is compliant with the following properties:

- (i) *sign restriction*: Inequality (3) is fulfilled
- (ii) *steady state condition*: V^* is in agreement with Eq. (5)

Still, the flux modes are not minimal and thus, not elementary. To ensure non-decomposability, V^* needs to be further restricted:

- (iii) *non-decomposability*: No vector \hat{V} with the following properties exists:

- \hat{V} satisfies (i) and (ii)
- $V_i^* = 0 \rightarrow \hat{V}_i = 0$ and it exists a j fulfilling: $\hat{V}_j = 0$ and $\hat{V}_j \neq V_j^*$

Therefore, a flux mode F is called elementary, if no valid vector \hat{V} exists, which contains more zero-positions or less active reactions than V^* . Finally, an EFM can be defined as:

Definition 1 *An elementary flux mode is a minimal set of reactions in a metabolic network that can operate at steady state.*

In contrast to basis vectors, EFMs are unique up to a scaling by $\lambda > 0$. Consequently, a complete set of EFMs corresponds to all possible minimal reaction pathways of a system. Naturally, feasible fluxes of a system can also be a superposition of EFMs:

$$V = \sum_k \phi_k e^{(k)}, \phi > 0, \quad (7)$$

where $e^{(k)}$ are feasible EFMs and ϕ ensures that the linear combination of EFMs satisfy the inequality (3).

Considering the formalism above, a noncyclic EFM can be understood as a non-decomposable path through a model converting one set of external metabolites into another, by consuming and producing a balanced set of internal metabolites. These information can be used to conduct e.g. molar yield studies (cf. Poolman et al. (2003); Trinh et al. (2008)). Nonzero cyclic EFMs are a special case of EFMs, where internal metabolites are transformed into each other in a circular manner and arise e.g. from futile cycles, where energy such as ATP is dissipated (Gebauer et al. 2012). A more thorough description of the concept of EFMs is given in Schauble et al. (2011b), chapter 3.

Fig. 1.6 represents a simple instructive model, demonstrating the definition of EFMs within a given network. The model, as defined by a model scheme and its corresponding stoichiometric matrix (Fig. 1.6a and 1.6b) gives rise to four EFMs (Fig. 1.6c-1.6f). Note that the complete network as represented by Fig. 1.6a is not a valid EFM itself, since it represents a linear combination of the four possible EFMs and can be decomposed.

A number of tools were developed, capable of EFM computation. The first algorithms were based on Gaussian elimination (Schuster 1994; Pfeiffer et al. 1999). After a number of improvements (Wagner 2004; Gagneur and Klamt 2004; Klamt et al. 2005; Wagner and Urbanczik 2005), the currently most efficient algorithm to enumerate all EFMs is given by Terzer and Stelling (2008). It makes use of a parallel computation of bit pattern trees to significantly speed up the calculation process. Despite the effort, the computation of all EFMs corresponding to notably large genome-scale metabolic models remains infeasible, regardless of the applied algorithm. Particularly genome-scaled networks comprise a multitude of branched pathways, which cause a combinatorial explosion of feasible minimal routes (see also Fig. 1.5). Consequently, the computation of EFMs increases exponentially with the network size. This becomes problematic when considering the increasing number of large or genome-scale metabolic networks that have been published in recent years (Kim et al. 2012). Here, either calculating a valid sample (Kaleta et al. 2009b) or enumerating the shortest valid EFMs (de Figueiredo et al. 2009) can circumvent this problem at the cost of deriving only a subset of valid EFMs.

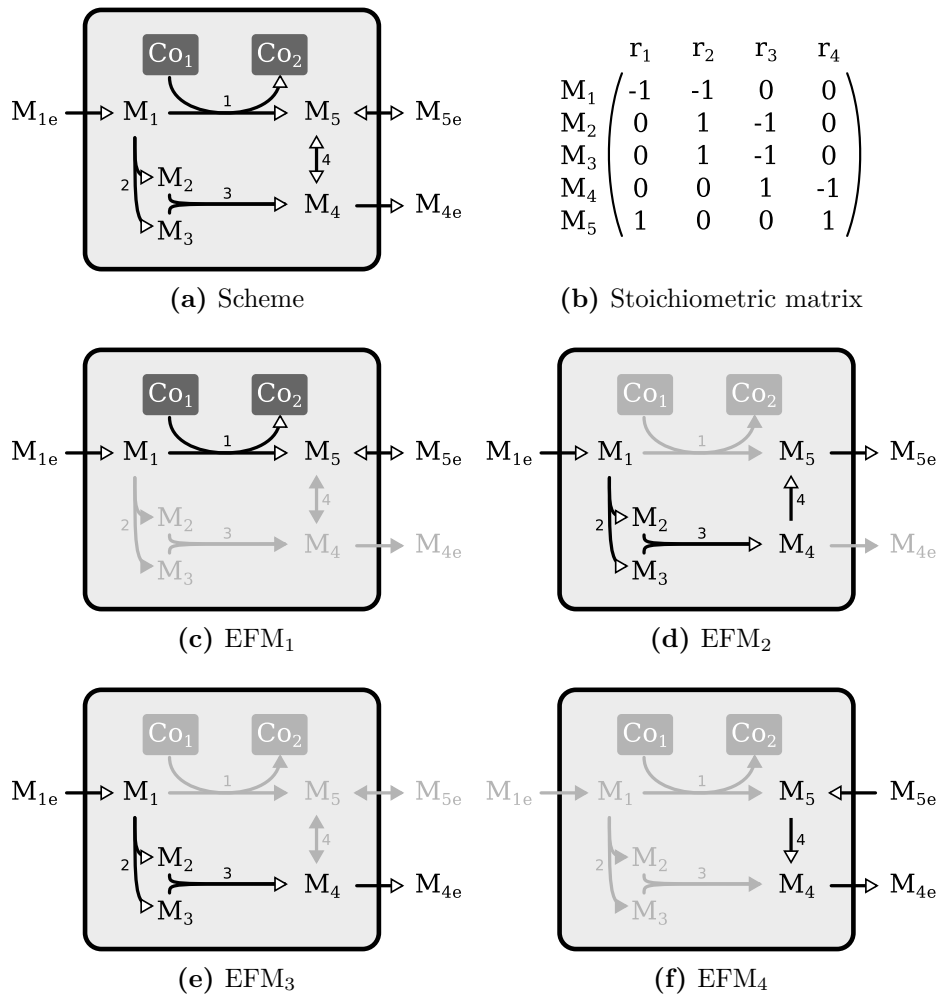


Figure 1.6.: Introductory model and associated EFMs. (a) Model scheme. Species M_{1e} , M_{4e} and M_{5e} are external versions of the respective internal metabolites. Co_1 and Co_2 resemble cofactors of reaction 1, which are external as well and need thus, not to be balanced. Arrows depict irreversible and reversible reactions. (b) Stoichiometric matrix associated to Fig. 1.6a. (c-f) All possible EFMs of the given model.

1.2.2. Flux Balance Analysis

Next to EFM analysis, flux balance analysis (FBA) can be considered one of the fundamental approaches of systems biology (Varma and Palsson 1994; Papin et al. 2004; Westerhoff and Palsson 2004).

When Fell and Small studied pathways in adipose tissues responsible for transforming glucose to fat in 1986, they and others started the successful story of FBA (Fell and Small 1986). They realized the effectiveness of an earlier suggestion made

by Watson (Watson 1984, 1986) who proposed to tackle possible biochemical fluxes by the means of linear programming. This approach is reasonable, since mass balance as well as capacity constraints restrict the solution to lie within a convex cone (cf. Fig. 1.4). The mass balance constraint is imposed directly by the stoichiometric matrix representation of the given network, whereas the capacity constraint is related to the lower and upper bounds of all reactions. Moreover, by manipulating the lower and upper bounds, uptake or excretion rates can be defined for specific metabolites (Savinell and Palsson 1992a). These can comprise ingredients of a given growth media or a desired product and can be matched to experimentally determined fluxes (Savinell and Palsson 1992b).

Having a reconstructed model at hand and adding sufficient knowledge about a feasible flux range, FBA can be used to predict flux through any objective reaction. A common objective is the optimization of growth. This can be quantified by the biomass function, which constitutes the yield at which metabolic compounds such as amino or nucleic acids, lipids and proteins are converted into biomass. More generally, the biomass objective function describes the rate at which all of the required biomass precursors are produced in the correct proportions (Feist and Palsson 2010). Altogether, the metabolic reactions are thus, defined by a system of linear equations and (in-)equality constraints. To compute an FBA a linear program is typically formulated as follows:

$$\begin{aligned} \max \quad & c^T V, \\ & SV = 0, \\ & lb \leq V \leq ub. \end{aligned} \tag{8}$$

Let n be the number of reactions contained in a given model. The objective vector c of length n is defined as a vector of weights, corresponding to the contribution of each reaction to the objective function. Vector V of length n represents the optimized flux through all reactions with respect to vector c . All feasible lower and upper bounds of the systems reactions are defined in the vectors lb and ub of length n , respectively. For irreversible reactions, the corresponding coefficients of lb are zero. Commonly, vector c is designed to maximize growth yield (maximizing the output of the biomass function). The linear programming formulation (8) is compliant with the steady state and irreversibility conditions as given by Eq. (2)

and the inequality (3) and can be solved efficiently by algorithms based on e.g. the simplex (Dantzig et al. 1955) or interior point approach (Karmarkar 1984). Note that the running time of the simplex approach is exponential at worst (compared to a polynomial runtime of interior point methods), but proved to be practically efficient in most cases. In fact, FBA is thus applicable to large-scaled metabolic networks up to genome-scale, in contrast to e.g. enumerating all EFMs.

Therefore it is of no surprise that FBA has been extensively used to test reconstructed metabolic models (Feist et al. (2007); Boyle and Morgan (2009); Oberhardt et al. (2009); Thiele et al. (2013)). Furthermore, FBA has been extended in several ways to account e.g. for dynamics of a metabolic network (Mahadevan et al. 2002), variable optimal fluxes (Mahadevan and Schilling 2003), minimization of metabolic adjustments (MOMA, Segrè et al. (2002)) or for flux balance in microbial communities (Khandelwal et al. 2013). These extensions are valuable, since particularly the classic approach of optimizing one objective function is questionable (Schuster et al. 2008; Feist and Palsson 2010). Obviously, microorganisms and more complex organisms are not simply optimizing one objective given an ever changing natural environment. In contrast, several optimal flux distributions potentially exist, which reflect e.g. robustness to nutritional availability or the capability of optimizing metabolism in communities of multiple bacteria (Schuetz et al. 2012).

1.3. Biological background

Taken together, modeling dynamic models with differential equations or constraint-based methods enables the study of a broad range of biological organisms and processes.

In chapter 2 and 3, I present five different studies, whereby two focus on the theoretical aspects of kinetic (Schäuble et al. (2013), chapter 2) and EFM based modeling (Schäuble et al. (2011b), chapter 3). In the remaining three studies, I applied the presented methods to study cellular senescence in primary human fibroblasts, circadian regulation in the green algae *Chlamydomonas reinhardtii* and metabolic costs of amino acid and protein production in *Escherichia coli*. In the remainder of this chapter, I will introduce the biological background of these studies in more detail.

1.3.1. Cellular senescence, the irreversible cell cycle arrest

DNA damage and mutations are thought to be key events in the manifestation of biological aging. The driving causes for prolonging life or longevity are still unknown, however, whereas particularly cellular maintenance systems are thought to recognize and repair damage, if possible (Vijg 2008).

It has been long known that *in vitro* cells show a limited replicative capacity, which ends up in a permanent and irreversible cell cycle lock (Hayflick and Moorhead 1961). This phenomena, termed cellular senescence, is regularly encountered along aging (Campisi and Sedivy 2009) and is known to potentially delay aging related diseases (Baker et al. 2011). In this context, cellular senescence might also play a tumor suppressing role (Krizhanovsky et al. 2008). Unique to senescent cells is their active metabolism alongside an indifference to mitogenic or apoptotic stimuli, despite their inability to proliferate (Chaturvedi et al. 1999; Marcotte et al. 2004). Senescent cells show notable changes in their cytoskeleton and feature an increased cell size, accompanied with additional numbers of lysosomes, vacuoles and mitochondria (Cristofalo et al. 2004). From a molecular perspective, specific biomarkers have been established and connected to cellular senescence (Dimri et al. 1995; Narita et al. 2003; Kuilman et al. 2010; Sikora et al. 2011). Among others, these include senescence-associated β -galactosidase (SA- β -Gal, Dimri et al. (1995)), telomere dysfunction-induced foci (TIF) (Herbig et al. 2004; Jeyapalan et al. 2007) and heterochromatin foci (SAHF). The latter are known to occur in primary human cell types after up-regulation of p-16, but not in all of them (Narita et al. 2003; Kosar et al. 2011). Moreover, a notable up-regulation of the cell cycle regulators p16, p21 and p53 is known (Robles and Adami 1998; Ressler et al. 2006). Finally, senescent cells express matrix-degrading proteases as well as inflammatory chemokines and cytokines, which are known as senescence-messaging secretome (SMS) or senescence-associated secretory phenotype (SASP) (Shelton et al. 1999; Acosta et al. 2008; Kuilman and Peeper 2009; Coppé et al. 2010).

Essential for the induction of senescence is a functioning cellular DNA damage response system and several pathways, including p53-p21 and p16-pRb (Kuilman et al. 2010). On the other hand, telomere-dependent cellular senescence is induced by an increased DNA damage response at dysfunctional telomeres (Herbig et al. 2004). Here, uncapped telomeres accumulate increased levels of phosphorylated H2AX, as well as several repair factors (d'Adda di Fagagna et al. 2003; Takai et al.

2003; Herbig et al. 2004; Jeyapalan et al. 2007; Nakamura et al. 2009). These factors initiate a number of downstream processes, leading to, among others, the p53 induced transcription of p21, which ultimately results in cell cycle arrest (Herbig et al. (2004); Gire et al. (2004); Goodarzi et al. (2008), reviewed in Ben-Porath and Weinberg (2004, 2005); Shay and Wright (2005); Cosme-Blanco and Chang (2008)). Whereas the p53-p21 pathway is always activated by cells with damaged telomeres, the p16-pRb pathway alone is not sufficient to induce DNA damage associated senescence (Herbig et al. 2004). Interestingly, DNA damaging treatments like irradiation trigger the same response cascade, if the DNA damage surpasses a certain threshold (Di Leonardo et al. 1994; Serrano et al. 1997; Robles and Adami 1998; Toussaint et al. 2000; Hwang 2002; Zhao et al. 2004; Debacq-Chainiaux et al. 2005; Havelka et al. 2007; Mallette et al. 2007). Overexpression of activated oncogenes, such as RAS, triggers oncogene-induced senescence (Serrano et al. 1997; Dimri et al. 2000; Pearson et al. 2000), which in turn is primarily dependent on the p16-pRb pathway (Pantoja and Serrano 1999). Due to their low or absent levels of p16, however, primary human fibroblasts are resistant to RAS-induced senescence (Benanti and Galloway 2004). Taken together, cellular senescence is understood as a systematic reaction to various stress inducing events, including DNA damage and telomere shorting (Rodier and Campisi 2011).

SA- β -Gal is generally considered to be a good quantitative senescence marker, particularly when combined with Ki67-negativity (absence of proliferation) and observation of developed DNA damage foci (Lawless et al. 2010). The activity of SA- β -Gal is detectable at high levels in senescent cells, but not in pre-senescent or reversibly cell cycle arrested (quiescent) mammalian cells (Dimri et al. 1995). Moreover, the marker can be induced by artificially triggering a cell cycle arrest in immortal cell lines, where otherwise the marker would be undetectable (Dimri et al. 1995).

Due to accumulating damage, the decreasing speed and finally stop of proliferation, an aged cell population can in fact be considered a mixture of proliferating, cell cycle arrested and senescent cells (Faragher et al. 1993; Kill et al. 1994). The cell cycle inhibitors p16 and p21 are suitable markers to reflect the arrested cell state, whereas SA- β -Gal potentially reflects a senescent cell. Together with the growth curves, these markers enable the study of mixed cell populations at any timepoint, which is the basis for the presented model in Schäuble et al. (2012) in chapter 2.

1.3.2. The green algae *Chlamydomonas reinhardtii* as model organism for nitrogen assimilation

Although an EFM based analysis results in a set of theoretically possible minimal fluxes, kinetic parameters are not considered. Thus, connecting regulatory events to EFM analysis is challenging. By including genomic sequence data, the dynamics of a system can still be investigated with a constraint-based modeling approach. To demonstrate the value of this procedure, we analyzed the nitrogen uptake system of the green algae *Chlamydomonas reinhardtii* in Schäuble et al. (2011a) presented in chapter 3. Here, we focused on the algae's capacity of integrating nitrogen into certain amino acids.

The haploid unicellular green algae is an exceptional model organism and has been used to study various biological processes such as photosynthetic mechanisms, microtubule assembly, flagella usage and movement, or mineral nutritions, among others (Rochaix et al. 1998). Moreover, its sequenced and annotated genome (<http://www.chlamy.org/>) allows for comparative genomics and functional studies.

Common to all phototrophic organisms is the necessity of incorporating inorganic nitrogen compounds into metabolic components, which is a key requisite of survival and growth (Daniel-Vedele et al. 1998). Although required in great abundances, plants must compete for nitrogen in the soil with (a-)biotic processes, such as microbial consumption or erosion. In consequence, plants have evolved efficient mechanisms to acquire different nitrogen sources already at low concentrations. Besides, whenever plants are harvested, soil nitrogen is removed and thus, remains only available at lower concentrations. Considering nitrate, whose concentration in natural soil can vary from 10 μM to 100 mM, it is either stored in the vacuoles or further reduced. Assimilating nitrogen from nitrate involves two transport and two reduction steps, occurring in the cytosol (nitrate reductase: $\text{NO}_3^- \rightarrow \text{NO}_2^-$) and the chloroplast (nitrite reductase: $\text{NO}_2^- \rightarrow \text{NH}_4^+$, cf. Fig.2, Schäuble et al. (2011a), chapter 3). Finally, ammonium is incorporated into carbon skeletons via the glutamine or glutamate pathway (Crawford and Arst 1993; Crawford 1995; Lam et al. 1995; Daniel-Vedele et al. 1998). The reducing energy is provided by NAD(P)H for nitrate reductase and reduced ferredoxin for nitrite reductase. Therefore, incorporating nitrogen is not only dependent on its availability in the phototrophic organism's environment, but also on sufficiently available energy. This is particularly challenging for plants at night. Notably, numerous genes and putative transporters

are involved in nitrogen assimilation of *C. reinhardtii*. Considering as well that small quantities of nitrite are toxic and perturb plant growth (Oke 1966), the importance of sensitive regulation of the nitrogen uptake pathway becomes apparent (Fernández and Galván 2007).

In fact, the nitrogen metabolism of the green algae is circadian-clock regulated, which in turn is influenced by the heteromer CHLAMY1, among others. This mRNA binding factor consists of two subunits, C1 and C3, whereby the latter is well conserved in humans (Zhao et al. 2004). CHLAMY1 is known to recognize and bind UG repeats of at least seven repetitions, which are located in the 3' UTR of a number of mRNAs of e.g. nitrite reductase or argininosuccinate lyase, both key enzymes in the nitrogen and arginine pathway, respectively (Mittag 1996; Waltenberger et al. 2001; Zhao et al. 2004). Moreover, by introducing UG-repeat motifs in the 3' UTR of reporter luciferase genes, these genes obtain circadian expression (Kiaulehn et al. 2007). The binding activity of CHLAMY1 is circadian-clock controlled, as is indicated by its activity increase at dusk, which decreases again at dawn. Since it has been shown that the activity of nitrite reductase and UG_{≥7}-repeat reporter constructs peak along sunrise, CHLAMY1 is assumed to prevent translation of recognized mRNAs during the night (Iliev et al. 2006; Kiaulehn et al. 2007). Furthermore, at least the subunit C3 has been shown to interact with further clock components, highlighting its importance to adjust to daily light and temperature cycles (Dathe et al. 2012).

To assess the nitrogen uptake capability of *C. reinhardtii*, we connected the nitrogen uptake pathway to energy providing pathways of the central metabolism and to possible sink pathways of suitable amino acids. To study the influence of circadian regulation, we considered partial as well as complete translation prevention of CHLAMY1 affected mRNAs. Although the kinetic parameters of all involved enzymes are not available, we can still show that the impact of circadian regulation on *C. reinhardtii* is substantial.

1.3.3. The microorganism *Escherichia coli* in a biotechnological context

Biotechnological applications have been used by mankind already for thousands of years. In general, biotechnology aims at modifying living organisms in order to improve various processes, such as milk production, or the cultivation of high yield

crops. Since the end of the last century, biotechnology significantly advanced with the advent of genomics and recombinant gene technologies, which allows the modification and industrial usage of microorganisms. Since *Escherichia coli*, a Gram-negative and facultative anaerobic bacteria, is comparably easy to genetically manipulate, it emerged as one of the most used bacteria to synthesize desired biological products (Lee 1996). Specifically, *E. coli* is used for the production of recombinant proteins, which in turn are essential for medical, technical as well as scientific purposes (Makrides 1996; Baneyx and Mujacic 2004; Kayser et al. 2005; Heizer et al. 2006; Carlson 2007). Moreover, *E. coli* is extensively used to produce amino acids, which in turn are used by the pharmaceutical or cosmetical industry or as food additives (Wendisch et al. 2006; Park and Lee 2010a,b). *E. coli*'s success in bioengineering is also based on the possibility to let it grow to high cell densities, particularly when providing defined medium with glucose or better, glycerol as carbon substrate (Korz et al. 1995; Rinas et al. 1995; Lee 1996; Shiloach and Fass 2005). On the other hand, at unbalanced conditions, together with an excessive availability of carbon sources, *E. coli* excretes acetate, which can prevent further growth, once acetate accumulates to toxic concentrations (Luli and Strohl 1990; Wolfe 2005; Shiloach and Rinas 2009). Acetate at moderate concentrations, however, can also serve as carbon source (Wolfe 2005).

Thus, from a biotechnological perspective it is of importance to elucidate the possible yields of desired end-products, like amino acids and proteins based on suitable major carbon compounds, such as glucose, glycerol or acetate. To allow for an unbiased prediction of amino acid production capabilities, the complete information of the entire network should be taken into account. The substantial influence of co-factors or precursors on metabolic production pathways is only then fully integrated in computed yield numbers. Therefore, the previously computed numbers, based on limited models need to be treated with caution (Varma et al. 1993; Burgard and Maranas 2001). Based on a genome-scale metabolic model (Feist et al. 2007), we updated these numbers and additionally computed measures of ATP consumption and per kilodalton costs of flux distributions (Kaleta et al. (2013), chapter 3). By comparing these numbers across different amino acids, we derived a comprehensive overview of metabolic costs in *E. coli*. This can potentially serve to setup further industrial designs of producing amino acids or proteins in the future.

2. Dynamic Modelling

Substrate competition and cellular senescence

Chapter summary. *This chapter is dedicated to the study of dynamic models based on ODE modeling approaches. The first study elucidates the impact of substrate competition on kinetic rate laws. We show that certain network structures result in notable concentration changes of modeled entities, if substrate competition is ignored. Therefore, our results should influence mechanistic modeling, whenever one of the comprised study cases are contained in the studied biological process. Furthermore, we derived reversible kinetic rate laws considering substrate competition that might simplify further studies. In the second study, we included experimental data of primary human fibroblasts to develop a dynamic model of cellular senescence. This model differentiates between proliferating, cell cycle arrested and senescent cells and allows for a quantitative description of respective growth curves. Moreover, by incorporating experimentally derived data of biomarkers, we estimate the transition rates between these cellular states. Thereby, our model enables the evaluation of the specificity of further biomarkers to reflect a particular cellular state.*

journal homepage: www.FEBSLetters.org

Effect of substrate competition in kinetic models of metabolic networks

Sascha Schäuble^{a,1}, Anne Kristin Stavrum^{b,1}, Pål Puntervoll^c, Stefan Schuster^d, Ines Heiland^{e,*}^a Theoretical Systems Biology Group, Friedrich-Schiller-University Jena, Germany^b Department of Informatics, University of Bergen, Bergen, Norway^c Computational Biology Unit, Uni Computing, Uni Research, Bergen, Norway^d Department of Bioinformatics, Friedrich-Schiller-University Jena, Germany^e Department of Arctic and Marine Biology, University of Tromsø, Norway

ARTICLE INFO

Article history:

Received 8 May 2013

Revised 18 June 2013

Accepted 18 June 2013

Available online 26 June 2013

Edited by Athel Cornish-Bowden

Keywords:

Substrate competition

Kinetic rate laws

Modelling

Metabolic pathway

ABSTRACT

Substrate competition can be found in many types of biological processes, ranging from gene expression to signal transduction and metabolic pathways. Although several experimental and *in silico* studies have shown the impact of substrate competition on these processes, it is still often neglected, especially in modelling approaches. Using toy models that exemplify different metabolic pathway scenarios, we show that substrate competition can influence the dynamics and the steady state concentrations of a metabolic pathway. We have additionally derived rate laws for substrate competition in reversible reactions and summarise existing rate laws for substrate competition in irreversible reactions.

© 2013 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

1. Introduction

Substrate competition has been reported to have implications in different biochemical processes, including degradation of polymeric carbohydrates [1], plant secondary metabolism [2], metabolic transport [3–5], signal transduction pathways [6] and gene regulation [7,8]. All these have in common that different substrates compete for the active site of the same enzyme. Substrate competition is also used to describe biochemical mechanisms where two different enzymes compete for the same substrate. L-Arginine, for example, is a substrate for both nitric oxide synthase and arginase, and competition between the enzymes plays a role in asthma development [9]. This type of competition has also been described as a possible mechanism behind changes in methylation patterns in cancer cells [10]. While the reaction rates of the latter type of competition can be described by standard Michaelis–Menten kinetics (MMK), descriptions of reaction rates of the first type of substrate competition require some modifications. Surprisingly, applicable rate laws describing competition between different substrates for the same enzyme are not available for reversible reactions. Although the rate laws described here can be used for

substrate competition in different cellular processes, our examples will focus on metabolic processes.

For irreversible reactions substrate competition is comparable to enzyme inhibition. The different substrates can be viewed as inhibitors of each others reactions, and hence, the mechanism of substrate competition can be described by adapting the kinetic rate laws from competitive inhibition. The mechanisms of enzyme inhibition have been thoroughly investigated for decades, and the kinetics are mostly based on the original reaction rate equation by Henri, Michaelis and Menten [11–13]. In 1977, Chou and Talaly published a generalised equation for the analysis of multiple inhibitors for various mechanisms of irreversible reactions [14]. Furthermore, they provided general rules for the different inhibition mechanisms that can be applied to various kinetic rate laws. Dingerkus et al. [15] used a rate law that is similar to irreversible MMK with competitive inhibition to describe the competition between tryptophan and other amino acids to get across the blood–brain barrier. Although the reverse transport rate might be low under physiological conditions, the amino acid transporters are indeed reversible. Thus, to describe the dynamics of these transport processes more accurately, reversible rate laws are required. To our knowledge, explicit kinetic rate laws that describe steady states of reversible reactions, which include substrate competition, are not available in the literature. In contrast to irreversible reactions, the competitive binding of the product must also be considered for reversible reactions. So far, only kinetic rate laws for the initial

* Corresponding author. Tel.: +47 77646889

E-mail address: ines.heiland@uit.no (I. Heiland).¹ These authors have contributed equally to this work.

velocity in the absence of products have been described [16,17]. However, these rate equations are not suitable for steady state calculations. To close this gap we applied the rules provided by Chou and Talaly [14] to derive rate laws for reversible reactions, based on the quasi steady state assumption.

To study the impact of substrate competition, we constructed three toy models that resemble real pathway scenarios. We used the models to simulate the impact of substrate competition on: (i) substrate accumulation over time, (ii) steady state concentrations of intermediates for increasing substrate concentrations, and (iii) the metabolic capacity of the system. Although it may be valid to neglect competition in some cases, we show that it is very difficult to safely judge whether this is appropriate in complex pathways.

2. Methods

Three generalised toy models were made to study the effect of substrate competition as a result of enzymes catalysing multiple reactions in (A) different pathways, (B) different branches within one pathway or (C) multistep reactions (schemes see Figs. 1–3). Four versions of each model were set up: the first two models describe an irreversible mechanism, of which one includes competition and the other neglecting it. The other two models contain reversible reactions, again one containing competition and the other neglecting it. To ensure that the observed effect was due to competition alone, we standardised the setup of the models. The input reactions were modelled using mass action kinetics (Eq. (1)), whereas transitions between intermediate species and output reactions were modelled using standard Michaelis–Menten kinetics for irreversible (Eq. (2)) and reversible (Eq. (3)) reactions. The Michaelis–Menten constant (K_m) was arbitrarily set to 0.02 mM and maximal velocity (V) was set to 1 mM/h, for the respective parameters in all reactions.

$$v = k_1 S \quad (1)$$

$$v = \frac{VS}{K_m + S} \quad (2)$$

$$v = \frac{V_f \frac{S}{K_m^S} - V_r \frac{P}{K_m^P}}{\frac{S}{K_m^S} + \frac{P}{K_m^P} + 1} \quad (3)$$

The full description of all used models can be found in the [Supplementary material](#).

2.1. Substrate competition

Based on the rules described by Chou and Talaly [14] the rate law for monomolecular irreversible reactions with any number of competing substrates can be described as follows:

$$v_1 = \frac{VS_1}{K_{m_1} \left(1 + \sum_{i=2}^n \frac{S_i}{K_{m_i}} \right) + S_1} \quad (4)$$

where S_1 competes with $n-1$ substrates S_2, \dots, n for the binding site of the catalysing enzyme. The variables K_{m_i} describe the Michaelis constants for the respective substrates S_i . An equally simple relationship could not be found in the literature for reversible reactions. A general formulation for reversible reactions including competition between multiple substrates can be deduced by recognising that not only the substrates but also the products compete for the binding site of the free enzyme. Thus the K_m -values have to be modified as follows:

$$K_{m_1} = K_{m_1} \left(1 + \sum_{i=2}^n \left(\frac{S_i}{K_m^S} + \frac{P_i}{K_m^P} \right) \right) \quad (5)$$

The resulting kinetic rate law for mono-molecular mechanism then becomes:

$$v_1 = \frac{V_f \frac{S_1}{K_m^S} + V_r \frac{P_1}{K_m^P}}{\sum_{i=1}^n \left(\frac{S_i}{K_m^S} + \frac{P_i}{K_m^P} \right) + 1} \quad (6)$$

with S_1 and P_1 competing with $n-1$ other substrates S_2, \dots, n and $n-1$ other products P_2, \dots, n for the binding site of the catalysing enzyme.

Kinetic rate laws for other reaction mechanisms as well as the derivation for the mono-molecular reaction (Eq. (6)) can be found in the [Supplementary materials](#).

2.2. Steady state analysis and time course simulations

Steady state analyses are commonly performed to predict species concentrations and reaction rates. To see the effect of competition on predicted species concentrations, we used COPASI [18] to analyse the influence of increasing input-species concentrations on the steady state of the respective toy model. External concentrations were varied from 0.001 to 1 mM with a step size of 0.001 mM. The results for the concentrations at which a steady state was found were used to calculate the difference between the corresponding models including or neglecting competition. The last concentration in the scan that yielded a steady state was considered to be the metabolic capacity of the model corresponding to the saturating concentration of the system.

Additionally, time course calculations were performed to study the differences over time. The uptake rate was set to 1 mM/h and the concentration of the external metabolites A_{ex} and B_{ex} was set to 0.05 mM.

3. Results

In irreversible reactions substrate competition can be described by competitive inhibition kinetics by substituting inhibitory Michaelis–Menten constants K_i with the respective K_m values of the competing substrates (see Section 2 for details). This is possible as reactions with substrate competition and reactions with competitive inhibition both have the same number of competing ligands. In reversible reactions, however, both substrates and the respective products compete for the active binding site of an enzyme. The number of compounds that influence enzyme kinetics is $2n$, which is the sum of n competing substrates and the n competing products. In contrast, an irreversible competitive inhibition describes n ligands influencing the kinetics of an enzyme, as it is accessible to $n-1$ inhibitors and one substrate. Hence, substrate competition in reversible reactions cannot be simulated by rate laws describing competitive inhibition.

Surprisingly, explicit kinetic rate laws for substrate competition of reversible reactions were not available in the literature. To derive these rate laws we modified the rule provided for irreversible competitive inhibition by Chou and Talaly [14]. The modification was based on the consideration that in reversible reactions both substrate and product compete for the binding to the active site of the enzyme (details see Section 2). This modified rule (Eq. (5)) was subsequently applied to derive kinetic rate laws for mono- and bimolecular reactions of different types (see Eq. (6) and [Supplementary material](#)). Our modified rule was proven to be correct for a monomolecular reaction, by deriving the kinetic rate law with an independent method ([Supplementary material](#)).

2820

S. Schäuble et al./FEBS Letters 587 (2013) 2818–2824

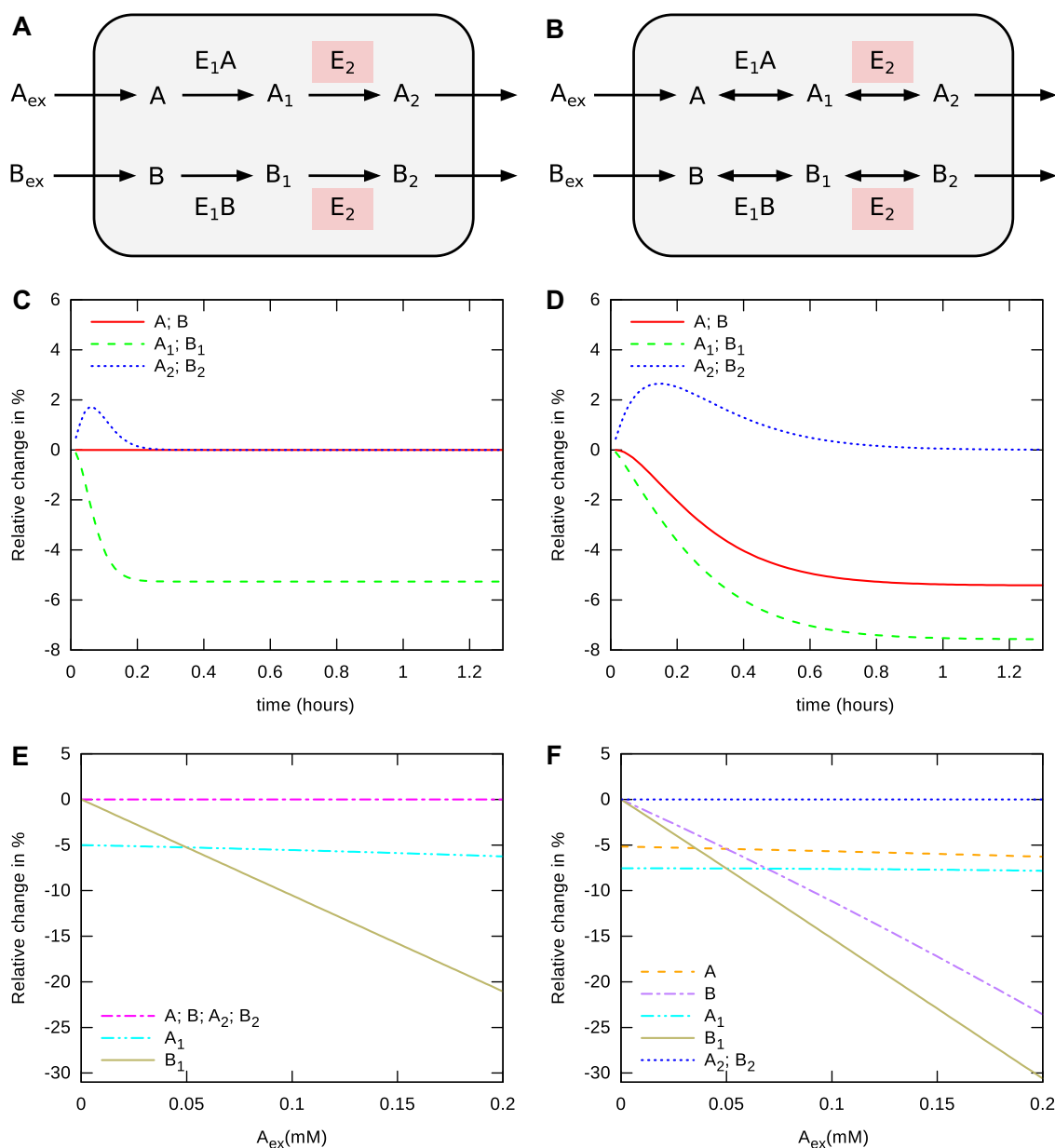


Fig. 1. Competition between different pathways. (A, B) Schematic diagram of two independent pathways sharing one common enzyme. One reaction in each branch is catalysed by the same enzyme (highlighted). The y-axis shows the relative difference in metabolic concentrations over time (C and D) or at steady state (E and F) between the model considering substrate competition and the model neglecting it. We considered two different sets of reactions. In the left panels (A, C, E) all reactions are irreversible. In the right panels (B, D, F) all reactions except inflow and outflow reactions (reactions leaving or entering the grey area) are reversible.

The effect of substrate competition was systematically studied in metabolic networks by mimicking three different types of substrate competition: competition between substrates in different pathways (Case A), competition between substrates of different branches within the same pathway (Case B) and multi-step reactions catalysed consecutively by two alternating enzymes (Case C). For each case we constructed four models: The first two resemble an irreversible reaction mechanism, where one includes and one neglects substrate competition. The next two models describe a reversible reaction scenario, again one including and one

neglecting substrate competition. The models were used to analyse the effects of competition in simulations, by observing: (i) changes over time for all substrates before reaching a steady state, (ii) changes in the actual steady state concentrations of the substrates, and (iii) changes in saturation levels for the respective pathways.

3.1. Case A – Competition between different pathways

The first scenario comprises two or more independent pathways that contain one or more common enzymes or transporters.

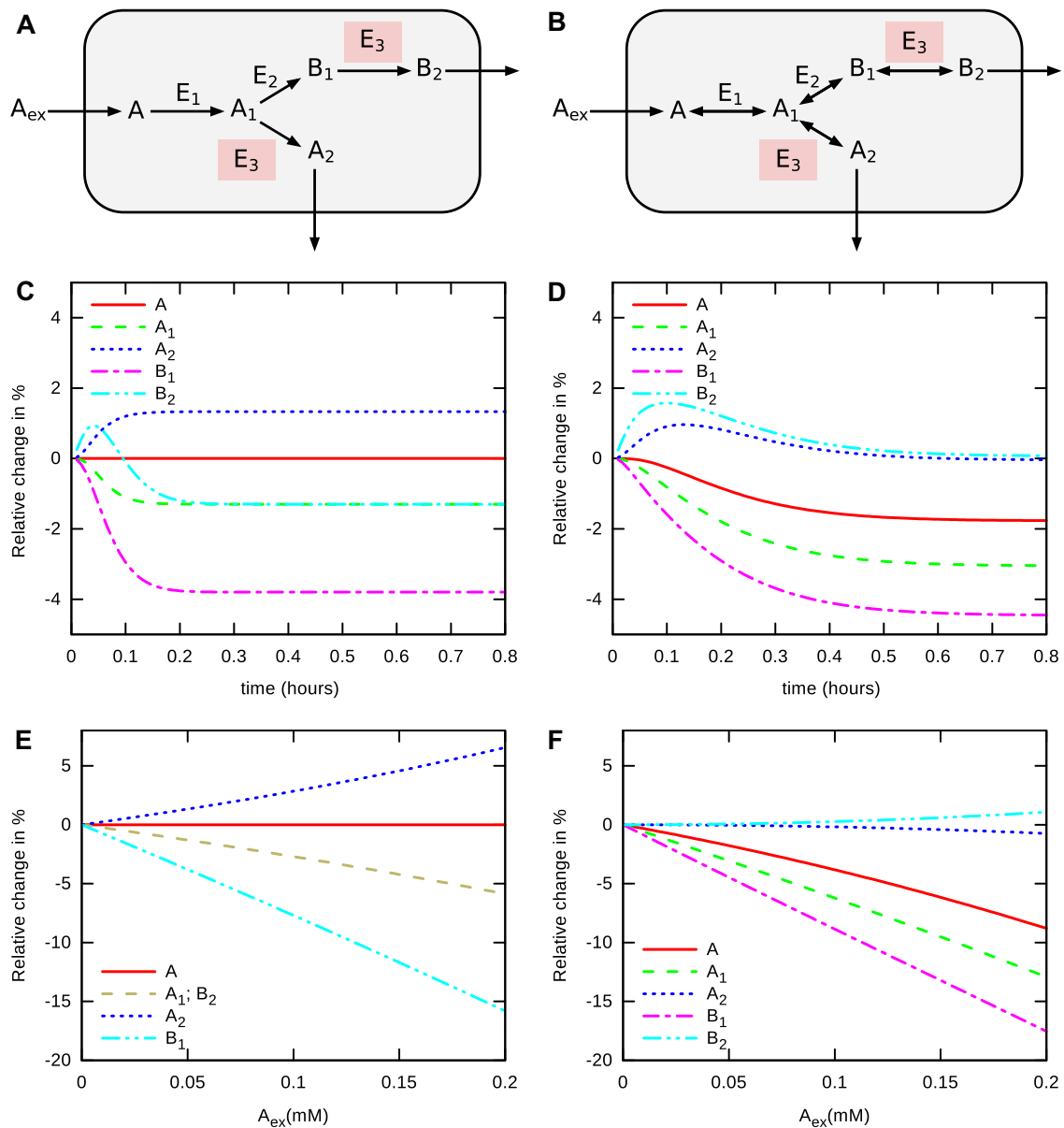


Fig. 2. Competition within a branched pathway. (A, B) Schematic diagram of a branched pathway. One reaction in each branch is catalysed by the same enzyme (highlighted). The y-axis shows the relative difference in metabolic concentrations over time (C and D) or at steady state (E and F) between the model considering substrate competition and the model neglecting it. In the left panels (A, C, E) all reactions are irreversible. In the right panels (B, D, F) all reactions except inflow and outflow reactions (reactions leaving or entering the grey area) are reversible.

As an example, neutral amino acids are metabolised by independent pathways, but depend on common transporters to get across the cell membrane. Experiments have shown that the competition between these amino acids for transport across membranes is significant [5].

To simulate effects of competition between pathways, we constructed two simple linear pathways (see Fig. 1A and B). One reaction in each pathway was set to be catalysed by the same enzyme. Consequently, its substrates will compete for the active site.

First, we investigated the influence of competition over time by analysing the accumulation of intermediate substrates (Fig. 1C and D). The external substrates A_{ex} and B_{ex} were both set to 0.05 mM,

which is above the half saturation points ($K_m = 0.02$ mM) used in the models. The plots show that neglecting competition led to underestimation of upstream competing substrate (A_1 and B_1) concentrations and overestimation of final product (A_2 and B_2) concentrations at early time points. In the reversible system, upstream substrate (A and B) concentrations were also underestimated.

Second, we scanned the dependency of the steady state concentration of the intermediates on the availability of external substrates. Because the two pathways are identical, a scan of external metabolite A_{ex} will provide the symmetric result of scanning B_{ex} . Therefore, we chose to only scan A_{ex} , and used values from far below (0.001 mM) to far above (1 mM) the K_m values

2822

S. Schauble et al./FEBS Letters 587 (2013) 2818–2824

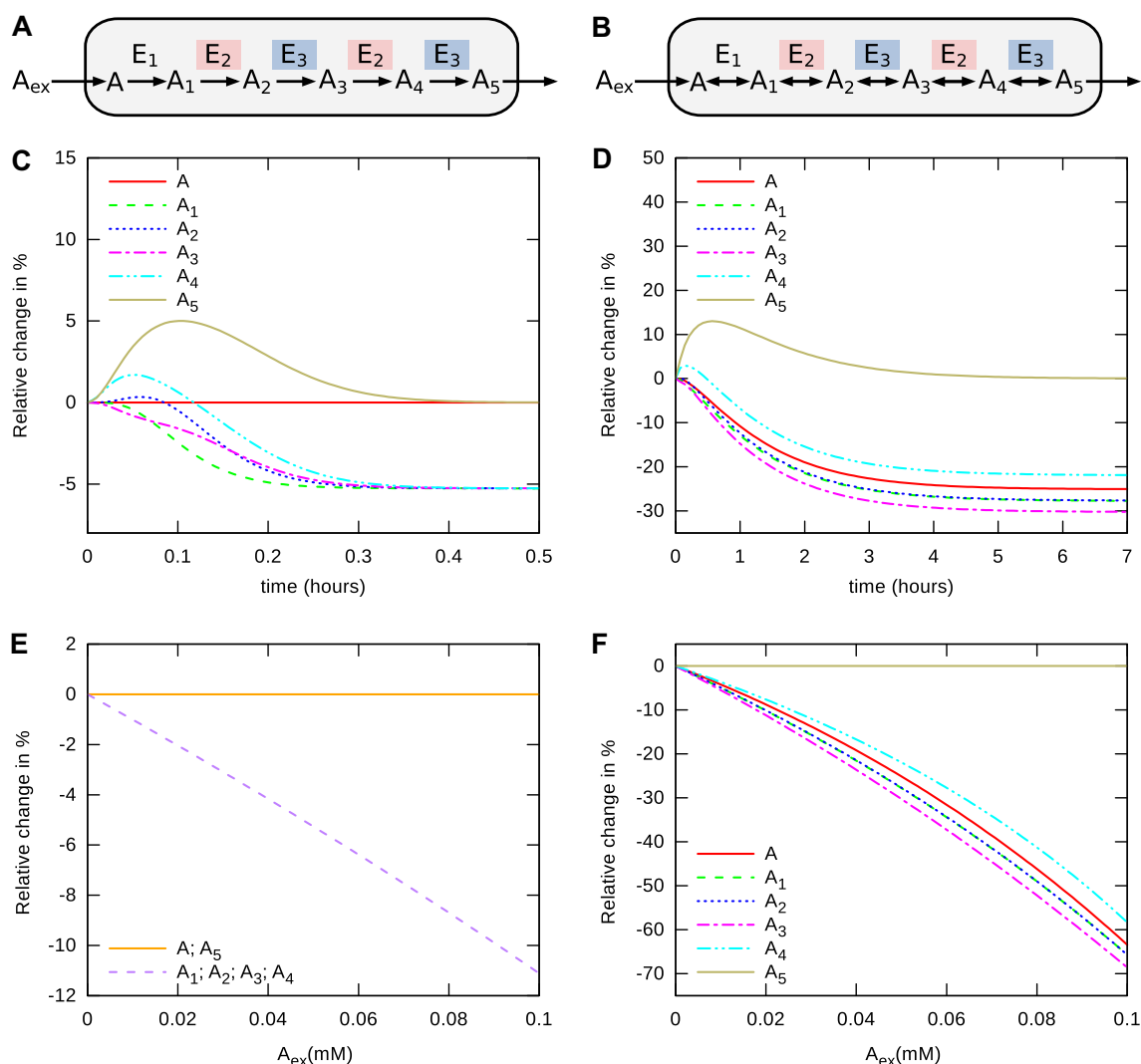


Fig. 3. Multistep pathway with alternating enzymes. As in the previous figures, we considered either a pathway with solely irreversible reactions (left panels) or with all internal reactions being reversible but inflow and outflow reaction being irreversible (right panels). A and B show the respective reaction schemes. The enzymes catalysing more than one reaction are highlighted. In C and D the relative differences of the final product over time has been calculated. In E and F the differences in steady state concentration between models that consider substrate competition and models that do not are plotted against the concentration of the external substrate A_{ex} .

(0.02 mM), while B_{ex} was kept constant at 0.05 mM. Fig. 1 E shows that only the concentration of the upstream competing metabolites (A_1 and B_1) were affected by substrate competition in the irreversible scenario.

In the reversible case, on the other hand, all metabolites upstream of the competing reaction (A , B , A_1 and B_1) showed differences in concentrations between the model including competition and the one neglecting it (Fig. 1F). In both the irreversible and the reversible models, neglecting competition led to a decrease in the steady state concentration of A_1 (and A in the reversible case) when the availability of external A_{ex} species increased. Thus, with increasing A_{ex} concentrations, A_1 predominantly occupied enzyme E_2 , resulting in accumulation of B_1 (and B for the reversible case) upstream of the competing reaction.

Finally, we analysed the effects of competition on the metabolic capacity of the modelled system. The capacity of the system can be measured by identifying saturation points, beyond which one or

more intermediates accumulate, and steady state can no longer be achieved. Simulations showed that the metabolic capacity was lower for models that considered substrate competition (see Table 1).

3.2. Case B – Competition in a branched pathway

The second scenario is a linear pathway that branches into two or more consecutive reactions, which lead to different final products (see Fig. 2A and B). The competition between substrates in different branches of a pathway occurs for instance in the tryptophan pathway. Tryptophan is an essential amino acid in the human diet and is involved in the synthesis of a variety of important metabolites, such as serotonin or NAD. The enzyme indoleamine-2,3-dioxygenase (IDO, EC: 1.13.11.52) catalyses four different reactions in different branches, and is one of several enzymes that catalyse multiple reactions in the tryptophan pathway.

Table 1

The values represent the highest external concentration that allow for a steady state.

	Case A		Case B		Case C	
	irrev. (mM)	rev. (mM)	irrev. (mM)	rev. (mM)	irrev. (mM)	rev. (mM)
with comp.	0.95	0.95	1	1	0.5	0.13
w/o comp.	1	1	1	1	1	0.95

Abbreviations: with comp., with competition; w/o comp., without competition; irrev., irreversible; rev., reversible.

To investigate the effects of competition within one pathway, we created a toy model where two intermediates in different branches compete for the enzyme E_3 (Fig. 2A and B).

The time-dependent changes of the intermediates of the pathway are shown in Fig. 2C and D. Investigating the model version with irreversible reactions, a transient phase needs to be distinguished from the state reached asymptotically. In both phases, neglecting substrate competition has an effect, though in different directions (see Fig. 2C). In the model with reversible reactions (Fig. 2D), the production of both A_2 and B_2 is initially faster in the model ignoring the substrate competition, since A_1 , A_2 and B_2 are connected by reversible reactions and are thus balanced.

By investigating the steady state concentrations, we see that all metabolites (except A in the irreversible case) are changed when ignoring substrate competition (Fig. 2E and F), but the effects are different for irreversible and reversible reactions. Although both the time dependent behaviour and the steady state concentration are different between the model considering substrate competition and model neglecting it, the saturation points for all models are equal (Table 1).

3.3. Case C – Multistep reactions with alternating enzymes

The last case we studied is a pathway, where multiple consecutive reactions are catalysed by the same enzyme. Prominent examples can be found in pathways that process polymers, such as starch and glycogen synthesis and degradation, and fatty acid degradation [19,20,1]. In all these cases, the enzymes accept substrates of different chain lengths, sometimes with increasing or decreasing affinities. In the case of fatty acid degradation, several enzymes process the substrate in an alternating manner. Based on the latter example, we constructed a toy model with two alternating enzymes in a linear pathway (Fig. 3A and B).

In this multistep scenario, the time-dependent simulations showed that the concentrations of most intermediates was lower in the models neglecting competition. However, some of them were transiently higher, including the production of the final product A_5 (Fig. 3C and D). In contrast, the steady state concentration of A_5 was the same for all models, regardless of substrate competition (Fig. 3E and F). The steady state concentrations of all other intermediates were lower in the models without competition, since the intermediates did not compete for the catalytic binding sites.

Finally, the saturation points, and thus the metabolic capacity of the pathway, was much lower in models with substrate competition (Table 1). Thus, for this pathway, neglecting substrate competition leads to a large overestimation of metabolic capacity.

4. Discussion

Recent studies show that the contribution of substrate competition to the dynamics of biomolecular networks can be significant [1,21,6,15,22,20,23]. Although elegant, most of these modelling approaches are mathematically complex and do not provide general, readily applicable rate laws for constructing models that consider substrate competition. We have therefore reviewed and derived explicit kinetic rate laws for various types of reaction mechanisms.

For irreversible systems, standard MMK rate laws can be extended to include competition between an arbitrary number of substrates [14]. We have complemented these results by deriving rate laws for reversible monomolecular reactions, and for multireactant reactions, such as ping-pong and bi-bi mechanisms. Moreover, we have provided a description of how to modify other existing rate laws, based on the modification of the rules originally proposed by Chou and Talaly [14].

To investigate the effect of including substrate competition, we constructed and investigated three toy models. The results show that ignoring competition can potentially influence the time dependent behaviour of the models, the steady state concentrations of intermediate and final products, and the metabolic capacity of a pathway. The degree of influence depends on the structure of the pathway(s), the position of enzymes that face competing substrates within a pathway, and the concentration of the competing substrates. Substrates with concentrations far below the enzyme's Michaelis–Menten (K_m) constant will result in very little competition, while at substrate concentrations above K_m , competition can be significant. In our time dependent simulations, external metabolite concentrations were set just above the K_m value used in the models. At this level we expected to see moderate effects from the competition between the two substrates. To assess the effects of increasing competition on the dynamics of the system, steady state simulations were performed with external concentrations ranging from far below to far above the model K_m value. The most pronounced effect was observed in case C, which resembles for example polymer chain prolongation. Here, simulations that ignored competition underestimated the time needed for substrate accumulation, the steady state concentrations of most intermediate substrates, as well as the metabolic capacity of the system. It has furthermore been noted earlier for a similar case that substrate competition influences metabolic control analysis of the system [24].

The other two cases displayed more moderate, but notable, changes. In case A, competition between two pathways, time dependent simulations showed a modest impact from competition, and a more notable effect on intermediate steady state concentrations. In contrast, the steady state concentrations of the final products were unaffected.

In the more complex case B, with competition between two pathway branches, substrate competition affected both steady state concentrations and time dependent changes in concentrations of most metabolites. The level of competition was similar to that observed in case A.

In general these result show that models ignoring substrate competition will, to a varying degree, give incorrect estimations of pathway dynamics. In certain scenarios, however, competition can apparently be safely ignored, such as in the study of steady state concentrations of final product of a pathway, that is structurally similar to case A and C. However, the simple toy models used here do not encompass the complexity of real biological pathways, and hence possible effects of substrate competition should be explored. One example that can illuminate this is the tryptophan metabolic pathway, which is highly branched. One of the pathway enzymes, aromatic L-amino acid decarboxylase (DDC, EC 4.1.1.28),

catalyses reactions both in different branches of the tryptophan pathway, and in other amino acid pathways (e.g. tyrosine metabolism and histidine metabolism). A model that includes all of the reactions catalysed by DDC would therefore be described by a combination of cases A and B. The effect of substrate competition on the behaviour and control of metabolism in this case could be more difficult to predict without explicitly modelling substrate competition.

A wide range of pathways, including glycolytic, fatty acid, and amino acid pathways, incorporate several enzymes that can process different substrates from the same or different pathways. In living cells, particular events of substrate competition may be prevented by compartmentalisation or tissue specific usage of certain pathways or branches of a pathway [25]. In such cases one might argue for ignoring competition in models. However, de-compartmentalised models are often assumed, for the sake of simplicity, or because information about compartmentalisation is not available. It may thus be difficult to judge whether substrate competition should be included or not. To help solve this issue, more experiments characterising metabolite transporters and the compartmentalisation of compounds and enzymes are needed. To capture the true dynamics of pathways where substrate competition is suspected or has been determined, competition should be explicitly modelled. The results presented here, together with the cited experimental work on substrate competition, strongly suggest that future modelling efforts should consider substrate competition.

Acknowledgements

We thank Christoph Kaleta for helpful discussions and Simon Oddy for reading the manuscript and providing valuable feedback. This work was supported by the Research Council of Norway through the Norwegian-German Collaboration Support Scheme [Grant No. 191721], the eVITA program [Grant No. 178885/V30], and funding from the German Federal Ministry of Education and Research [BMBF, Project Nos. 0315591D and 0315890C] within the Research Initiative in Systems Biology of Ageing (GerontoSys), from the German academic exchange service (DAAD) within the German-Norwegian Collaborative Research Support Scheme.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.febslet.2013.06.025>.

References

- [1] Kartal, Ö., Mahlow, S., Skupin, A. and Ebenhöf, O. (2011) Carbohydrate-active enzymes exemplify entropic principles in metabolism. *Mol. Syst. Biol.* 7, 542.
- [2] Knoke, B., Textor, S., Gershenzon, J. and Schuster, S. (2009) Mathematical modelling of aliphatic glucosinolate chain length distribution in *Arabidopsis thaliana* leaves. *Phytochem. Rev.* 8 (1), 39–51.
- [3] Marquez, B. and Van Bambeke, F. (2011) ABC multidrug transporters: target for modulation of drug pharmacokinetics and drug–drug interactions. *Curr. Drug Targets* 12, 600–620.
- [4] Pardridge, W.M. and Choi, T.B. (1986) Neutral amino acid transport at the human blood–brain barrier. *Fed. Proc.* 45, 2073–2078.
- [5] Salter, M., Knowles, R.G. and Pogson, C.I. (1986) Quantification of the importance of individual steps in the control of aromatic amino acid metabolism. *Biochem. J.* 234, 635–647.
- [6] Kim, Y., Andreu, M.J., Lim, B., Chung, K., Terayama, M., Jiménez, G., Berg, C.A., Lu, H. and Shvartsman, S.Y. (2011) Gene regulation by MAPK substrate competition. *Dev. Cell* 20 (6), 880–887.
- [7] Lee, S.P., Censullo, M.L., Kim, H.G. and Han, M.K. (1995) Substrate-length-dependent activities of human immunodeficiency virus type 1 integrase in vitro: differential DNA binding affinities associated with different lengths of substrates. *Biochemistry* 34, 10215–10223.
- [8] Jöres, L. and Wagner, R. (2003) Essential steps in the ppGpp-dependent regulation of bacterial ribosomal RNA promoters can be explained by substrate competition. *J. Biol. Chem.* 278, 16834–16843.
- [9] Bratt, J.M., Zeki, A.A., Last, J.A. and Kenyon, N.J. (2011) Competitive metabolism of L-arginine: arginase as a therapeutic target in asthma. *J. Biomed. Res.* 25, 299–308.
- [10] Ulanovskaya, O.A., Zuhl, A.M. and Cravatt, B.F. (2013) NNMT promotes epigenetic remodeling in cancer by creating a metabolic methylation sink. *Nat. Chem. Biol.* 9 (5), 300–306.
- [11] Henri, V. (1902) Théorie générale de l'action de quelques diastases. *Compt. Rend. Hebd. Acad. Sci. Paris* 135, 916–919.
- [12] Henri, V. (1903) Lois générales de l'action des diastases, Herman.
- [13] Michaelis, L. and Menten, M. (1913) Die Kinetik der Invertinwirkung. *Biochem. Z.* 49, 333–369.
- [14] Chou, T.C. and Talaly, P. (1977) A simple generalized equation for the analysis of multiple inhibitions of Michaelis–Menten kinetic systems. *J. Biol. Chem.* 252, 6438–6442.
- [15] Dingerkus, V.L.S., Gaber, T.J., Helmbold, K., Bubenzer, S., Eis-ert, A., Sánchez, C.L. and Zepf, F.D. (2012) Acute tryptophan depletion in accordance with body weight: influx of amino acids across the blood–brain barrier. *J. Neural Transm.* 119, 1037–1045.
- [16] Segel, I.H. (1993) *Enzyme Kinetics*, John Wiley & Sons.
- [17] Cornish-Bowden, A. (2012) *Fundamentals of Enzyme Kinetics*, Wiley-VCH.
- [18] Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Men des, P. and Kummer, U. (2006) Copasi – a complex pathway simulator. *Bioinformatics* 22, 3067–3074.
- [19] Emery, J.A., Hermon, K., Hamid, N.K.A., Donald, J.A. and Turchini, G.M. (2013) Δ-6 Desaturase substrate competition: dietary linoleic acid (18:2n-6) has only trivial effects on α-linolenic acid (18:3n-3) bioconversion in the teleost rainbow trout. *PLoS One* 8, e57463.
- [20] Baks, T., Janssen, A.E.M. and Boom, R.M. (2006) A kinetic model to explain the maximum in α-amylase activity measurements in the presence of small carbohydrates. *Biotechnol. Bioeng.* 94, 431–440.
- [21] De Vos, D., Bruggeman, F.J., Westerhoff, H.V. and Bakker, B.M. (2011) How molecular competition influences fluxes in gene expression networks. *PLoS ONE* 6 (12), e28494.
- [22] Pomeroy, J.R. (2008) Uncovering mechanisms of bistability in biological systems. *Curr. Opin. Biotechnol.* 19, 381–388.
- [23] Venkatraman, L., Li, H., Dewey, C.F., White, J.K., Bhowmick, S.S., Yu, H. and Tucker-Kellogg, L. (2011) Steady states and dynamics of urokinase-mediated plasmin activation in silico and in vitro. *Biophys. J.* 101, 1825–1834.
- [24] Cascante, M., Canela, E.I. and Franco, R. (1990) Control analysis of systems having two steps catalyzed by the same protein molecule in unbranched chains. *Eur. J. Biochem.* 192, 369–371.
- [25] McKenna, M.C. (2012) Substrate competition studies demonstrate oxidative metabolism of glucose, glutamate, glutamine, lactate and 3-hydroxybutyrate in cortical astrocytes from rat brain. *Neurochem. Res.* 37, 2613–2626.

Quantitative Model of Cell Cycle Arrest and Cellular Senescence in Primary Human Fibroblasts

Sascha Schäuble^{1,2}, Karolin Klement^{3a}, Shiva Marthandan³, Sandra Münch^{3b}, Ines Heiland², Stefan Schuster², Peter Hemmerich³, Stephan Diekmann^{3*}

1 Research Group Theoretical Systems Biology, Friedrich-Schiller-University, Jena, Germany, **2** Department of Bioinformatics, Friedrich-Schiller-University, Jena, Germany, **3** FLI, Beutenbergstr. 11, Jena, Germany

Abstract

Primary human fibroblasts in tissue culture undergo a limited number of cell divisions before entering a non-replicative “senescent” state. At early population doublings (PD), fibroblasts are proliferation-competent displaying exponential growth. During further cell passaging, an increasing number of cells become cell cycle arrested and finally senescent. This transition from proliferating to senescent cells is driven by a number of endogenous and exogenous stress factors. Here, we have developed a new quantitative model for the stepwise transition from proliferating human fibroblasts (P) via reversibly cell cycle arrested (C) to irreversibly arrested senescent cells (S). In this model, the transition from P to C and to S is driven by a stress function γ and a cellular stress response function F which describes the time-delayed cellular response to experimentally induced irradiation stress. The application of this model based on senescence marker quantification at the single-cell level allowed to discriminate between the cellular states P, C, and S and delivers the transition rates between the P, C and S states for different human fibroblast cell types. Model-derived quantification unexpectedly revealed significant differences in the stress response of different fibroblast cell lines. Evaluating marker specificity, we found that SA- β -Gal is a good quantitative marker for cellular senescence in WI-38 and BJ cells, however much less so in MRC-5 cells. Furthermore we found that WI-38 cells are more sensitive to stress than BJ and MRC-5 cells. Thus, the explicit separation of stress induction from the cellular stress response, and the differentiation between three cellular states P, C and S allows for the first time to quantitatively assess the response of primary human fibroblasts towards endogenous and exogenous stress during cellular ageing.

Citation: Schäuble S, Klement K, Marthandan S, Münch S, Heiland I, et al. (2012) Quantitative Model of Cell Cycle Arrest and Cellular Senescence in Primary Human Fibroblasts. PLoS ONE 7(8): e42150. doi:10.1371/journal.pone.0042150

Editor: Elad Katz, University of Edinburgh, United Kingdom

Received: March 29, 2012; **Accepted:** July 2, 2012; **Published:** August 7, 2012

Copyright: © 2012 Schäuble et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by a grant from Bundesministerium fuer Bildung und Forschung (BMBF) (grant no. 0315581D). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: diekmann@fli-leibniz.de

^a Current address: Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Calgary, Calgary, Alberta, Canada

^b Current address: Medical Research Institute, Ninewells Hospital and Medical School, Dundee, Scotland, United Kingdom

Introduction

Ageing is an omnipresent process observed throughout all organisms, yet its fundamental driving forces remain unclear. Some aspects of ageing are believed to be recapitulated during cellular senescence of some types of primary mammalian cells in cell culture systems [1]. Notably, experimental clearance of cellular senescent cells in mice delays ageing-related pathologies in at least some tissues [2]. *In vitro*, cellular senescence manifests as a permanent cell cycle arrest resulting from the replicative exhaustion of cultured normal diploid cells [3]. Senescence also acts as an efficient tumor suppressor mechanism [4].

Although senescent cells are unable to proliferate, they are still viable and metabolically active, but resistant to mitogenic or apoptotic stimuli [5,6]. To identify senescent cells *in vitro* and *in vivo*, specific biomarkers have been established [7,8] reviewed in [9,10]. Senescent cells are characterized by an increased cell size associated with higher numbers of lysosomes, vacuoles and mitochondria, and major changes in the cytoskeleton (reviewed in [11]). Diagnostically important senescence markers include a high activity of senescence-associated β -galactosidase (SA- β -Gal)

[8], telomere dysfunction-induced foci (TIF) [12,13] as well as up-regulation of the cell cycle regulators p16, p21 and p53 [14,15]. Many, but not all primary human cell types develop senescence associated heterochromatin foci (SAHF) [7,16]. Furthermore, senescent cells express matrix-degrading proteases and inflammatory chemokines and cytokines, known as the senescence-messaging secretome (SMS) or senescence-associated secretory phenotype (SASP) [17–21].

Several mechanisms are essential for the induction of senescence including an intact DNA damage response and functional p53-p21, p16-pRb, and p38-MAPK pathways [9]. Telomere-dependent cellular senescence is induced by an increased DNA damage response (DDR)-activity at dysfunctional telomeres [12]. Uncapped telomeres focally accumulate high levels of phosphorylated H2AX and repair factors such as p-ATM, 53BP1, MDC1, and pNBS1 [12,13,22–24]. Activated ATM at dysfunctional telomeres induces the downstream effectors Chk1 and Chk2. These kinases activate p53, which induces transcription of p21 to promote cell cycle arrest [12,25,26], reviewed in [27–30]. While cells with damaged telomeres always activate the p53-p21-pathway, the p16-pRb-pathway alone is not sufficient to

trigger DNA damage-induced senescence [12]. DNA damaging interventions such as irradiation or drug treatment induce the same cascade of responses if DNA damage exceeds a certain threshold [15,31–39]. Overexpression of activated oncogenes, e.g. RAS, induces oncogene-induced senescence (OIS) [35,40,41], probably also through signaling from hyperproliferation-induced DNA damage sites, in this case at collapsed replication forks [35,42–44]. OIS appears to be mainly dependent on the p16-pRb-pathway [45]. Consistent with these observations, freshly isolated normal human epidermal fibroblasts are resistant to RAS-induced senescence because of their low or absent levels of p16 [46]. In general, cellular senescence is considered to be a stress-response program which can be activated by various stressors most prominently by excessive DNA damage and telomere shortening [47].

Cultured senescent cells can be easily identified by the absence of proliferation (Ki67-negativity) combined with the identification of DNA damage foci. Additionally, in particular when being combined with the two markers, SA- β -Gal activity is a good quantitative indicator of senescence [48]. SA- β -Gal activity is detectable at high levels in senescent, but not in pre-senescent or quiescent mammalian cells in culture [8]. Furthermore, the marker is undetectable in immortal cell lines, yet is induced when cells are growth-arrested by genetic manipulations [8]. Recently, we have identified accumulation of annexin A5 at the nuclear envelope as an additional robust and quantitative marker for cellular senescence [49].

With increased age, cell populations, either in tissues or in culture, will be a mixture of proliferating, cell cycle arrested and senescent cells [50,51]. Cell cycle inhibitors, in particular p21 and p16, diagnostically indicate the presence of arrested human fibroblast cells. On the other hand, robust senescence markers such as SA- β -Gal identify senescent cells. In parallel, growth curves of cell populations yield the information on total cell number so that at any time point, three cellular states can be estimated: proliferating, cell cycle arrested and senescent cells. This enables the establishment of quantitative computer models describing cellular growth and the transition to senescence. Here we present such a model, depict its properties and apply the new model to the analysis of growth rates. The cellular response to low dose irradiation (low stress) driving primary cells into cell cycle arrest without senescence, can be quantitatively described by the model. At high dose irradiation, higher population doublings and thus higher stress levels, cells become senescent, a transition also well described by the model. From the quantitative fit of the growth curves by our model, the population of the arrested C and senescent S states can be estimated and compared to experimental values. We also employed the model to quantitatively describe subtle differences in the cellular ageing process of different primary human fibroblasts.

Materials and Methods

Cell Lines

Primary human fibroblasts MRC-5 (primary cells, *Homo sapiens*, 14 weeks gestation male, fibroblasts from normal lung, normal diploid karyotype), WI-38 (primary cells, *Homo sapiens*, 3 months gestation female, fibroblasts from normal lung, normal diploid karyotype) and BJ (primary cells, *Homo sapiens*, newborn male, fibroblasts from normal foreskin, normal diploid karyotype), IMR-90 (primary cells, *Homo sapiens*, fibroblasts from normal lung, normal diploid karyotype) were obtained from ATCC (LGC Standards GmbH, Wesel, Germany). HFF (primary cells, *Homo sapiens*, fibroblasts from foreskin, normal diploid karyotype) cells

were kind gifts of T. Stamminger (University of Erlangen, Kronschnabl and Stamminger [52]).

Cell Culture

Cells were cultured as recommended by ATCC in Dulbeccos modified Eagles low glucose medium (DMEM) with L-glutamine (PAA Laboratories, Pasching, Austria), supplemented with 10% fetal bovine serum (FBS) (PAA Laboratories). Cells were grown under normal air conditions in a 9.5% CO₂ atmosphere at 37°C. For sub-culturing, the remaining medium was discarded and cells were washed in 1xPBS (pH 7.4) (PAA Laboratories) and detached using trypsin/EDTA (PAA Laboratories). Primary fibroblasts were sub-cultured in a 1:4 (= 2 PDs) or 1:2 (= 1 PD) ratio.

Immuno-fluorescence

Analysis of senescence (p21, p16) and heterochromatin markers (recognizing SAHFs) as well as the DNA damage indicator γ H2AX were determined by immuno-fluorescence at the single cell level. For this analysis in young and senescent cells, cells were fixed in 4% paraformaldehyde (in 1xPBS, pH 7.4) for 10 min. Cells were permeabilized using 0.25% Triton-X 100 (in 1xPBS, pH 7.4) for 3 min. Primary antibodies (anti-p21 (H-164, Santa Cruz), anti-p16 (BD Pharmingen), anti- γ H2AX (16–193, Upstate, USA)) were diluted in 1xPBS (pH 7.4) and incubated on the cells for 1 hr at RT. SAHFs (“dense DAPI regions”) were visualized either by DAPI staining or by immune-fluorescence using antibodies against H3K9me3 (ab8898, Abcam, Cambridge, UK) or H4K20me3 (ab9053, Abcam). The secondary fluorescently labeled antibodies were incubated for 1 hr at RT. The DNA was stained with 4'-6-diamidino-2-phenyl indole (DAPI, Invitrogen, Carlsbad, USA) by mounting the slides in Prolong gold antifade reagent (Invitrogen). Cells were imaged using a Zeiss LSM 510 or LSM 710 microscope with a 63x/1.40 oil immersion objective. Immuno-fluorescence results were obtained by determining the number of cells positive for the respective marker in relation to the total cell number. 100 cells, identified by DAPI-stained nuclei, were visually investigated by fluorescence microscopy in each experiment. Marker-positive cells were identified by detecting the specific fluorescence-labeled antibodies, cells without the respective marker did not show fluorescence. Mean values and standard deviations represent a minimum of three independent experiments.

Induction of Cellular Senescence

Cellular senescence was induced by γ -irradiation: human fibroblast cells were irradiated by ionizing radiation in a Gammacell GC40 (MDS Nordion, Ottawa, Canada) by ¹³⁷Cs as radioactive isotope. Exposure time was determined by correcting the irradiation dose of 1.23 Gy/min with decay time. Cells were seeded 1 day before exposure, irradiated at RT, and subsequently cultured at 37°C.

Detection of SA- β -galactosidase

The SA- β -galactosidase (SA- β -Gal) assay was performed as described by Dimri et al. [8]. Cells were washed in 1xPBS (pH 7.4) and fixed in 4% paraformaldehyde (pH 7.4), 10 min at RT. After washing the cells in 1xPBS (pH 7.4), staining solution consistent of 1 mg/ml X-Gal, 8 mM citric acid/sodium phosphate pH 6.0, 5 mM K₃Fe(CN)₆, 5 mM K₄Fe(CN)₆, 150 mM NaCl, 2 mM MgCl₂, was added. The enzymatic reaction occurred without CO₂ for 4–16 hours at 37°C. After incubation the cells were washed in 1xPBS (pH 7.4) and, in order to visualize cell nuclei, DNA and

SAHFs, mounted with DAPI-containing Prolong Gold antifade reagent (Invitrogen).

Model Simulation

We simulated our model of cell cycle arrest and cellular senescence as a system of ordinary differential equations using standard ODE solvers from the MATLAB environment (<http://www.mathworks.com>; Natick, USA). For simplicity, linear kinetics were chosen for the transition between different cell states. Experimental data were fitted by the model using MATLAB's `lsqnonlin` function. The quality of the fit was evaluated by solving the least squares optimization problem.

Model Adjustments for γ -irradiation Induced DNA Damage

After a stress situation, cells will always continue to proliferate if a single functional cell survived. We thus considered a cell concentration below one as growth termination criteria in our simulation. For high irradiation doses, cells are critically damaged and transformed into a senescent cell state, with the number of proliferating cells reduced below one. However, if cells are only slightly damaged by a low irradiation dose, proliferating cells survived and the cell population recovered.

Results

Model Deduction

Constant growth. A number of immortal cell lines show constant and stable growth (Figure 1). Freshly isolated primary fibroblasts share this feature during the early periods of their lifespan, although their growth rate can be lower. In a most simple approach, this situation can be described by a cell population solely consisting of proliferating cells (Figure 2A). With a population of healthy, proliferating cells, $P(t)$, constantly growing with time t , the time-dependent change of $P(t)$ depends on the constant growth factor r (Eq. 1):

$$\frac{dP(t)}{dt} = rP(t) \quad (1)$$

Eq. 1 results in exponential growth and yields in a linear graph when plotting population doublings (PD) versus time (Figure 1). Fitting this minimal model to experimental data on various rodent species from the literature [53] and our own data (HeLa cells), starting with a single cell, we derive unique division rates r for different cell lines which vary up to a factor of 7.8 between minimal and maximal growth rates (Table 1). Cells not only from different species but also cells from the same species but at different age (e.g. growth rates of adult versus embryonic squirrel

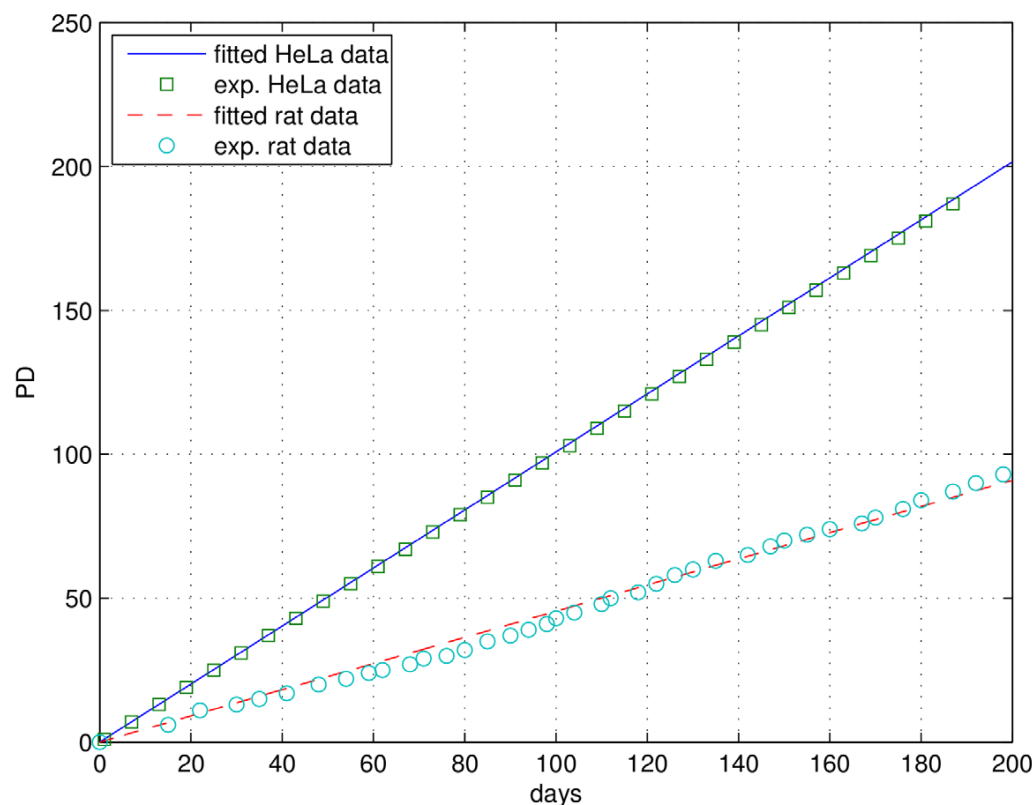


Figure 1. Fit of Eq. 1 to constant growth for HeLa (own data: green squares, fit: blue line) and rat fibroblast cells (data: blue circles [53], fit: red dashed line). See also Table 1.
doi:10.1371/journal.pone.0042150.g001

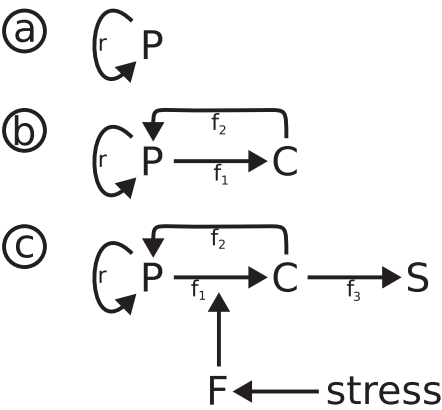


Figure 2. Model scheme. a, proliferation; b, extension with second cell state species; c, complete final model (upper part: P-C-S transitions; lower part: stress induction and cell response function F). doi:10.1371/journal.pone.0042150.g002

fibroblasts, see Table 1) show a significant difference in their unhampered constant growth speed. The fastest growth rate was measured for the cancer cell line HeLa which is supposed to be made possible by neglecting cellular maintenance [54–56].

Complete model. Mild stress can lead to short term reversible cell cycle arrest [57,58]. We quantitatively analyzed this effect by irradiating MRC-5 fibroblasts with 0.5 Gy, inducing low levels of DNA damage as indicated by increased numbers of γ H2AX DNA repair foci determined using immuno-fluorescence (Figure 3A). Within 16 hrs after irradiation, the number of p21-positive cells (determined by immuno-fluorescence) increased indicating short term cell cycle arrest (Figure 3B); within the following hours, the number of p21-positive cells decayed to background levels (73 hrs) indicating successful DNA repair and return into the cell cycle. This low dose irradiation did neither result in an increase of the number of p16-positive cells (investigated using immuno-fluorescence, Figure 3C) nor in the up-regulation of the cellular senescence marker SA- β -Gal

(percentage of SA- β -Gal positive cells, Figure 3D). The cell population continued to grow with a slight time lag, consistent with cell cycle re-entry after a transient cell cycle arrest (Figure 4A). In order to quantitatively describe this reversible cell cycle arrest, we introduced the additional cell cycle state “C” (Figure 2B) with the rate f_1 for the transition from proliferating cells P to cell cycle arrested cells C and the rate f_2 for the transition back into the cell cycle. This model can be described by the following equations:

$$\frac{dP(t)}{dt} = rP(t) - f_1P(t) + f_2C(t) \tag{2a}$$

$$\frac{dC(t)}{dt} = f_1P(t) - f_2C(t) \tag{2b}$$

After high dose irradiation, MRC-5 (20 Gy) (or WI-38 (15 Gy)) fibroblasts showed a very different response (Figure 3A–D). Immediately after irradiation, all cells had a highly elevated number of γ H2AX repair foci which decreased only slightly within the following 6 days (Figure 3A). The number of p21, and now also the number of p16-positive cells increased after irradiation, associated with a complete stop of cell proliferation and a constant increase of the number of SA- β -Gal-positive cells (Figure 3B, C, D). SA- β -Gal, a quantitative marker of cellular senescence [8,48], indicates the presence of an increased number of cells irreversibly arrested in the cell cycle. This senescent phenotype extends our model by an additional state “S” which is populated by rate f_3 from the C state (Eq. 3c); S cannot be derived directly from the P state, nor can it return into the C or P state (no back reaction, Figure 2C, upper part).

The simple model (Eq. 2a, b) with time-independent parameters f_1 and f_2 describes an immediate response to mild stress. However, we only observed an immediate response for the appearance of γ H2AX repair foci (Figure 3A). For other markers, we observed a slightly delayed response, with p21 levels reaching maximal levels about 16 hrs after irradiation when γ H2AX DNA damage foci already had declined close to background levels (Figure 3A, B). We thus have to consider a delayed cellular response to stress. Furthermore, the transition from reversibly (C) to irreversibly (S) arrested cells depends on the level of irradiation, or, in general, on the stress level applied to the cells. Thus, the P-C-S transition model presented in Figure 2C (upper part) must be extended to include a stress factor and a cellular stress response function F, sensing this difference: F allows for continued population growth under mild stress, but emphasizes the transition to the senescent phenotype when faced with harsh stress, resulting ultimately in replicative senescence (Figure 2C). This model can be described by the equations:

$$\frac{dP(t)}{dt} = rP(t) - f_1F(t)P(t) + f_2C(t) \tag{3a}$$

$$\frac{dC(t)}{dt} = f_1F(t)P(t) - f_2C(t) - f_3C(t) \tag{3b}$$

$$\frac{dS(t)}{dt} = f_3C(t) \tag{3c}$$

Table 1. Constant growth, model parameter r fitted to experimental data.

Species	Parameter r
House mouse	0.28
Embryonic squirrel	0.20
Squirrel	0.09
Naked mole rat	0.16
Gerbil	0.25
Chinchilla	0.13
Rat	0.32
Chipmunk	0.13
Guinea pig	0.26
Muskrat	0.15
Hamster	0.28
HeLa	0.70

Experimental data from the literature (rodent species [53]) and our own data (HeLa cells). doi:10.1371/journal.pone.0042150.t001

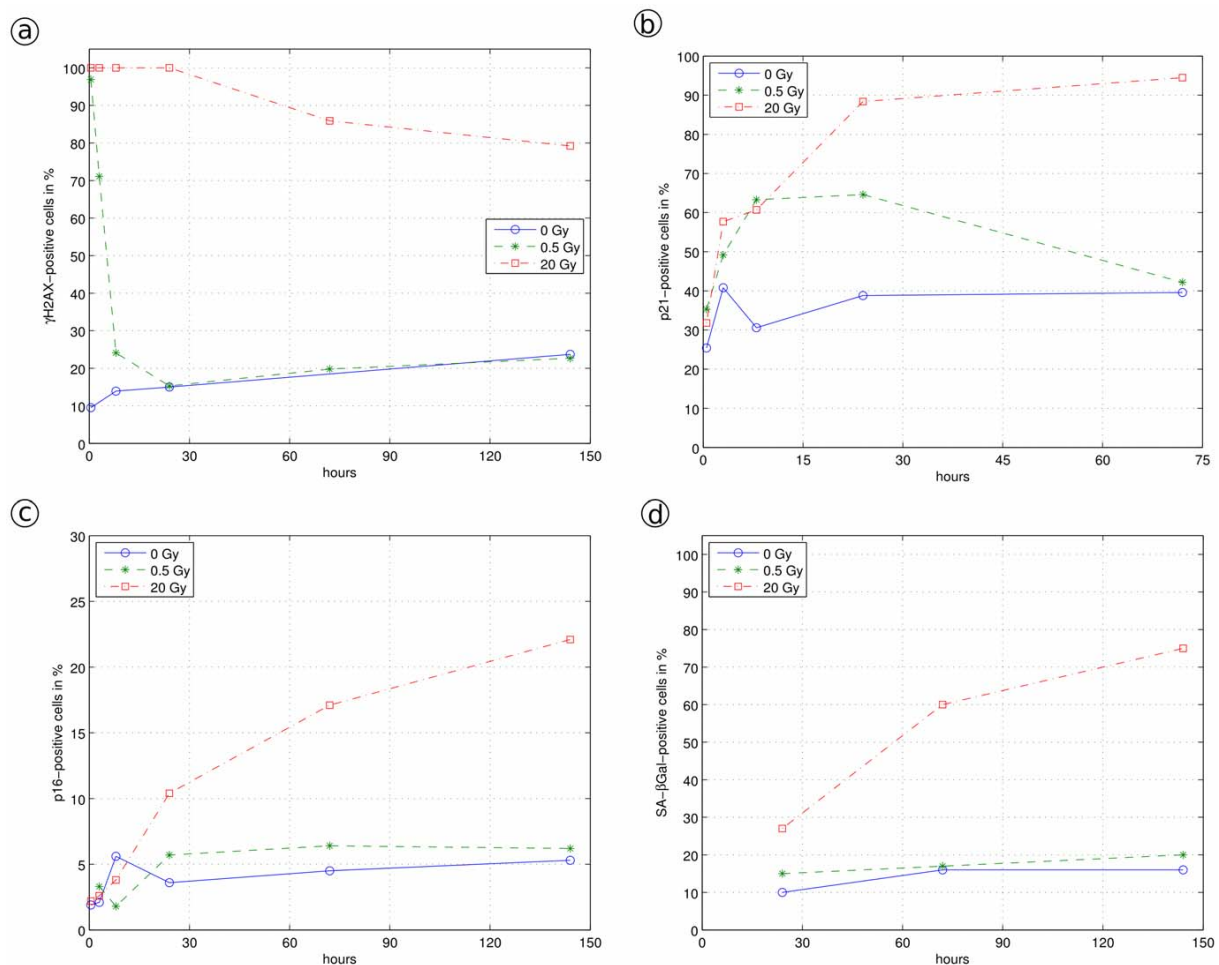


Figure 3. Relative number MRC-5 fibroblast cells positive for cellular markers after irradiation by the doses 0, 0.5 and 20 Gy. MRC-5. a, DNA damage marker γ H2AX; b, cell cycle arrest marker p21; c, cycle arrest marker p16; d, senescence marker SA- β -Gal. The experimental error in such experiments is less than $\pm 5\%$. doi:10.1371/journal.pone.0042150.g003

$$\frac{dF(t)}{dt} = \gamma F(t) \left(1 - \frac{F(t)}{K} \right) - \frac{F(t)^2}{1 + F(t)^2} \quad (3d)$$

The stress response function $F(t)$ (Eq. 3a, b, d), multiplied by the stress expression γ , describes the influence of cellular stress phenomena, such as exposure to irradiation, oxidative stress, telomere shortening, etc. While the external stress itself can be short, function F describes the cellular response, which can be delayed relative to the various stresses: function F can “remember” that a stress event has occurred, as its value might not decrease rapidly, once the stress itself has disappeared.

Function F , adapted from Ludwig et al. [59] was chosen to have a bi-stable behavior (Eq. 3d): below a certain stress threshold, the F value is low (resembling a healthy cell), but F rapidly turns to high values once this threshold is exceeded (resembling an endangered cell). Once the stress function γ dropped below threshold values, F decreases slowly, describing a slow return to the normal cell state. Indeed, a bifurcation analysis confirmed that function F , chosen in this form, is bi-stable and, depending on parameter values, switches between two states (Supplement S1). F is defined by two parameters,

K and γ . Whereas K is the maximal value F can reach, γ denotes various forms of stress, ranging from sudden stress such as irradiation pulses, to accumulating stress resulting e.g. from telomere shortening or non-physiological oxygen concentrations which represent long term effects. When irradiation is applied, $\gamma > 0$ for the time cells are exposed to irradiation, and $\gamma = 0$ otherwise. In contrast, when simulating long-term stress, we substituted a slowly increasing function for γ in order to resemble accumulating effects (see below), according to Equation 4:

$$\frac{d\gamma(t)}{dt} = \alpha + \beta\gamma(t), \gamma_0 = 0.001 \quad (4)$$

Introducing a stress (γ) and a response (F) function separately, offers the advantage that the effect of any kind of stress and the cell-specific response to this stress can be modeled individually or in selected combinations, as large variations had been observed [60], and cells from different tissue origin might respond differently [46]. Here we analyzed cells from the same tissue source (lung; MRC-5, WI-38), the same donor age (MRC-5, WI-38, BJ) and the same/different gender (male: MRC-5, BJ; female: WI-38).

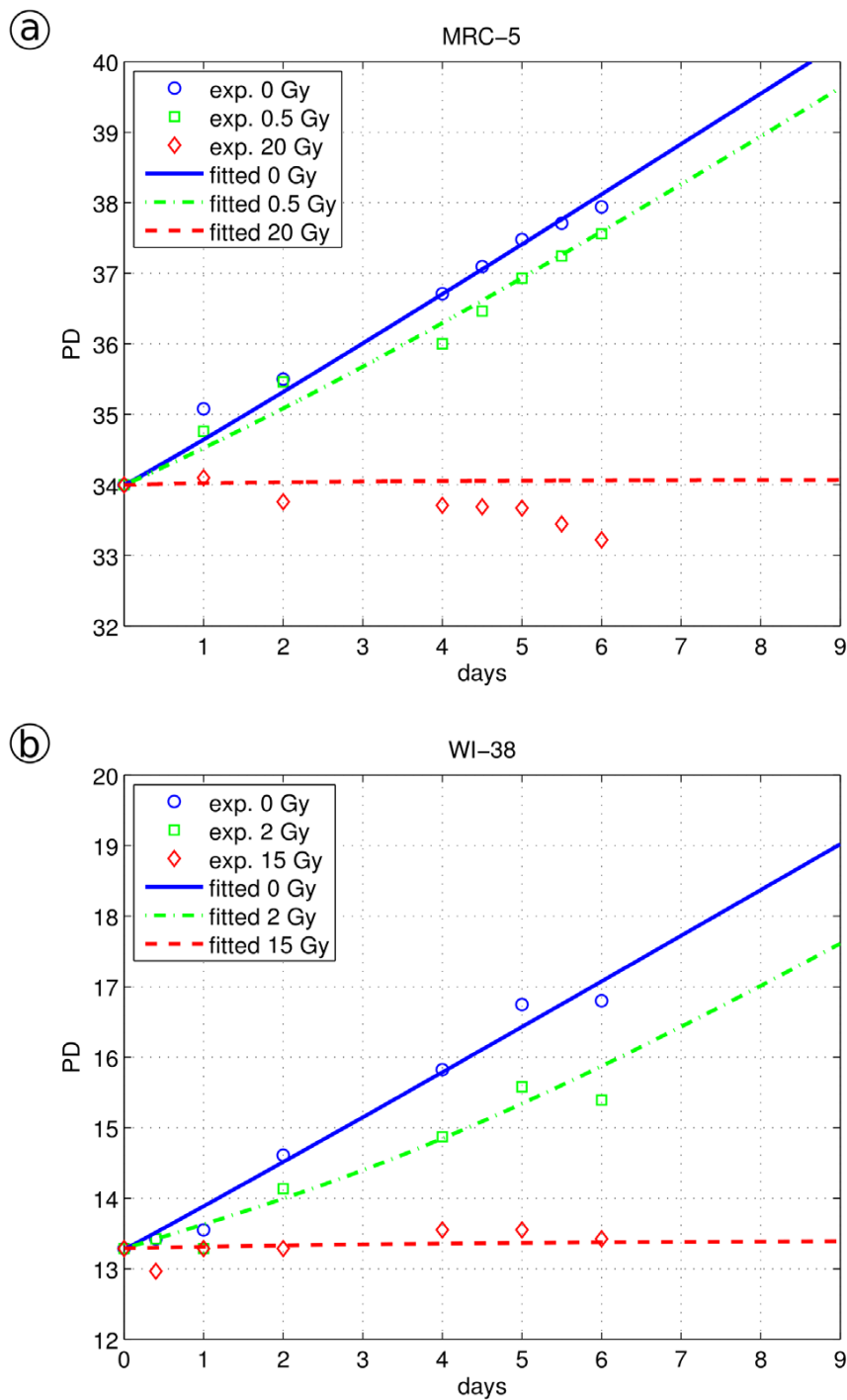


Figure 4. PD curves of human fibroblasts. Model fitting of different radiation doses and experimentally derived PDs. The data were fitted using the same parameter set for all radiation doses (differing only in the applied amount of irradiation time) applying the model described by Eq. 3a–d: a, MRC-5 fibroblast, different radiation doses 0, 0.5 and 20 Gy; b, WI-38 fibroblast, radiation doses 0, 2 and 15 Gy. doi:10.1371/journal.pone.0042150.g004

Sensitivity analysis. We investigated the sensitivity of the model parameter values regarding the resulting PD curves and cell state distributions in the culture (Supplement S2). The

analysis clearly showed that our model is most sensitive to the variation of the growth rate r (resembling its particular importance for the maximal replicative capacity of the fibroblast

cells) and much less sensitive to the other parameter f_{1-3} , K , α and β . Interestingly, WI-38 cells are notably less sensitive to f_2 and f_3 variation than BJ cells.

Model Application to Differently Stressed Cells

Irradiation induced DNA damage in human fibroblast WI-38 cells with 0, 2, or 15 Gy. Primary human WI-38 fibroblasts were γ -irradiated with 2 Gy or 15 Gy inducing low or high levels of DNA damage, respectively. The cellular stress response was analyzed by immunofluorescence before and 0.5 hours, 1 day and 6 days after irradiation [61]. Immediately after irradiation, we observed increased γ H2AX levels, which, after low irradiation, decayed to background levels within one day, whereas they decayed to background levels only after three days when irradiated with a high dose. One and 6 days after irradiation we noticed an increase in p21 levels which was only slight after low, but strong after high levels of irradiation, while p16 levels increased only after 6 days and high dose irradiation. The number of SA- β -Gal positive cells within the population increased from low levels ($13 \pm 3\%$) for untreated cells to $30 \pm 1\%$ and $67 \pm 6\%$ after 6 days for 2 Gy and 15 Gy irradiated cells, respectively, as revealed by the SA- β -Gal staining. Compared to untreated cells, low level irradiation delayed growth but cells continued to proliferate after about one day (Figure 4B). In contrast, high irradiation levels resulted in severe DNA damage so that cells stopped proliferation irreversibly (Figure 4B).

Applying our model for stress-induced cell cycle arrest and senescence (Figure 2C), we simulated the irradiation experiment by selecting a value for γ in the range of values $700 < \gamma < 6,000$. The value of γ is crucial for the quantitative description of the low irradiation growth curve. Without irradiation, we apply here $\gamma = 0$, and for strong irradiation, γ exceeds a threshold value and is not well-defined above. For $\gamma < 700$, we observe that cells are able to recover even from strong irradiation, and cells under mild irradiation are only insignificantly hampered in growth, indicating that its value is set too low. In contrast, for $\gamma > 6,000$, our simulation does not allow cells suffering from mild irradiation to recover, showing that such high γ values cannot resemble sensitivity of WI-38 cells to irradiation. Consequently, setting $\gamma = 2,800$ arbitrary units for the irradiation time (zero for no irradiation, 108 sec for 2 Gy, 13.5 min for 15 Gy), we were able to fit one set of model parameters (r , f_1 , f_2 and f_3) to the three different measured growth curves. For the values $r = 0.46$, $f_1 = 3.8$, $f_2 = 16$ and $f_3 = 0.26$ and three different irradiation times for γ , we obtained a convincing description of the experimental behavior (see Figure 4B).

Irradiation induced DNA damage in human fibroblast MRC-5 cells with 0, 0.5, or 20 Gy. A similar irradiation experiment was conducted using MRC-5 cells, γ -irradiated with 0.5 Gy or 20 Gy, as described above (see Figure 3A–D). Applying our model, we successfully fitted the parameter set $r = 0.51$, $f_1 = 4.0$, $f_2 = 16$ and $f_3 = 0.42$ to the measured growth curves (Figure 4A). In order to resemble irradiation-induced damage, γ was selected within the range of values $800 < \gamma < 29,000$. For $\gamma < 800$, we again observed that mildly stressed cells grew almost without delay and that cells suffering from strong irradiation were still able to recover. For $\gamma > 29,000$, we observed that even low dose irradiated cells were unable to recover, indicating again that such high γ values are outside of the relevant range. We set $\gamma = 0$ for no irradiation, $\gamma = 8,000$ over 27 sec for 0.5 Gy, and over 18 min for 20 Gy. Choosing the same γ value for the time of irradiation as for WI-38 fibroblasts ($\gamma = 2,800$), we were unable to explain the MRC-5 data. Although we could fit well the 20 Gy irradiation data with $\gamma = 2,800$, the mildly irradiated MRC-5 cells

appeared to be not sufficiently stressed, resulting in a growth rate almost as high as cells that were not irradiated. Nevertheless, both values of γ indicate a rapid change of F to the endangered state, triggering subsequent cell cycle arrest. The different γ values suggest that WI-38 fibroblasts are considerably more sensitive to (this) stress compared to MRC-5 cells (see also [60]), consistent with a lower γ value for the exposure time. Choosing a similar sensitivity for MRC-5 cells (γ values similar to WI-38 levels), would reduce the quality of the fitted curves considerably.

“S” feedback on proliferation. Another hallmark of cellular senescence is the formation of a senescence associated secretory phenotype (SASP) which is marked by the secretion of inflammatory cytokines [20,21,62,63]. The secretion of these factors is mainly dependent on an increased and persistent DNA damage response and is therefore specific for senescent cells [64]. These secreted factors might have an effect on the growth behavior of the cell population. To investigate the effect of SASP secreted from “S” cells on the growth behavior (rate r) of proliferating “P” cells and their transition to “C” cells (f_1 , f_2), early passage MRC-5 cells were cultured in medium harvested from cells which were senescent for at least one week. This senescence-conditioned medium was then constantly kept on the young cells and replaced with fresh senescence-conditioned medium every week. In this preliminary experiment, young cells did not show impaired proliferation compared to control cells. Thus, senescence-conditioned medium did not induce senescence in young passage fibroblasts. Only at very late passages, cells cultured with the senescence-conditioned medium tend to senesce approximately two population doublings earlier than control cells. In a more extensive study, influences of short cytokine half-lives or strong dilution effects should be excluded. Nevertheless, our preliminary data indicate that the presence of “S” cells seem to influence the growth of the whole cell population in its entirety only slightly (“S” seems to have little influence on r , f_1 and f_2).

Replicative senescence induced by 20% oxygen and telomere shortening. When human fibroblast WI-38, BJ and MRC-5 cells continue to grow, they stop proliferation when reaching their Hayflick limit [3]. This state of replicative senescence may be determined by both, cell-specific stress responses (due to for example constant oxidative stress) and telomere shortening [65]. The growth behavior and the number of cells with up-regulated markers p21, p16, SA- β -Gal and SAHF were determined for MRC-5, WI-38 and BJ cell lines [49]. Immuno-fluorescence analyses of all three cell lines showed, after a certain time in culture, an elevated number of cells with higher levels of p16, p21 and SA- β -Gal. Our studies additionally revealed that senescent cells accumulated persistent DNA damage, as shown by increased γ H2AX levels (data not shown) and loss of the proliferation marker Ki67 [49]. WI-38 cells showed a low number of cells with up-regulated markers for p16, p21, SA- β -Gal and SAHFs at PD 40 but high numbers at PD 60 (here, all cells were SA- β -Gal positive), with very little delay between the various markers indicating an immediate transition from proliferation to senescence (see Figure 5B). To quantitatively describe these data, we substituted γ according to Equation 4 using $\alpha = 0.002$ and $\beta = 0.028$ for the description of the long term accumulating stress. With the fitted values $r = 0.47$, $f_1 = 0.40$, $f_2 = 0.43$ and $f_3 = 0.66$, we obtained a quantitative fit of the WI-38 growth curve (Figure 5A) and the marker increase (Figure 5B).

For BJ fibroblast cells at higher PDs, the timing of the up-regulation of senescence markers varied strongly: p21 levels started to increase already at PD 30 while the number of SA- β -Gal positive cells started to increase at PD 50 (Figure 6B). Even at high PD values, BJ cells did not show SAHF formation. Highly

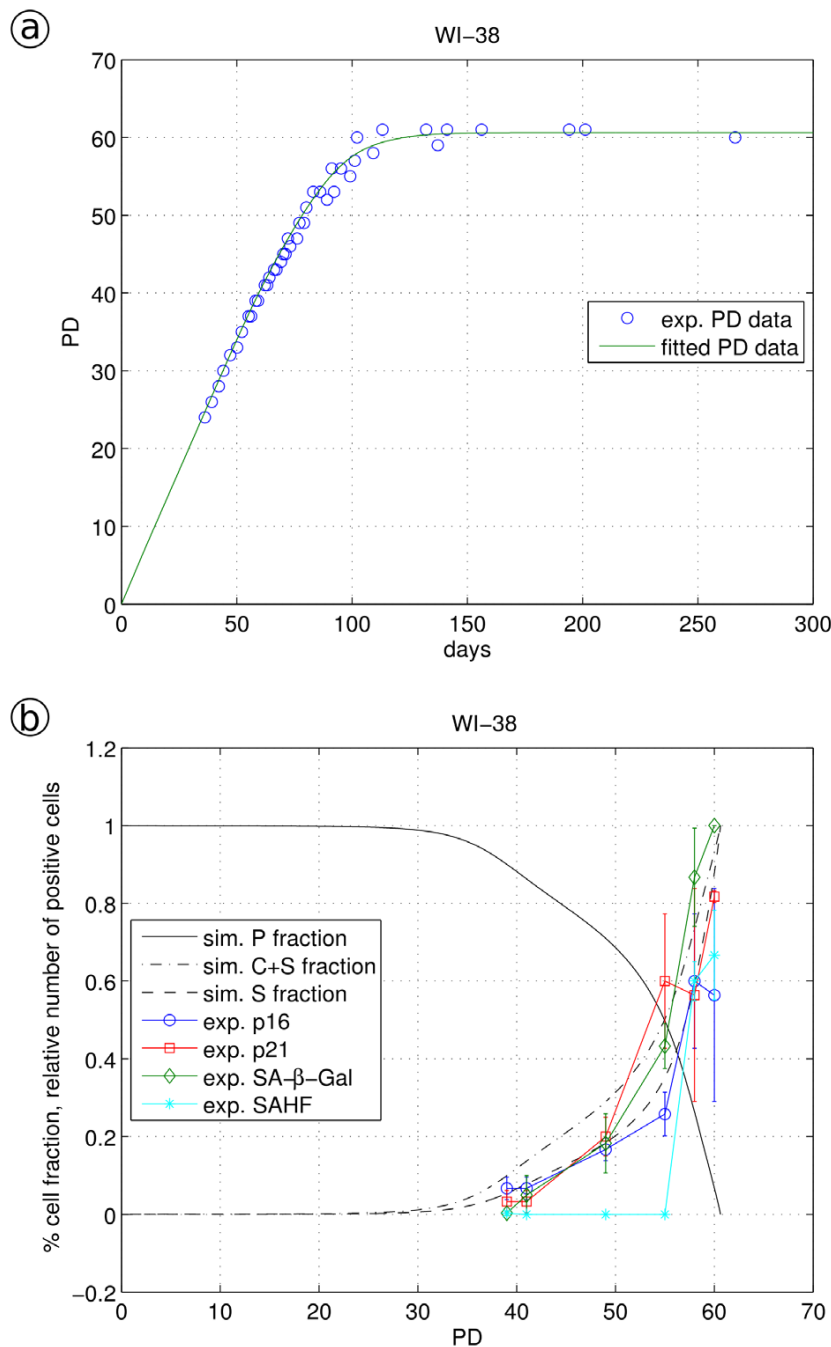


Figure 5. Simulation of WI-38 fibroblast data. a, Experimental growth data (circled) were fitted by model Eq. 3a–d using Eq. 4 as an expression for monotonically increasing stress γ ; b, the fraction of proliferating cells P, cells showing a cell cycle arrest or a senescent phenotype (C+S) and solely the fraction of senescent cells S are shown together with the appearance of biomarkers. Biomarker values (p16, p21, SA-β-Gal and SAHF) were measured by immune-fluorescence as number of positive cells [49].
doi:10.1371/journal.pone.0042150.g005

compacted heterochromatin foci, termed senescence associated heterochromatin foci (SAHF), develop during senescence in many, but, as also shown here, not all primary human cell types [7,16]. Applying γ (Eq. 4) with $\alpha = 0.001$ and $\beta = 0.016$ for the description of long-term accumulating stress and the fitted values $r = 0.30$,

$f_1 = 1.1$, $f_2 = 2.45$ and $f_3 = 0.27$ we obtained a quantitative fit of the BJ growth curve (Figure 6A) with a good fit for p21 and SA-β-Gal increase (Figure 6B), in particular at higher PD values. Compared to WI-38 cells, BJ cells showed a much broader transition from partial cell cycle arrest to senescence, indicated by the increase of

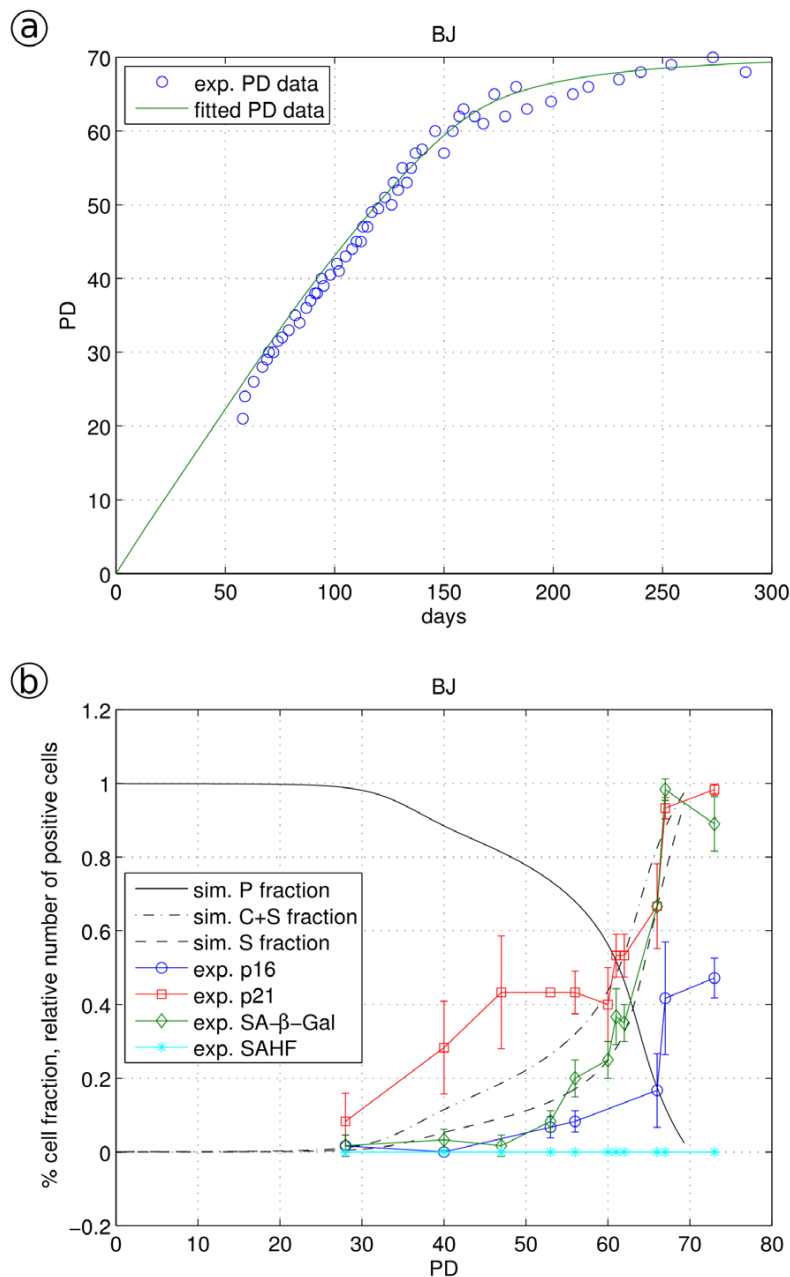


Figure 6. Simulation of BJ fibroblast data. a, experimental data (circled) were fitted by model Eq. 3a–d using Eq. 4 as an expression for monotonically increasing stress γ ; b, the fraction of proliferating cells P, cells showing a cell cycle arrest or a senescent phenotype (C+S) and solely the fraction of senescent cells S are shown together with the appearance of biomarkers. Biomarker values (p16, p21, SA-β-Gal and SAHF) were measured by immune-fluorescence as number of positive cells [49].
doi:10.1371/journal.pone.0042150.g006

ageing markers and growth curves. We noticed for BJ cells other than in WI-38 cells, the number of p21 up-regulated cells is not a quantitative indicator of cell cycle arrested C cells: between PDs 30 and 55, this number is considerably higher than the calculated number of C cells (Figure 6B). Indeed, cell cycle arrest is regulated by CDKs (inhibited by p21); thus, p21 is only an indirect marker of cell cycle arrest. p21 seems to titrate CDKs differently in different fibroblasts.

This is consistent with our observations in MRC-5 cells (Figure 7A, B). The up-regulation of senescence markers started already at low PDs: the number of p16 and SA-β-Gal positive cells started to increase at PD 48 (Figure 7B), immediately followed by SAHF formation. Applying γ (Eq. 4) with $\alpha=0.004$ and $\beta=0.0042$ and the fitted values $r=0.489$, $f_1=6.64$, $f_2=5.99$ and $f_3=0.26$, we obtained a quantitative fit of the MRC-5 growth curve (Figure 7A) with a good fit for p21 at low PDs and an

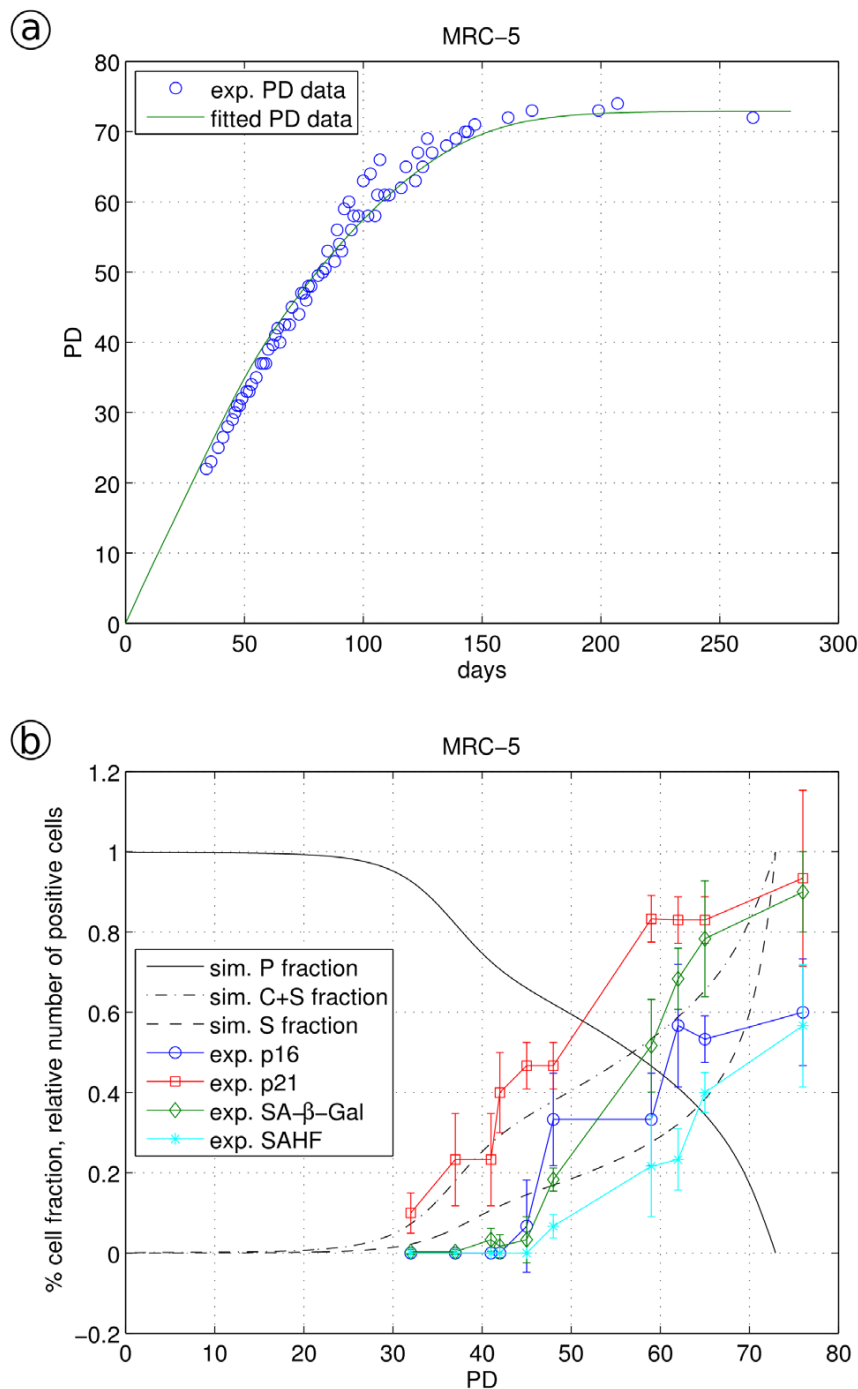


Figure 7. Simulation of MRC-5 fibroblast data. a, experimental data (circled) were fitted by model Eq. 3a–d using Eq. 4 as an expression for monotonically increasing stress γ ; b, the fraction of proliferating cells P, cells showing a cell cycle arrest or a senescent phenotype (C+S) and solely the fraction of senescent cells S are shown together with the appearance of biomarkers. Biomarker values (p16, p21, SA- β -Gal and SAHF) were measured by immune-fluorescence as number of positive cells [49].
doi:10.1371/journal.pone.0042150.g007

unsatisfying fit at higher PD values. Other than for WI-38 and BJ cells, the number of SA- β -Gal positive MRC-5 cells increased considerably while the cells still continued to grow (with only a

slightly reduced rate). We were not able to find a set of model parameters which quantitatively describes both, the marker up-regulation as S cells as well as the growth curve. This suggests that

in MRC-5 cells SA- β -Gal might increase not solely due to the cell transition into cellular senescence but additionally also due to other reasons, a finding consistent with a previous report [66]. Thus, identified by our model, SA- β -Gal seems to be a good quantitative senescence marker in WI-38 and BJ cells but not in MRC-5 cells.

Although applying similar irradiation stress to WI-38 and MRC-5 cells, the cellular response is described by a roughly 3-fold lower γ value for WI-38 compared to MRC-5 cells. Obviously, WI-38 cells are more sensitive to irradiation than MRC-5 cells. This corresponds to WI-38 cells showing a rather sharp transition to senescence at low PD values. In contrast, BJ and MRC-5 fibroblasts undergo a considerably smoother transition towards cellular senescence at higher PD values. When in these three fibroblast cell lines, reactive oxygen species (ROS) production and telomere shortening accumulate to a similar extent, then the early and fast transition into senescence would indicate WI-38 cells indeed being more sensitive compared to the other two cell lines, consistent with published reports [60]. This agrees with results from the analysis of the f_1 , f_2 and f_3 parameter values and the γ expression influenced by α and β : for BJ cells, f_1 and f_2 are notably larger than for WI-38 cells, resembling a relatively high flux towards C cells and back to proliferating P cells. We interpret this as a higher maintenance workload for BJ than for WI-38 cells. In contrast, f_3 is more than 2-fold smaller for BJ cells than for WI-38 cells resulting in a considerable delay for the BJ population to build-up a significant senescence cell fraction.

Influence from tissue origin and age. The observed cellular proliferation and the transition into senescence can vary depending on the tissue of origin and the age of the source tissue. Indeed, our data clearly indicate that for example cells from the same tissue source and the same donor age, but different gender (MRC-5 and WI-38 cell-lines) show clear differences in growth curves and transition to senescence. In order to determine if these observed differences are significant, we measured the experimental variance of our data by comparing the growth curves and the transition into senescence of primary human fibroblasts from the same tissue source (lung, MRC-5, WI-38), same donor age (MRC-5, WI-38 and BJ) and same or different gender (male: MRC-5, BJ; female: WI-38) as biological and technical replicates. When two aliquots of the same cell line but from different (commercial) sources were analyzed in parallel (biological replicates), we noticed small but significant differences between the growth curves of biological replicates for BJ, HFF, MRC-5 and WI-38 cells, but not IMR-90 cells. These small growth curve differences between two biological replicates, when fitted by our model, yielded in variations of transition rates between proliferating and cell cycle arrested cells (model parameters f_1 and f_2). In contrast, when cell-lines were separated into technical triplicates at early PD and analyzed in parallel, we observed no differences in their growth behavior: technical replicates showed excellent quantitative agreement between their growth curves yielding in identical model parameter values. Thus, the differences between the five human fibroblast cell-lines described above are significant.

Reduced growth rates can be explained by the presence of C and S cells with high r values. Experimentally determined cellular markers and our model identify the presence of cell cycle arrested and senescent cells in cell populations, as reported by others [60]. Thus, population growth might be determined by a constant (maximal) growth rate r for the proliferating P cells in the population, while the cell cycle arrested C and S cells do not contribute, i.e. the overall growth rate of the whole population is reduced relative to the maximal r value since it is a mixture of proliferating and non-proliferating cells. Fastest growth ($r = 0.7$,

see Table 1) was observed for cancerous HeLa cells assumed not to be involved in cellular maintenance. We thus asked if we could explain the various growth rates presented in Table 1 by our model (Eq. 2a-b), now keeping r constant ($r = r_{\max}$ addressing the high value of $r_{\max} = 0.7$ observed for HeLa cells) and fitting the transfer rate f_1 . Indeed, we obtained constant growth fitting the experimental values by a constant value for r and different values for f_1 , as given in Table 2. Thus, the observed variation in population growth could be explained by a constant high growth rate for the proliferating cells and for example by a varying amount of cell cycle arrested cells, which are assumed to be involved in cellular maintenance functions. This agrees with published observations on oxidative damage, for example for human fibroblast IMR-90 cells which divide faster under low level (3%) oxygen, with a doubling time 4–20% shorter, than for cells cultured under high (20%) oxygen, inducing more damage and thus requiring more maintenance [67]. However, different growth rates could also be explained by alternative cellular processes. The analysis of the biochemical origin of growth rate variation requires additional studies.

Proliferation at constant stress depends on r relative to f_1 , $f_3/f_1 + f_2$. Finally, we analytically investigated our model assuming a constant stress level (see Supplement S3), which is given, for instance, for constantly high (20%) oxygen levels. Our analysis shows that for $r \leq f_1$, $f_3/(f_2 + f_3)$ the population is not able to grow, as the growth factor r is too small to overcome the cell cycle arresting and senescence effects. f_1 and f_2 have opposing effects: low f_1 and high f_2 promote cell growth, while high f_1 and low f_2 values result in senescence. While low f_3 values result in a growing population, even high values of f_3 can be overcome by sufficiently high proliferation rates r when the stress response function F is sufficiently small. Thus, regardless of the transition rate from C to S cells, the population can still grow under high C to S transitions as long as the C fraction of the population is low (mediated by low f_1 and high f_2 values). High f_3 values could be important for populations living in environments that show high risks for rapid stress promoting events, where a fast transition into a senescent phenotype might be necessary for survival. In contrast, cells that rarely encounter dangerous events can afford rather low f_3 values, as cell cycle arrested cells might well be rescued.

Table 2. Values of the parameter f_1 in the model extension.

Species	Parameter f_1
House mouse	2.00
Embryonic squirrel	3.02
Squirrel	8.31
Naked mole rat	3.83
Gerbil	2.28
Chinchilla	5.01
Rat	1.61
Chipmunk	4.84
Guinea pig	2.11
Muskrat	4.12
Hamster	1.88

f_1 values for model extension with cell cycle arrest species. Same experimental data as in Table 1. For all simulations parameter $r = r_{\max}$ is set to 0.70 determined for HeLa cell growth rate and parameter f_2 to 1.
doi:10.1371/journal.pone.0042150.t002

Discussion

DNA damage is detected by cellular checkpoints. Checkpoints use a signaling mechanism either to stall the cell cycle until the damage has been repaired successfully or, if repair is unsuccessful, to target the cell for destruction via apoptosis. Under strong stress, some cell types cannot return to proliferation and enter the senescence state [47,68,69]. Here we present a quantitative model to simulate transient cell cycle arrest and cellular senescence. Next to the proliferating state P, we introduce a cell cycle arrested state C which is populated by rate f_1 and returns to proliferation by rate f_2 . Once arrested, cells can become senescent (state S) with rate f_3 . In our model, the transition to the senescent state is considered as being irreversible (no back reaction). To some extent, the arrested state C is experimentally identified by the cell cycle inhibition marker p21, and the senescent state S by SA- β -Gal (and partly by p16). Our model combines these values of the different cellular states with the growth curves of these cells, quantitatively describing these.

Since the cellular response to stress is delayed, we introduced a stress response function F describing this behavior, as analogously used by [59]. F was designed to allow for a switch-like behavior distinguishing reversible cell cycle arrest as a consequence of low stress from irreversible arrest due to high or long-term stress. In its switched state, F cannot change back immediately after the stress source has decayed. Its nonlinear form is necessary to model two possible and well distinguishable steady states, describing here proliferating and strongly stressed cells. Moreover, we required F to enable a switch-like behavior from one state into the other in a rapid sigmoidal manner, once a critical point is exceeded. Switch-like behavior can be modeled also by other formulations (like, for example, by double negative feedback loops), however, choosing F in the form presented here offers the advantage of convenient analytical properties combined with a minimal amount of free parameters.

The stress itself can either be permanent (e.g. constantly high levels of oxygen), transient (due to, for example, induction of DNA damage by γ -irradiation pulses) or might increase with time (like telomere shortening). We therefore introduced a stress induction factor of the form $\gamma = \alpha + \beta t$ with a time-independent influence α and a time-dependent contribution βt .

A rigorous analytical analysis of the model for constant stress resulted in a relation between growth rate r and the transition rate quotient $f_1 f_3 / (f_2 + f_3)$ which determines the fate of the cell population: with r smaller than the quotient, proliferation will stop while the population will continue to grow for $r > f_1 f_3 / (f_2 + f_3)$.

We compared cellular data obtained for different stress levels in three different human fibroblast cell types. First, we induced two different levels of DNA damage by short and long term γ -irradiation, and successfully fitted our model by a single parameter set to the three growth curves. Then, we successfully fitted our model to cells growing into replicative senescence, regarded here to be mainly induced by oxidative stress and telomere shortening. Again, our model is able to explain the observed behavior by quantitatively describing the growth curves.

Applying our model to the experimental growth curves, we detected response differences between the WI-38 and the BJ or MRC-5 cell lines, consistent with published data on fibroblast line-specific properties [46,60]. Such differences are not obvious from inspecting the experimental data, however, it could be well detected by our quantitative model analysis. WI-38 cells seem to be more sensitive to stress compared to MRC-5 cells since they reacted to smaller γ stress values. This is combined with a higher flux between P and C cells and a low transition into S cells for BJ

cells, compared to lower transitions between P and C combined with a higher flux to S for WI-38 cells. This suggests considerable maintenance workload in BJ cells, which might explain its higher resistance to stress. Consistently, WI-38 cells show a faster transition to senescence at lower PD values than BJ and MRC-5 cells.

The observed differences in C (seen in BJ and MRC-5) and S cell fractions (seen in MRC-5) compared to p21 and SA- β -Gal marker levels, respectively, can be explained by the qualitative nature of these markers. It cannot be ruled out that their up-regulation has additional side-effects, which ultimately influence the growth behavior of the respective cell line. Therefore, our work highlights the need for further exploration of more specific cell cycle arrest- and senescence-specific biomarkers [9,10].

Earlier, the influence of telomere shortening on cell proliferation was analyzed in network models of cell senescence [70–74]. Sozou and Kirkwood [72] applied a stochastic model for human diploid fibroblasts in which telomere reduction, oxidative stress linked to mitochondrial damage and nuclear somatic mutations were considered. Their model resulted in simulations that were in good agreement with data on intra-clonal variability in cell doubling potential published by [75]. Modeling the influence of telomere length on cellular senescence, Golubev et al. [73] conclude that telomere length decrease is a correlate of cell proliferation that, however, cannot alone account for senescence, instead also free radical damage influences have to be taken into account, consistent with statements by [72]. Portugal et al. [74] presented a stochastic growth model based on cell divisions in each time interval being a random process the probability of which decreases linearly with telomere shortening. The authors observed a good approximation of the qualitative growth of cultured human mesenchymal stem cells. In these models, theoretical parameters were not fitted to experimental data, in particular not to cellular growth curves. Lawless et al. [48] presented an analytical model fitting two cellular states, proliferation and senescence, to human fibroblast growth curves. Their model was successfully used to evaluate markers of cellular senescence. However, the model does not consider the intermediate state of transiently cell cycle arrested cells. When faced with mild stress, we observed here that a portion of the cells entered a temporary and reversible cell cycle arrest and not a senescence state, as indicated by the lack of SA- β -Gal up-regulation (see Figure 3D). In a series of sophisticated analyses, B. Novak, J. J. Tyson and collaborators modeled the protein interaction network for the regulation of DNA synthesis and mitosis [76,77]. Their approach provides a theoretical framework for the understanding of cell cycle regulation and presents increasingly complex models of the networks controlling cell growth and division. However, these authors did not model the cellular transition into senescence. Cellular senescence is maintained and reinforced by a DNA damage-ROS production feedback loop [78]. Passos et al. [78] presented a biochemically detailed stochastic model for this feedback loop on a single cell basis. Applying this model to our data, we were able to qualitatively simulate the cellular response to low as well as high irradiation. However, this model does not quantitatively compare cellular response to growth curve, a strength of our model.

In addition to entering the cell cycle arrested state C or the senescent state S as described here by our model, fibroblast cells may become apoptotic, enter the quiescent state or terminally differentiate. To address this question, specific markers for these cellular states should be quantified in order to further extend our current model.

Supporting Information

Supplement S1 Bifurcation analysis of the stress response F(t). (PDF)

Supplement S2 Sensitivity analysis. (PDF)

Supplement S3 Analytical Analysis of the complete cellular senescence model. (PDF)

References

- Campisi J, Sedivy J (2009) How does proliferative homeostasis change with age? What causes it and how does it contribute to aging? *J Gerontol A Biol Sci Med Sci* 64: 164–166.
- Baker DJ, Wijshake T, Tchkonia T, LeBrasseur NK, Childs BG, et al. (2011) Clearance of p16Ink4a-positive senescent cells delays ageing-associated disorders. *Nature* 479: 232–236.
- Hayflick L, Moorhead PS (1961) The serial cultivation of human diploid cell strains. *Exp Cell Res* 25: 585–621.
- Krizhanovsky V, Xue W, Zender L, Yon M, Hernando E, et al. (2008) Implications of cellular senescence in tissue damage response, tumor suppression, and stem cell biology. *Cold Spring Harb Symp Quant Biol* 73: 513–522.
- Chaturvedi V, Qin JZ, Denning MF, Choubey D, Diaz MO, et al. (1999) Apoptosis in proliferating, senescent, and immortalized keratinocytes. *J Biol Chem* 274: 23358–23367.
- Marcotte R, Lacelle C, Wang E (2004) Senescent fibroblasts resist apoptosis by downregulating caspase-3. *Mech Ageing Dev* 125: 777–783.
- Narita M, Nunez S, Heard E, Narita M, Lin AW, et al. (2003) Rb-mediated heterochromatin formation and silencing of E2F target genes during cellular senescence. *Cell* 113: 703–716.
- Dimri GP, Lee X, Basile G, Acosta M, Scott G, et al. (1995) A biomarker that identifies senescent human cells in culture and in aging skin in vivo. *Proc Natl Acad Sci U S A* 92: 9363–9367.
- Kuilman T, Michaloglou C, Mooi WJ, Peeper DS (2010) The essence of senescence. *Genes Dev* 24: 2463–2479.
- Sikora E, Arendt T, Bennett M, Narita M (2011) Impact of cellular senescence signature on ageing research. *Ageing Res Rev* 10: 146–152.
- Cristofalo VJ, Lorenzini A, Allen RG, Torres C, Tresini M (2004) Replicative senescence: a critical review. *Mech Ageing Dev* 125: 827–848.
- Herbig U, Jobling WA, Chen BPC, Chen DJ, Sedivy JM (2004) Telomere shortening triggers senescence of human cells through a pathway involving ATM, p53, and p21(CIP1), but not p16(INK4a). *Mol Cell* 14: 501–513.
- Jeyapalan JC, Ferreira M, Sedivy JM, Herbig U (2007) Accumulation of senescent cells in mitotic tissue of aging primates. *Mech Ageing Dev* 128: 36–44.
- Ressler S, Bartkova J, Niederegger H, Bartek J, Scharfetter-Kochanek K, et al. (2006) p16INK4A is a robust in vivo biomarker of cellular aging in human skin. *Aging Cell* 5: 379–389.
- Robles SJ, Adami GR (1998) Agents that cause DNA double strand breaks lead to p16INK4a enrichment and the premature senescence of normal fibroblasts. *Oncogene* 16: 1113–1123.
- Kosar M, Bartkova J, Hubackova S, Hodny Z, Lukas J, et al. (2011) Senescence-associated heterochromatin foci are dispensable for cellular senescence, occur in a cell type- and insult-dependent manner and follow expression of p16(ink4a). *Cell Cycle* 10: 457–468.
- Coppé J-P, Desprez P-Y, Krtolica A, Campisi J (2010) The senescence-associated secretory phenotype: the dark side of tumor suppression. *Annu Rev Pathol* 5: 99–118.
- Kuilman T, Peeper DS (2009) Senescence-messaging secretome: SMS-ing cellular stress. *Nat Rev Cancer* 9: 81–94.
- Shelton DN, Chang E, Whittier PS, Choi D, Funk WD (1999) Microarray analysis of replicative senescence. *Curr Biol* 9: 939–945.
- Acosta JC, O'Loughlin A, Banito A, Guijarro MV, Augert A, et al. (2008) Chemokine signaling via the CXCR2 receptor reinforces senescence. *Cell* 133: 1006–1018.
- Wajapeyee N, Serra RW, Zhu X, Mahalingam M, Green MR (2008) Oncogenic BRAF induces senescence and apoptosis through pathways mediated by the secreted protein IGFBP7. *Cell* 132: 363–374.
- d'Adda di Fagnola F, Reaper PM, Clay-Farrace L, Fiegler H, Carr P, et al. (2003) A DNA damage checkpoint response in telomere-initiated senescence. *Nature* 426: 194–198.
- Takai H, Smogorzewska A, de Lange T (2003) DNA damage foci at dysfunctional telomeres. *Curr Biol* 13: 1549–1556.
- Nakamura AJ, Redon CE, Bonner WM, Sedelnikova OA (2009) Telomere-dependent and telomere-independent origins of endogenous DNA damage in tumor cells. *Aging (Albany NY)* 1: 212–218.
- Gire V, Roux P, Wynford-Thomas D, Brondello J-M, Dulic V (2004) DNA damage checkpoint kinase Chk2 triggers replicative senescence. *EMBO J* 23: 2554–2563.
- Goodarzi AA, Noon AT, Deckbar D, Ziv Y, Shiloh Y, et al. (2008) ATM signaling facilitates repair of DNA double-strand breaks associated with heterochromatin. *Mol Cell* 31: 167–177.
- Ben-Porath I, Weinberg RA (2004) When cells get stressed: an integrative view of cellular senescence. *J Clin Invest* 113: 8–13.
- Ben-Porath I, Weinberg RA (2005) The signals and pathways activating cellular senescence. *Int J Biochem Cell Biol* 37: 961–976.
- Shay JW, Wright WE (2005) Senescence and immortalization: role of telomeres and telomerase. *Carcinogenesis* 26: 867–874.
- Cosme-Blanco W, Chang S (2008) Dual roles of telomere dysfunction in initiation and suppression of tumorigenesis. *Exp Cell Res* 314: 1973–1979.
- Hwang ES (2002) Replicative senescence and senescence-like state induced in cancer-derived cells. *Mech Ageing Dev* 123: 1681–1694.
- Zhao W, Lin ZX, Zhang ZQ (2004) Cisplatin-induced premature senescence with concomitant reduction of gap junctions in human fibroblasts. *Cell Res* 14: 60–66.
- Havelka AM, Berndtsson M, Olofsson MH, Shoshan MC, Linder S (2007) Mechanisms of action of DNA-damaging anticancer drugs in treatment of carcinomas: is acute apoptosis an “off-target” effect? *Mini Rev Med Chem* 7: 1035–1039.
- Di Leonardo A, Linke SP, Clarkin K, Wahl GM (1994) DNA damage triggers a prolonged p53-dependent G1 arrest and long-term induction of Cip1 in normal human fibroblasts. *Genes Dev* 8: 2540–2551.
- Serrano M, Lin AW, McCurrach ME, Beach D, Lowe SW (1997) Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. *Cell* 88: 593–602.
- Toussaint O, Medrano EE, von Zglinicki T (2000) Cellular and molecular mechanisms of stress-induced premature senescence (SIPS) of human diploid fibroblasts and melanocytes. *Exp Gerontol* 35: 927–945.
- Debaq-Chainiaux F, Borlon C, Pascal T, Royer V, Eliaers F, et al. (2005) Repeated exposure of human skin fibroblasts to UVB at subcytotoxic level triggers premature senescence through the TGF-beta1 signaling pathway. *J Cell Sci* 118: 743–758.
- Mallette FA, Gaumont-Leclerc M-F, Ferbeyre G (2007) The DNA damage signaling pathway is a critical mediator of oncogene-induced senescence. *Genes Dev* 21: 43–48.
- Maslov AY, Vijg J (2009) Genome instability, cancer and aging. *Biochim Biophys Acta* 1790: 963–969.
- Dimri GP, Itahana K, Acosta M, Campisi J (2000) Regulation of a senescence checkpoint response by the E2F1 transcription factor and p14(ARF) tumor suppressor. *Mol Cell Biol* 20: 273–285.
- Pearson M, Carbone R, Sebastiani C, Cioce M, Fagioli M, et al. (2000) pML regulates p53 acetylation and premature senescence induced by oncogenic Ras. *Nature* 406: 207–210.
- Bartkova J, Rezaei N, Liontos M, Karakaidos P, Kletsas D, et al. (2006) Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. *Nature* 444: 633–637.
- Lee AC, Fenster BE, Ito H, Takeda K, Bae NS, et al. (1999) Ras proteins induce senescence by altering the intracellular levels of reactive oxygen species. *J Biol Chem* 274: 7936–7940.
- Di Micco R, Fumagalli M, Cicalese A, Piccinin S, Gasparini P, et al. (2006) Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. *Nature* 444: 638–642.
- Pantoja C, Serrano M (1999) Murine fibroblasts lacking p21 undergo senescence and are resistant to transformation by oncogenic Ras. *Oncogene* 18: 4974–4982.
- Benanti JA, Galloway DA (2004) Normal human fibroblasts are resistant to RAS-induced senescence. *Mol Cell Biol* 24: 2842–2852.
- Rodier F, Campisi J (2011) Four faces of cellular senescence. *J Cell Biol* 192: 547–556.
- Lawless C, Wang C, Jurk D, Merz A, von Zglinicki T, et al. (2010) Quantitative assessment of markers for cell senescence. *Exp Gerontol* 45: 772–778.

Acknowledgments

We thank Christian Bodenstern and Christoph Kaleta for fruitful discussions and helpful remarks.

Author Contributions

Conceived and designed the experiments: S. Schauble KK S. Marthandan S. Münch IH PH SD. Performed the experiments: S. Schauble KK S. Marthandan S. Münch IH PH. Analyzed the data: S. Schauble KK S. Marthandan S. Münch IH PH SD. Contributed reagents/materials/analysis tools: KK S. Marthandan S. Münch PH. Wrote the paper: S. Schauble IH S. Schuster PH SD.

49. Klement K, Melle C, Murzik U, Diekmann S, Norgauer J, et al. (2012) Accumulation of annexin A5 at the nuclear envelope is a biomarker of cellular aging. *Mech Ageing Dev*: in press.
50. Kill IR, Faragher RG, Lawrence K, Shall S (1994) The expression of proliferation-dependent antigens during the lifespan of normal and progeroid human fibroblasts in culture. *J Cell Sci* 107 (Pt 2): 571–579.
51. Faragher RG, Kill IR, Hunter JA, Pope FM, Tannock C, et al. (1993) The gene responsible for Werner syndrome may be a cell division “counting” gene. *Proc Natl Acad Sci U S A* 90: 12030–12034.
52. Kronschnabl M, Stamminger T (2003) Synergistic induction of intercellular adhesion molecule-1 by the human cytomegalovirus transactivators IE2p86 and pp71 is mediated via an Sp1-binding site. *J Gen Virol* 84: 61–73.
53. Seluanov A, Hine C, Bozzella M, Hall A, Sasahara THC, et al. (2008) Distinct tumor suppressor mechanisms evolve in rodent species that differ in size and lifespan. *Aging Cell* 7: 813–823.
54. Thornley JH (1970) Respiration, growth and maintenance in plants. *Nature* 227: 304–305.
55. Penning de Vries FWT (1975) The Cost of Maintenance Processes in Plant Cells. *Annals of Botany* 39: 77–92.
56. Pirt SJ (1965) The maintenance energy of bacteria in growing cultures. *Proc R Soc Lond B Biol Sci* 163: 224–231.
57. Kim CS, Kim JK, Nam SY, Yang KH, Jeong M, et al. (2007) Low-dose radiation stimulates the proliferation of normal human lung fibroblasts via a transient activation of Raf and Akt. *Mol Cells* 24: 424–430.
58. Fumagalli M, Rossiello F, Clerici M, Barozzi S, Cittaro D, et al. (2012) Telomeric DNA damage is irreparable and causes persistent DNA-damage-response activation. *Nat Cell Biol* 14: 355–365.
59. Ludwig D, Jones DD, Holling CS (1978) Qualitative analysis of Insect Outbreak Systems: Spruce Budworm and Forest. *The Journal of Animal Ecology* 47: 315–332.
60. Itahana K, Zou Y, Itahana Y, Martinez J-L, Beausejour C, et al. (2003) Control of the replicative life span of human fibroblasts by p16 and the polycomb protein Bmi-1. *Mol Cell Biol* 23: 389–401.
61. Münch S (2011) Die Rolle der PML-Kernkörperchen in der DNA-Schadensantwort [Dissertation]: Friedrich Schiller University.
62. Kuilman T, Michaloglou C, Vredevelde LCW, Douma S, van Doorn R, et al. (2008) Oncogene-induced senescence relayed by an interleukin-dependent inflammatory network. *Cell* 133: 1019–1031.
63. Coppé J-P, Kauser K, Campisi J, Beausejour CM (2006) Secretion of vascular endothelial growth factor by primary human fibroblasts at senescence. *J Biol Chem* 281: 29568–29574.
64. Rodier F, Coppé J-P, Patil CK, Hoeijmakers WAM, Muñoz DP, et al. (2009) Persistent DNA damage signalling triggers senescence-associated inflammatory cytokine secretion. *Nat Cell Biol* 11: 973–979.
65. Shin J-S, Hong A, Solomon MJ, Lee CS (2006) The role of telomeres and telomerase in the pathology of human cancer and aging. *Pathology* 38: 103–113.
66. Lee BY, Han JA, Im JS, Morrone A, Johung K, et al. (2006) Senescence-associated beta-galactosidase is lysosomal beta-galactosidase. *Aging Cell* 5: 187–195.
67. Chen Q, Fischer A, Reagan JD, Yan LJ, Ames BN (1995) Oxidative DNA damage and senescence of human diploid fibroblast cells. *Proc Natl Acad Sci U S A* 92: 4337–4341.
68. Toussaint O, Remacle J, Dierick J-F, Pascal T, Fripiat C, et al. (2002) Stress-induced premature senescence: from biomarkers to likelihood of in vivo occurrence. *Biogerontology* 3: 13–17.
69. Toussaint O, Weemaels G, Debacq-Chainiaux F, Scharffetter-Kochanek K, Wlaschek M (2011) Artefactual effects of oxygen on cell culture models of cellular senescence and stem cell biology. *J Cell Physiol* 226: 315–321.
70. Tan Z (1999) Intramitotic and intraclonal variation in proliferative potential of human diploid cells: explained by telomere shortening. *J Theor Biol* 198: 259–268.
71. Rubelj I, Vondracek Z (1999) Stochastic mechanism of cellular aging—abrupt telomere shortening as a model for stochastic nature of cellular aging. *J Theor Biol* 197: 425–438.
72. Sozou PD, Kirkwood TB (2001) A stochastic model of cell replicative senescence based on telomere shortening, oxidative stress, and somatic mutations in nuclear and mitochondrial DNA. *J Theor Biol* 213: 573–586.
73. Golubev A, Khrustalev S, Butov A (2003) An in silico investigation into the causes of telomere length heterogeneity and its implications for the Hayflick limit. *J Theor Biol* 225: 153–170.
74. Portugal RD, Land MGP, Svaiter BF (2008) A computational model for telomere-dependent cell-replicative aging. *Biosystems* 91: 262–267.
75. Smith JR, Whitney RG (1980) Intracolon variation in proliferative potential of human diploid fibroblasts: stochastic mechanism for cellular aging. *Science* 207: 82–84.
76. Csikász-Nagy A, Battogtokh D, Chen KC, Novák B, Tyson JJ (2006) Analysis of a generic model of eukaryotic cell-cycle regulation. *Biophys J* 90: 4361–4379.
77. Tyson JJ, Novak B (2008) Temporal organization of the cell cycle. *Curr Biol* 18: R759–R768.
78. Passos JF, Nelson G, Wang C, Richter T, Simillion C, et al. (2010) Feedback between p21 and reactive oxygen production is necessary for cell senescence. *Mol Syst Biol* 6: 347.

3. Constraint-Based Modelling

Applications of EFM analysis and FBA

Chapter summary. *The previous chapter comprised ODE based studies, where kinetic rate laws were applied to study the dynamics of the given systems. Once a system becomes considerably complex, however, the knowledge of specific kinetic rate laws or the computational power to simulate is commonly not available. Consequently, a simulation via deterministic ODE systems is then infeasible. An approach to circumvent these issues is represented by constraint-based modeling methods. In the first study of this chapter, our work on introducing the established concept of EFM analysis is presented. Next to giving a detailed description of the method, we demonstrate its usefulness on a study case of the photosynthate metabolism. The second study in this chapter comprises an elaborate network of the nitrogen uptake metabolism of the green algae *C. reinhardtii*. Here, we investigated the role of circadian regulation by analyzing the EFM set arising from the modeled network by integrating knowledge about sequence motifs that are recognized by a circadian clock regulated element. In the last study, a comprehensive summary of metabolic costs of amino acid and protein production in *E. coli* is presented. We calculated different measures of costs based on a manual calculation and an FBA of a genome-scale model of *E. coli*. This potentially enables an efficient choice of growth media for future industrial designs in order to optimize amino acid or protein production.*

CHAPTER TWENTY-TWO

HANDS-ON METABOLISM: ANALYSIS OF COMPLEX BIOCHEMICAL NETWORKS USING ELEMENTARY FLUX MODES

Sascha Schäuble, Stefan Schuster, *and* Christoph Kaleta

Contents

1. Introduction	438
2. Elementary Flux Modes	438
2.1. Mathematical background	440
3. Application	443
3.1. Network reconstruction	443
3.2. Application to photosynthate metabolism	444
3.3. Overview of further applications of EFM analysis	449
3.4. Pathway analysis in genome-scale metabolic networks	451
4. Conclusion	452
Acknowledgments	452
References	454

Abstract

The aim of this chapter is to discuss the basic principles and reasoning behind elementary flux mode analysis (EFM analysis)—an important tool for the analysis of metabolic networks. We begin with a short introduction into metabolic pathway analysis and subsequently outline in detail fundamentals of EFM analysis by way of a small example network. We discuss issues arising in the reconstruction of metabolic networks required for EFM analysis and how they can be circumvented. Subsequently, we analyze a more elaborate example network representing photosynthate metabolism. Finally, we give an overview of applications of EFM analysis in biotechnology and other fields and discuss issues arising when applying methods from metabolic pathway analysis to genome-scale metabolic networks.

Department of Bioinformatics, Friedrich Schiller University Jena, Jena, Germany

Methods in Enzymology, Volume 500

ISSN 0076-6879, DOI: 10.1016/B978-0-12-385118-5.00022-0

© 2011 Elsevier Inc.

All rights reserved.



1. INTRODUCTION

Exploring the structure of metabolic networks is a key step in order to understand the fundamental properties of living systems (Price *et al.*, 2004; Ruppin *et al.*, 2010). Such networks summarize the capabilities of a subsystem or the entire metabolism of an organism and show, for instance, how a set of source compounds the organism can find in its environment can be converted into the constituent metabolites and macromolecules of which it consists.

However, it is often not possible to identify important routes in metabolic networks from their reactions alone (Schuster *et al.*, 1999). Due to this problem a large array of methods that allow one to analyze such networks based on network stoichiometry and constraints upon fluxes and concentrations have been developed (Price *et al.*, 2004). These methods are often referred to as constraint-based methods. One important method for identifying pathways in metabolic networks is the concept of elementary flux modes (EFMs; Schuster *et al.*, 1999) and the related extreme pathways (Schilling *et al.*, 2000). EFMs correspond to minimal sets of reactions that can work together at steady state while obeying thermodynamical constraints on the direction of reaction fluxes that make some reactions practically irreversible at physiological conditions. The fundamentals of that approach are based on earlier work by Clarke (1981). An EFM is minimal in the sense that removing one reaction will preclude any steady-state flux through the remaining set of reactions of which it consists.

In this work, we will give a detailed introduction into EFM analysis and the concepts upon which this method is based. We will outline this method by way of several example networks that also allow us to demonstrate principal avenues of its application. Further, we will give a brief overview on recent works applying the concept of EFMs. Finally, we will address challenges that arise from the advent of genome-scale metabolic networks that summarize the metabolic capabilities of entire organisms and how they can be met in the context of EFM analysis.



2. ELEMENTARY FLUX MODES

The concept of EFMs (Schuster and Hilgetag, 1994; Schuster *et al.*, 1999) allows one to decompose a reaction network into well-defined metabolic pathways. As stated above, they are defined as minimal sets of reactions that allow for a steady-state flux that uses irreversible reactions only in the thermodynamically feasible direction. To apply EFM analysis, only information on the reaction stoichiometry of the network and information on reaction directionality is required, information that is often much

easier to obtain than details on the precise kinetics governing reactions. The reaction stoichiometry indicates for each reaction the number of metabolite molecules consumed or produced: the stoichiometric coefficient is negative if a metabolite is a substrate of a reaction and positive if it is a product.

The information of reaction stoichiometry is gathered in the stoichiometric matrix \mathbf{N} , where m_{ij} of \mathbf{N} is the coefficient of the i th metabolite taking part in the j th reaction:

$$\mathbf{N} = \begin{pmatrix} m_{11} & m_{12} & \cdots & m_{1m} \\ m_{21} & m_{22} & \cdots & m_{2m} \\ \vdots & & \ddots & \\ m_{n1} & m_{n2} & \cdots & m_{nm} \end{pmatrix}$$

Note that external metabolites are often not included in \mathbf{N} , as they are not considered in the context of the steady-state condition that will be discussed more thoroughly below. Instead they are often part of a parameter vector \mathbf{p} as their concentrations are assumed to be constant.

At this point it is important to note that while most applications of EFM analysis focus on small-molecule metabolism an application to more complex compounds such as DNA and proteins is not precluded. Thus, EFM and extreme pathway analysis have also been used to investigate signaling and regulatory networks (Behre and Schuster, 2009; Gianchandani *et al.*, 2006, 2009).

In order to define a systems boundary, the set of metabolites of a reaction network is decomposed into internal and external metabolites. While internal metabolites are required to be balanced within the network and thus subjected to the steady-state condition, external metabolites are assumed to be buffered. Mechanisms that buffer these metabolites can be, for instance, supply from a growth medium, drain through dilution or participation in a large number of reactions beyond the scope of the metabolic network. Metabolites that are, thus, often considered as external are metabolites of the growth medium, biomass precursors such as amino acids and nucleotides as well as energy currency cofactors such as ATP, NADH, or NADPH.

Under these premises, an EFM can be understood as a path through the network that transforms a set of external (substrate) metabolites into another set of external (product) metabolites over the intermediate of a set of balanced internal metabolites. Figure 22.1 illustrates the principle of EFMs in a biologically relevant pathway—the TCA cycle, the powerplant that provides many organisms with sufficient energy.

Note that for two reasons it is of utter importance to determine the state of a metabolite, that is, which metabolite can be defined as external or internal. First, the simulated uptake or production of particular metabolites, or the maintenance of energy greatly depends on this definition. Second, especially in

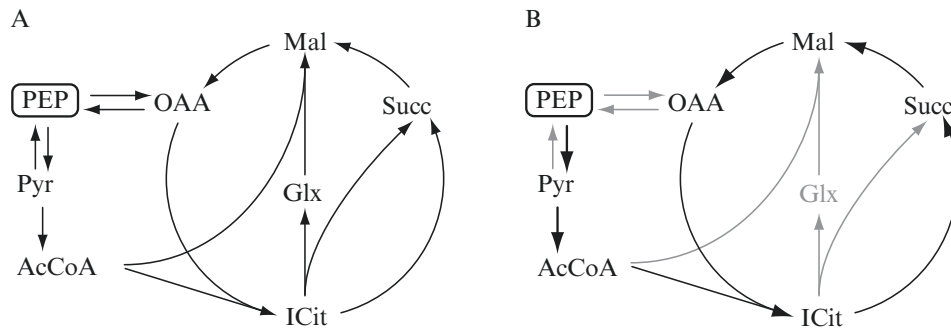


Figure 22.1 EFMs in a simplified model of the TCA cycle including adjacent reactions of glycolysis. The model comprises the metabolites phosphoenolpyruvate (PEP), pyruvate (Pyr), acetyl-Coenzyme-A (AcCoA), isocitrate (ICit), succinate (Succ), malate (Mal), oxaloacetate (OAA), and glyoxylate (Glx) where only the metabolite PEP is set to external status (boxed metabolite). (A) Shows all reactions considered, while (B) displays one valid EFM that completely oxidises PEP. Release and fixation of CO₂ has been omitted for clarity.

dense networks, where the number of EFMs grows exponentially with the size of the network (Klamt and Stelling, 2002), determining the state of metabolites has a very strong influence on the number of resulting EFMs (Dandekar *et al.*, 2003; Gagneur and Klamt, 2004; Schuster *et al.*, 2002b).

In fact, examining particularly compact networks is rather common, since metabolic networks often contain bi- or even multimolecular reactions. By transforming a multitude of metabolites into another, these reactions significantly increase the complexity of the network and hence the number of EFMs. However, as computer performance increases and more efficient algorithms to compute EFMs are developed, networks of increasing size can be investigated using EFM analysis (Gagneur and Klamt, 2004; Pfeiffer *et al.*, 1999; Terzer and Stelling, 2008; Urbanczik and Wagner, 2005).

2.1. Mathematical background

Although specific kinetic data is not required, some general assumptions about the kinetics of the system under consideration have to be made for EFM analysis.

For simplicity, it is reasonable to assume that the metabolites are homogeneously distributed and that the cell does not exhibit a time-dependent inflow or outflow behavior. Thus, the time-course of the concentrations of metabolites can be described formally by the differential equation:

$$\frac{d\mathbf{S}}{dt} = \mathbf{N}\mathbf{V}(\mathbf{S}, \mathbf{p}) - \mu\mathbf{S} \quad (22.1)$$

where $\mathbf{S} = (S_1, S_2, \dots, S_n)^T$, \mathbf{V} , and \mathbf{p} are the vectors of the concentrations of internal metabolites, net reaction rates, and parameters, respectively.

The growth rate μ can be taken into account using two approaches. The first approach is to consider dilution by growth to be so small for the time frame of the analysis that it can be neglected. Alternatively, growth can be modeled within the stoichiometric matrix \mathbf{N} through addition of an artificial biomass reaction that drains components of the cell, or their precursors, in their relative amounts.

Under the time frame considered, it is reasonable to assume that the concentrations of internal metabolites stay constant (Clarke, 1981; Pfeiffer *et al.*, 1999; Schuster and Hilgetag, 1994). Hence, their production and consumption need to be balanced. Thus, Eq. (22.1) can be simplified by setting it equal to zero:

$$\mathbf{N}\mathbf{V}(\mathbf{S}, \mathbf{p}) = 0 \quad (22.2)$$

Obviously, as organisms undergo dynamic processes, no concentration will be completely constant over time. Nevertheless, this assumption also holds in the approximate context as long as no intermediates accumulate or are depleted to a considerable extent over time, for instance, in the case of oscillations.

A first simple approach to analyze the steady-state condition (Eq. (22.2)) is the computation of the null space of \mathbf{N} . The null space or kernel refers to the Euclidean subspace of all vectors \mathbf{V} fulfilling Eq. (22.2). It already features some simple pathways in the network and can be computed by utilizing standard methods of linear algebra (Strang, 2009) such as Gaussian elimination. Note that basis vectors can be understood as steady-state flux distributions across the system.

However, this set of basis vectors is not unique, that is, there can be several sets of basis vectors. Further, they do not take into account the irreversibility of some reactions and might use them in a thermodynamically infeasible direction. To overcome these drawbacks, it is required that a subvector of \mathbf{V}^{irr} of \mathbf{V} , in which the coefficients correspond to irreversible reactions, satisfies

$$\mathbf{V}^{\text{irr}} \geq 0 \quad (22.3)$$

Hence, a linear inequality system is formed by the Eqs. (22.2) and (22.3).

Now, a *flux mode* $\mathbf{V}^* \in \mathbb{R}^r$ with r being the number of reactions within a network is defined as follows:

- (i) *steady-state condition*: \mathbf{V}^* satisfies Eq. (22.2)
- (ii) *sign restriction*: \mathbf{V}^* satisfies Eq. (22.3)

A first important observation is that if a flux vector \mathbf{V} fulfils Eqs. (22.2) and (22.3), also $\lambda\mathbf{V}$ with $\lambda > 0$ fulfils both equations. Hence, the analysis

should be restricted to a set of flux modes in which none can be derived as a simple scalar/multiple of another.

Another important observation from Eqs. (22.2) and (22.3) is that two flux modes \mathbf{V}_1 and \mathbf{V}_2 can be linearly combined by $\lambda_1 \mathbf{V}_1 + \lambda_2 \mathbf{V}_2$ with $\lambda_1, \lambda_2 > 0$ and we again obtain a flux mode. Thus, the analysis can be restricted to *elementary* flux modes. A flux mode is called elementary if suppressing the flux through any reaction used by it implies that there is no flux through the remaining reactions that satisfies Eqs. (22.2) and (22.3) (Schuster and Hilgetag, 1994). This is equivalent to stating that a flux mode \mathbf{V}' is called elementary if there is no other flux mode \mathbf{V}'' that uses a proper subset of the reactions of \mathbf{V}' (Schuster *et al.*, 2002a). In mathematical terms, this statement can be formulated by defining the support “supp” of a flux mode \mathbf{V} , which includes all nonzero elements of \mathbf{V} as

$$\text{supp}(\mathbf{V}) = \{i | V_i \neq 0\} \quad (22.4)$$

Then, a flux mode \mathbf{V}' is called elementary if there exists no other flux mode \mathbf{V}'' such that

$$\text{supp}(\mathbf{V}'') \subseteq \text{supp}(\mathbf{V}'). \quad (22.5)$$

Thus, EFMs can be defined as suggested by Schuster and Hilgetag (1994):

Definition 1 An EFM is a minimal set of enzymes that can operate at steady state with all irreversible reactions used in the correct direction.

An interesting property of the set of EFMs of a reaction network is that they are unique up to scaling by a factor $\lambda > 0$. Moreover, every flux mode \mathbf{V} (which hence satisfies Eqs. (22.2) and (22.3)) can be written as positive linear combination of EFMs:

$$\mathbf{V} = \sum_k \eta_k \mathbf{e}^{(k)}, \quad \eta_k > 0 \quad (22.6)$$

where $\mathbf{e}^{(k)}$ refers to the set of EFMs and η_k represents a scaling factor.

In Fig. 22.2, all of the EFMs of the example network in Fig. 22.1 are displayed. Again, the metabolite PEP is the only external species, whereas all other metabolites are considered to be intermediates and thus set to internal status (balanced at steady state). The model gives rise to four EFMs, of which two correspond to the interconversion of PEP and OAA and the interconversion of PEP and Pyr, respectively (Fig. 22.2B and C). Additionally to these two trivial EFMs, we find two EFMs that constitute all possible pathways on which PEP can be oxidized by this network. The first EFM

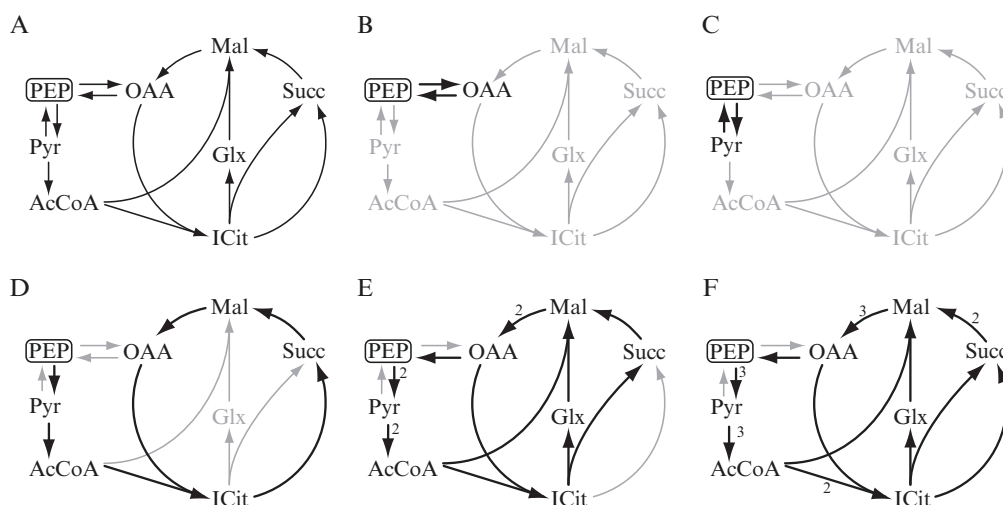


Figure 22.2 The simplified TCA cycle in (A), as depicted [Fig. 22.1](#) gives rise to four EFMs, shown in (B)–(E). The flux in (F) would utilize all reactions, which is a violation of condition (iii) and thus not a regular EFM, since it uses the combined reactions of the EFMs in (D) and (E). Abbreviations are explained in [Fig. 22.1](#).

([Fig. 22.2D](#)) corresponds to the classical scheme of PEP oxidation along the TCA cycle. The other pathway, in contrast, utilizes reactions from the glyoxylate shunt to oxidize PEP. Until some years ago, the enzymes of the glyoxylate shunt have only been considered to play a role in gluconeogenesis/anaplerosis rather than catabolism of glucose ([Fig. 22.2E](#)). However, they have been found to be used for PEP oxidation during growth of *Escherichia coli* ([Fischer and Sauer, 2003](#)) and *Mycobacterium tuberculosis* ([Beste and McFadden, 2010](#)) on low glucose concentrations. The corresponding pathway has been called PEP-glyoxylate cycle. Why are there no more EFMs? One might argue that in principle there should exist at least one more EFM that uses the complete TCA cycle and the PEP-glyoxylate cycle simultaneously ([Fig. 22.2F](#)). While this can be observed in *E. coli* ([Fischer and Sauer, 2003](#)), the resulting set of reactions does not constitute an EFM since it is not minimal, that is, it is just a superposition of two pathways. In consequence, any steady-state flux that uses the TCA cycle and the PEP-glyoxylate cycle simultaneously corresponds to a linear combination of two EFMs.

3. APPLICATION

3.1. Network reconstruction

In order to apply EFM analysis, an accurate model is required. In the most simple case, the particular metabolic network of interest is already published and ready for an EFM analysis, as is the case for classic model organisms, such as *Escherichia coli* or *Bacillus subtilis*. Although a continuously rising

number of models are published, it is often necessary to reconstruct a model from scratch if data concerning an organism of interest is not publicly available or insufficient. One should not underestimate the notable efforts and complications that come along with a model reconstruction effort. This becomes even more crucial, as more and more sequenced genomes are publicly available and allow for a genome-scale reconstruction and analysis of metabolic networks. Even in these cases, the reconstruction task can consume months until a high-quality model is developed (Feist *et al.*, 2009; Rupp *et al.*, 2010; Thiele and Palsson, 2010).

For instance, one might encounter the obstacle that databases like KEGG (Kanehisa *et al.*, 2010) or MetaCyc (Caspi *et al.*, 2010) that comprise metabolic pathway data contain contradicting information. This can be due to different conventions concerning the protonation of metabolites or differences in the reversibility assignment of reactions. The latter can be thermodynamically feasible in forward and backward direction or only in one of them. This issue is of critical importance, as those databases commonly form the starting point for the reconstruction process. Since data from different databases often do not agree with each other, including data from multiple sources, such as genomic, proteomic, or bibliomic sources, is most recommended if not mandatory, in order to reconstruct a high-quality model.

Note that genome-scale metabolic models are nevertheless abstractions from real world biological pathways and do not cover the complete set of reactions featured by the underlying organism *per se*, for instance, if some reactions have not yet been identified. Moreover, if no gene ID can be found in public databases, it is rather common that corresponding reactions, for instance, transporters are artificially introduced into the system in order to fill open gaps. This occurs very often in eukaryotes that feature different compartments and cell types. Although reasonable, this should naturally be regarded as a danger to the correctness of a model and hence be considered when reconstructing and ultimately analyzing a metabolic network. Therefore, subsequent analysis can only be considered accurate if the underlying model is as complete as sources will allow. If information is missing or inconsistent, sometimes biochemical expertise can be used to improve on this situation.

One good starting point for a reconstruction are central metabolic pathways as they are described in biochemistry textbooks. Although these networks represent only a small fraction of a complete biological system, they are commonly found in most organisms and have already been investigated for decades.

3.2. Application to photosynthate metabolism

In order to demonstrate how EFM analysis can be used in detail we will analyze a metabolic model of photosynthesis. This model (Fig. 22.3; Appendix) includes the photosynthate metabolism of the chloroplast stroma and

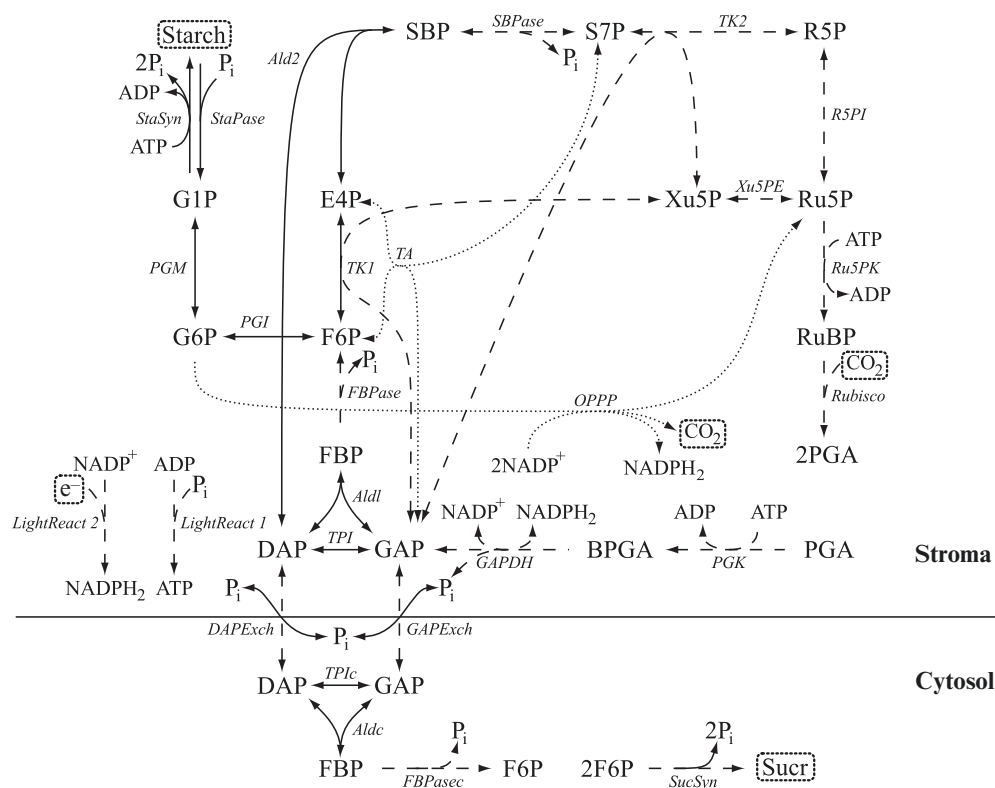


Figure 22.3 Reactions of the Calvin cycle, including parts of glycolysis. Unidirectional (bidirectional) arrows indicate irreversible (reversible) reactions. Dashed reactions are activated whereas dotted reactions are downregulated during the day. Boxed metabolites are set to external and all others to internal status for EFM analysis. Abbreviations of metabolites: PGA, 3-phosphoglycerate; BPGA, glycerate-1, 3-bisphosphate; GAP, glyceraldehyde-3-phosphate; DAP, dihydroxyacetone phosphate; FBP, fructose-1, 6-bisphosphate; F6P, fructose-6-phosphate; E4P, erythrose-4-phosphate; SBP, sedoheptulose-1,7-bisphosphate; S7P, sedoheptulose-7-phosphate; R5P, ribose-5-phosphate; Ru5P, ribulose-5-phosphate; RuBP, ribulose-1, 5-bisphosphate; X5P, xylulose-5-phosphate; G6P, glucose-6-phosphate; G1P, glucose-1-phosphate.

comprises reactions from the Calvin cycle, which is primarily regulated by the thioredoxin system (Schürmann and Jacquot, 2000). In contrast to the work of Poolman *et al.* (2003), who analyzed a similar model with respect to triose phosphate export in detail, we will investigate the production of sucrose and starch. Sucrose and starch are the major carbohydrate storage compounds in plants. Moreover, sucrose is the carbohydrate transported in the phloem and also corresponds to the sugar commonly used in everyone's kitchen.

If all reactions are taken into account, the system gives rise to 42 EFMs in total. As we will limit the analysis on EFMs during daytime (where the two dotted reactions in Fig. 22.3 are inactive), only 12 EFMs are relevant. These EFMs are listed in Table 22.1 and characterize three groups of EFMs of which some are shown in Fig. 22.4. These groups can be differentiated

Table 22.1 Twelve EFMs active during the day

Group	No.	Overall reaction	Enzymes
I	(1)	No net transformation	-TPI GAPEXch -DAPEXch TPIc
I	(2)	No net transformation	StaSyn StaPase Lightreact1
II	(3)	$6\text{CO}_2 + 12\text{e}^- \rightarrow \text{Starch}$	6 Rubisco 12 PGK 12 GAPDH 5 TPI 3 Ald1 3 FB Pase PGI PGM StaSyn 2 TK1 2 Ald2 2 SB Pase 2 TK2 2 R5PI 4 Xu5PE 6 Ru5PK 19 Lightreact1 12 Lightreact2
II	(4)	$6\text{CO}_2 + 12\text{e}^- \rightarrow \text{Starch}$	6 Rubisco 12 PGK 12 GAPDH 3 Ald1 3 FB Pase PGI PGM StaSyn 2 TK1 2 Ald2 2 SB Pase 2 TK2 2 R5PI 4 Xu5PE 6 Ru5PK 5 GAPEXch -5 DAPEXch 19 Lightreact1 5 TPIc 12 Lightreact2
III	(5)	$12\text{CO}_2 + 24\text{e}^- \rightarrow \text{Sucr}$	12 Rubisco 24 PGK 24 GAPDH 8 TPI 4 Ald1 4 FB Pase 4 TK1 4 Ald2 4 SB Pase 4 TK2 4 R5PI 8 Xu5PE 12 Ru5PK 4 GAPEXch 36 Lightreact1 2 TPIc 2 Aldc 2 FB Pasec SucSyn 24 Lightreact2
III	(6)	$12\text{CO}_2 + 24\text{e}^- \rightarrow \text{Sucr}$	12 Rubisco 24 PGK 24 GAPDH 10 TPI 4 Ald1 4 FB Pase 4 TK1 4 Ald2 4 SB Pase 4 TK2 4 R5PI 8 Xu5PE 12 Ru5PK 2 GAPEXch 2 DAPEXch 36 Lightreact1 2 Aldc 2 FB Pasec SucSyn 24 Lightreact2
III	(7)	$12\text{CO}_2 + 24\text{e}^- \rightarrow \text{Sucr}$	12 Rubisco 24 PGK 24 GAPDH 12 TPI 4 Ald1 4 FB Pase 4 TK1 4 Ald2 4 SB Pase 4 TK2 4 R5PI 8 Xu5PE 12 Ru5PK 4 DAPEXch 36 Lightreact1 -2 TPIc 2 Aldc 2 FB Pasec Suc Syn 24 Lightreact2

III	(8)	$12\text{CO}_2 + 24\text{e}^- \rightarrow \text{Sucr}$	12 Rubisco 24 PGK 24 GAPDH 4 Ald1 4 FB Pase 4 TK1 4 Ald2 4 SB Pase 4 TK2 4 R5PI 8 Xu5PE 12 Ru5PK 12 GAPEXch -8 DAPEXch 36 Lightreact1 10 TPIc 2 Aldc 2 FB Pasec SucSyn 24 Lightreact2
III	(9)	$12\text{CO}_2 + 4 \times \text{Starch} + 24\text{e}^- \rightarrow 3 \times \text{Sucr}$	12 Rubisco 24 PGK 24 GAPDH 10 TPI -4 PGI -4 PGM 4 TK1 4 Ald2 4 SB Pase 4 TK2 4 R5PI 8 Xu5PE 12 Ru5PK 6 GAPEXch 6 DAPEXch 4 StaPase 36 Lightreact1 6 Aldc 6 FB Pasec 3 SucSyn 24 Lightreact2
III	(10)	$12\text{CO}_2 + 4 \times \text{Starch} + 24\text{e}^- \rightarrow 3 \times \text{Sucr}$	12 Rubisco 24 PGK 24 GAPDH 16 TPI -4 PGI -4 PGM 4 TK1 4 Ald2 4 SB Pase 4 TK2 4 R5PI 8 Xu5PE 12 Ru5PK 12 DAPEXch 4 StaPase 36 Lightreact1 -6 TPIc 6 Aldc 6 FB Pasec 3 SucSyn 24 Lightreact2
III	(11)	$12\text{CO}_2 + 4 \times \text{Starch} + 24\text{e}^- \rightarrow 3 \times \text{Sucr}$	12 Rubisco 24 PGK 24 GAPDH 4 TPI -4 PGI -4 PGM 4 TK1 4 Ald2 4 SB Pase 4 TK2 4 R5PI 8 Xu5PE 12 Ru5PK 12 GAPEXch 4 StaPase 36 Lightreact1 6 TPIc 6 Aldc 6 FB Pasec 3 SucSyn 24 Lightreact2
III	(12)	$12\text{CO}_2 + 4 \times \text{Starch} + 24\text{e}^- \rightarrow 3 \times \text{Sucr}$	12 Rubisco 24 PGK 24 GAPDH -4 PGI -4 PGM 4 TK1 4 Ald2 4 SB Pase 4 TK2 4 R5PI 8 Xu5PE 12 Ru5PK 16 GAPEXch -4 DAPEXch 4 StaPase 36 Lightreact1 10 TPIc 6 Aldc 6 FB Pasec 3 SucSyn 24 Lightreact2

The group classification is given in the first column, while the net transformation is given in the third and the fluxes of the reactions in the fourth column. An overview of the reaction scheme is given in [Fig. 22.3](#).

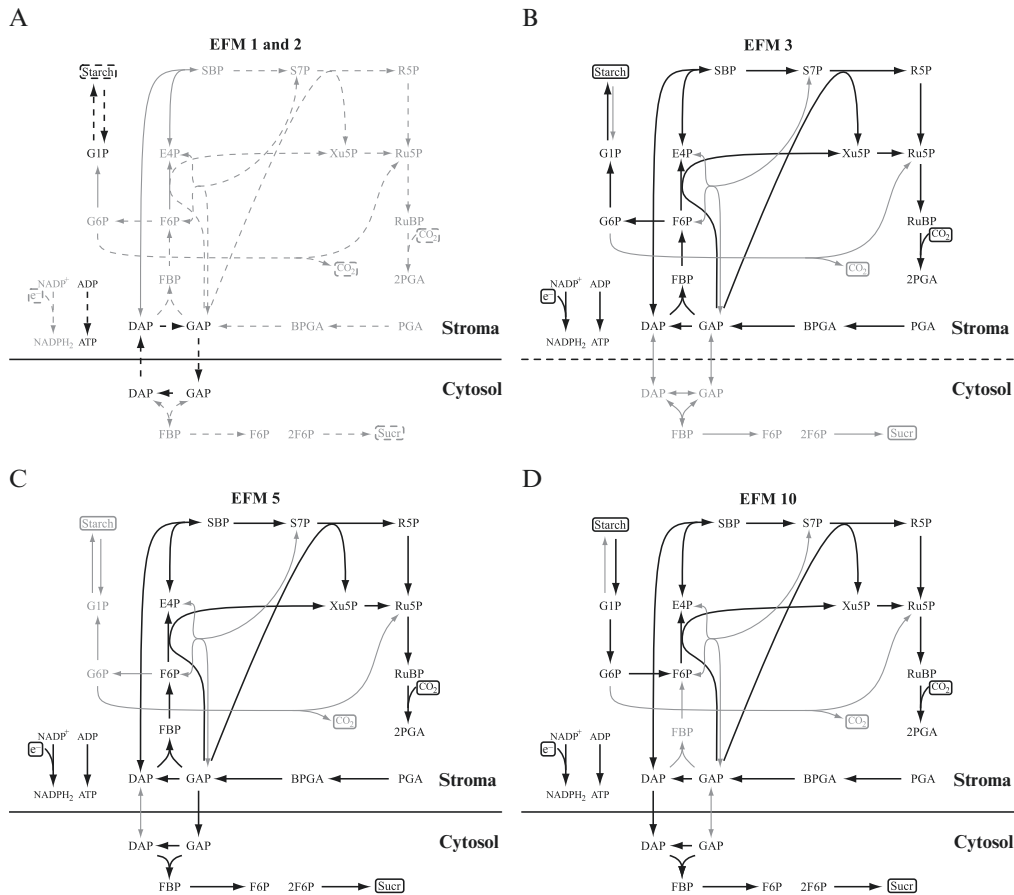


Figure 22.4 Selected EFMs of the model of Fig. 22.3. (A) Displays two cycling EFMs that show no net transformation (EFM 1—solid arrows, EFM 2—dashed arrows). The EFM in (B) describes starch production, via Rubisco. Sucrose synthesis from CO₂ alone or from CO₂ and starch is displayed in (C) and (D), respectively. The selected EFMs as well as further EFMs are described in Table 22.1 and in the text. Names of enzymes are given in Fig. 22.3.

according to their net transformation of external metabolites: in the first group, no net transformation occurs while EFMs of the second and third group produce starch and sucrose, respectively.

The first group comprises two cycles that are present during the day. The first reflects reversible DAP and GAP exchange across the cytosolic membrane with concurrent antiport of P_i (EFM 1, Fig. 22.4A). As this cycle is not driven thermodynamically by external metabolites, it is detailed balanced and can, thus, not carry any steady-state flux due to thermodynamical reasons. The second EFM is a futile cycle, as it consumes ATP from the light reactions, while interconverting starch and G1P back and forth (EFM 2, Fig. 22.4A). As energy is available in excess through the light reactions, both starch and sucrose can be produced at daytime.

The EFMs synthesizing starch (EFMs 3 and 4, see also [Fig. 22.4](#)) can be differentiated by their location. While EFM 3 ([Fig. 22.4B](#)) occurs solely in the stroma, EFM 4 (not shown in [Fig. 22.4](#)) comprises reactions of the stroma and the cytosol. Nevertheless, both EFMs require 19 mol of ATP and 12 mol of NADPH, provided by the light reactions in order to produce 1 mol of starch (i.e., the metabolite that represents one hexose unit in starch) and are thus equally “expensive” in the context of this model.

Sucrose can be produced solely from carbon dioxide (EFMs 5–8) or both, from carbon dioxide and starch (EFMs 9–12). An interesting question is which of the pathways is energetically more efficient, since both use the same amount of ATP and NADPH provided by the light reactions, but produce different quantities of sucrose from different sets of precursors. Producing 3 mol of sucrose from carbon dioxide alone requires an investment of three times 36 mol of ATP and 24 mol of NADPH provided by the light reactions and hence an overall consumption of 108 mol of ATP and 72 mol of NADPH. Producing 3 mol of sucrose from starch and carbon dioxide requires an investment of 36 mol of ATP, 24 mol of NADPH, and 4 mol of starch. Producing 4 mol of starch through EFMs of group II requires an additional investment of four times 19 mol of ATP and 12 mol of NADPH. In consequence, an overall $36 + 4 \times 19 = 112$ mol of ATP and $24 + 4 \times 12 = 72$ mol of NADPH are consumed to produce 3 mol of sucrose through EFMs 9–12. Thus, producing sucrose through EFMs 5–8 is energetically more favorable. Hence, while sucrose biosynthesis from light and starch is in principle possible, it is slightly more expensive for plants to rely also on starch deposits, if the light source is not sufficient. For further selected EFMs, see [Fig. 22.4](#).

Since [Table 22.1](#) comprises 12 EFMs, there are 30 EFMs left. All these 30 EFMs have in common that they either use reactions of the oxidative pentose phosphate pathway (OPPP in [Fig. 22.3](#)) or the transaldolase. These reactions are downregulated during the day and, thus, not available ([Schürmann and Jacquot, 2000](#)). All of these EFMs additionally use Rubisco or sedoheptulose-bisphosphatase (SBPase), reactions which are downregulated during the night ([Schürmann and Jacquot, 2000](#)). In consequence, these 30 EFMs are also downregulated during the night.

Hence, neither starch production nor sucrose synthesis is possible during night in the context of this model. Even if an ATP source were available during night, additional reactions that are downregulated during this phase of the day would be required. A detailed description and analysis of triose phosphate biosynthesis in plants is given in [Poolman *et al.* \(2003\)](#).

3.3. Overview of further applications of EFM analysis

EFM analysis has been used intensively to analyze the metabolic capabilities of organisms. One important example is the analysis of metabolic pathways in the central metabolism of *Escherichia coli* that provided evidence for the

existence of an alternative pathway for the complete oxidation of glucose apart from the TCA cycle (Liao *et al.*, 1996; Schuster *et al.*, 1999). This pathway, now called PEP-glyoxylate cycle, after it had been experimentally confirmed (Fischer and Sauer, 2003), makes use of the glyoxylate shunt and skips several reactions of the TCA cycle (Fig. 22.2E). Recently, this pathway has also been reported in *Mycobacterium tuberculosis* (Beste and McFadden, 2010). Another example is the analysis of a long-standing issue in biochemistry: the question whether even-chain fatty acids can be converted into carbohydrates in humans at steady state. Even though there exists a connected route between acetyl-CoA—the product of β -oxidation of even-chain fatty acids—and glucose along the TCA cycle, there can be no steady-state flux along this route at steady state. Analyzing a metabolic network comprising relevant reactions, de Figueiredo *et al.* (2009b) could indeed show that there exists no elementary mode within this network that converts fatty acids into glucose thus confirming earlier results (Weinman *et al.*, 1957).

Another important application of EFM analysis is the investigation of the susceptibility of metabolic networks to perturbations. A useful property of EFMs is that the knockout of an enzyme can be simulated by removing all EFMs that contain a reaction that is catalyzed by that enzyme. In consequence, enzyme deficiencies can be easily analyzed and EFM analysis has been used to understand metabolic pathways within a mutant of *E. coli* that lacks an outer membrane and the cell wall (Kenanov *et al.*, 2010). Moreover, it has been used to analyze medical implications of enzyme deficiencies in human erythrocytes (Çakir *et al.*, 2004; Schuster and Kenanov, 2005). On a larger scale, EFM analysis has been used to analyze and compare different networks with regard to their susceptibility to random perturbations (Behre *et al.*, 2008; Stelling *et al.*, 2002; Wilhelm *et al.*, 2004).

EFM analysis can also be used to facilitate the interpretation and integration of large-scale experimental data sets. The analysis of such data sets is and has been an important focus in Systems Biology since data analysis is currently lagging behind data generation (Palsson and Zengler, 2010). EFM analysis is a suitable tool in this endeavor as has been shown in a study which investigates regulatory adaptations of *E. coli* to different carbon sources (Stelling *et al.*, 2002) and in another work in which transcriptomic changes during the response to stresses in yeast have been analyzed (Schwartz *et al.*, 2007).

One particular successful area of application of EFM analysis is biotechnology. It has been used in the design and implementation of several strains of *E. coli* overproducing various biotechnological products of interest (Carlson *et al.*, 2002; Trinh and Sreenc, 2009; Trinh *et al.*, 2006; Unrean *et al.*, 2010). The motivation behind these applications is to reduce the space of admissible fluxes toward a smaller space in which the production of the

desired product is presumably coupled to a cellular objective such as maximizing the growth rate. This is achieved by knocking out genes that remove EFMs that have low yields in the desired product. Such knockout strategies can be identified by using the concept of minimal cut sets (Klamt, 2006)—minimal sets of reactions that allow to block a given set of EFMs. Also other theoretical tools based on EFM analysis that allow one to identify genetic modifications that increase the flux to a desired product have been proposed recently (Boghigian *et al.*, 2010; Bohl *et al.*, 2010; Hädicke and Klamt, 2010).

3.4. Pathway analysis in genome-scale metabolic networks

EFMs can only be enumerated in small to medium-scale metabolic networks. These restrictions come from the exponential increase in the number of EFMs with network size (Klamt and Stelling, 2002). In recent years, the algorithms to compute EFMs have been considerably improved (Gagneur and Klamt, 2004; Terzer and Stelling, 2008; Urbanczik and Wagner, 2005) such that networks with up to 25 million EFMs could be analyzed (Terzer and Stelling, 2008). However, the estimated number of EFMs in genome-scale metabolic networks is much larger. For instance, it has been estimated that the number of extreme pathways, a subset of the EFMs, in a genome-scale metabolic network of humans is approximately 10^{29} (Yeung *et al.*, 2007). Thus, even with drastic improvements in algorithms and hardware the enumeration of all EFMs in most genome-scale metabolic networks appears to be infeasible, apart from the difficulty to analyze such a large set of EFMs.

To counter this problem, recently, several approaches that allow for a pathway-based analysis even in genome-scale metabolic networks have been developed. They can be divided into two types. The first type aims to enumerate a subset of the EFMs or extreme pathways that fulfill certain biotechnological criteria (de Figueiredo *et al.*, 2009a; Kaleta *et al.*, 2009a; Xi *et al.*, 2009) and the second analyses sets of reactions that are part of a pathway of the entire system within a specific subsystem, so-called elementary flux patterns (Kaleta *et al.*, 2009b).

The advantage of the first methodology is that a subset of EFMs can be enumerated without requiring to compute the entire set as was necessary using classical algorithms for EFM computation. If one is interested in the shortest EFMs using a specific reaction, a mixed-integer linear programming formulation can be used to compute EFMs (de Figueiredo *et al.*, 2009a). Alternatively, large sets of EFMs using a specific reaction can be obtained using a genetic algorithm that allows one to sample EFMs randomly (Kaleta *et al.*, 2009a). Also random sampling methods have been used to identify subsets of extreme pathways in large-scale networks (Xi *et al.*, 2009).

The second methodology, elementary flux pattern analysis ([Kaleta *et al.*, 2009b](#)), allows one to identify pathways within a subsystem of metabolism that are compatible with a steady-state flux of the entire system. Due to this strong link of elementary flux patterns with the entire system it is further possible to analyze how a subsystem integrates into the remaining system. Using this approach, many methods building on EFM analysis can also be applied to subsystems of genome-scale metabolic networks. Another interesting property of elementary flux patterns is that they put no constraints on the connectivity between the reactions of the subsystem that are considered. In consequence, it is even possible to analyze subsystems that comprise two sets of reactions that do not interface each other through common substrates or products. This can be of importance when analyzing, for instance, the dependencies between a set of reactions within a parasite and a set of reactions within the host in order to understand how both organisms interact on a metabolic level. Thus, critical metabolic dependencies between the host and the parasite can be identified in order to develop new drugs that affect the parasite but have only a small deleterious effect on the host. The analysis of such consortium pathways going across two or more organisms is a topic of high current interest ([Bordbar *et al.*, 2010](#); [Raghunathan *et al.*, 2009](#)).



4. CONCLUSION

EFM analysis is a useful tool that allows the decomposition of biochemical networks into minimal constituent pathways. Thus, EFM analysis has already seen a wide array of applications ranging from theoretical works to biotechnology and has made significant contributions to research in these fields. However, EFM analysis in its classical form cannot be applied to genome-scale metabolic networks which has been considered as a downturn of this method. But with the development of new tools and concepts that port EFM analysis to such networks some hurdles have already been vanquished. In consequence, even in the era of genome-scale metabolic networks, EFM analysis is still a central tool for the analysis of metabolic pathways.

ACKNOWLEDGMENTS

Financial support from the German Ministry for Research and Education (BMBF) to C. Kaleta within the framework of the Forsys Partner initiative (grant FKZ 0315285E) and to S. Schäuble within the framework of the GerontoSys initiative (grant FKZ 0315581D) is gratefully acknowledged. Further, we thank Luís Filipe de Figueiredo and Ines Heiland for useful discussions on the topic of metabolic pathway analysis.

APPENDIX

Reaction scheme of the photosynthate metabolism in metatool format (discussed in [Section 3.2](#)).

```
-METEXT
CO2 Starch e Sucr
-CAT
#reactions of the Calvin cycle (down at night)
Rubisco : CO2 + RuBP => 2 PGA
PGK : PGA + ATP = BPGA + ADP
GAPDH : BPGA + NADPH = NADP + GAPs + Pis
FBPase : FBPs => F6Ps + Pis
SBPase : SBP => S7P + Pis
Ru5PK : Ru5P + ATP => RuBP + ADP
StaSyn : G1P + ATP => ADP + 2 Pis + Starch
Lightreact2 : ADP + Pis => ATP
Lightreact1 : NADPred : NADP + e => NADPH
#stroma reactions
TPI : GAPs = DAPs
Ald1 : DAPs + GAPs = FBPs
TK1 : F6Ps + GAPs = E4P + X5P
Ald2 : E4P + DAPs = SBP
TK2 : GAPs + S7P = X5P + R5P
R5PI : R5P = Ru5P
Xu5PE : X5P = Ru5P
PGI : F6Ps = G6Ps
PGM : G6Ps = G1P
StaPase : Starch + Pis => G1P
#cytoplasmatic reactions
TPIc : GAPc = DAPc
Aldc : GAPc + DAPc = FBPc
FBPasec : FBPc => F6Pc + Pic
SucSyn : 2 F6Pc => Sucr + 2 Pic
#transport reactions
GAPEXch : GAPs + Pic = Pis + GAPc
DAPEXch : DAPs + Pic = Pis + DAPc
#unique in oxPPP #down at day
TA : F6Ps + GAPs = E4P + S7P
OPPP : G6Ps + 2 NADP => 2 NADPH + CO2 + Ru5P
```

REFERENCES

- Behre, J., and Schuster, S. (2009). Modeling signal transduction in enzyme cascades with the concept of elementary flux modes. *J. Comput. Biol.* **16**, 829–844.
- Behre, J., Wilhelm, T., von Kamp, A., Rupp, E., and Schuster, S. (2008). Structural robustness of metabolic networks with respect to multiple knockouts. *J. Theor. Biol.* **252**, 433–441.
- Beste, D. J. V., and McFadden, J. (2010). Systems biology of the metabolism of *Mycobacterium tuberculosis*. *Biochem. Soc. Trans.* **38**, 1286–1289.
- Boghigian, B. A., Shi, H., Lee, K., and Pfeifer, B. A. (2010). Utilizing elementary mode analysis, pathway thermodynamics, and a genetic algorithm for metabolic flux determination and optimal metabolic network design. *BMC Syst. Biol.* **4**, 49, doi: 10.1186/1752-0509-4-49.
- Bohl, K., de Figueiredo, L. F., Hädicke, O., Klamt, S., Kost, C., Schuster, S., and Kaleta, C. (2010). CASOP GS: Computing intervention strategies targeted at production improvement in genome-scale metabolic networks. In “Lecture Notes in Informatics,” (D. Schomburg and A. Grote, eds.), Vol. P-173, pp. 71–80. Gesellschaft für Informatik, Bonn.
- Bordbar, A., Lewis, N. E., Schellenberger, J., Palsson, B. O., and Jamshidi, N. (2010). Insight into human alveolar macrophage and *M. tuberculosis* interactions via metabolic reconstructions. *Mol. Syst. Biol.* **6**, 422, doi: 10.1038/msb.2010.68.
- Çakir, T., Tacer, C. S., and Ülgen, K. O. (2004). Metabolic pathway analysis of enzyme-deficient human red blood cells. *Biosystems* **78**, 49–67.
- Carlson, R., Fell, D., and Sreenc, F. (2002). Metabolic pathway analysis of a recombinant yeast for rational strain development. *Biotechnol. Bioeng.* **79**, 121–134.
- Caspi, R., Altman, T., Dale, J. M., Dreher, K., Fulcher, C. A., Gilham, F., Kaipa, P., Karthikeyan, A. S., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L. A., *et al.* (2010). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* **38**, D473–D479, doi: 10.1093/nar/gkp875.
- Clarke, B. L. (1981). Complete set of steady states for the general stoichiometric dynamical system. *J. Chem. Phys.* **75**, 4970–4979.
- Dandekar, T., Moldenhauer, F., Bulik, S., Bertram, H., and Schuster, S. (2003). A method for classifying metabolites in topological pathway analyses based on minimization of pathway number. *Biosystems* **70**, 255–270.
- de Figueiredo, L. F., Podhorski, A., Rubio, A., Kaleta, C., Beasley, J. E., Schuster, S., and Planes, F. J. (2009a). Computing the shortest elementary flux modes in genome-scale metabolic networks. *Bioinformatics* **25**, 3158–3165.
- de Figueiredo, L. F., Schuster, S., Kaleta, C., and Fell, D. A. (2009b). Can sugars be produced from fatty acids? A test case for pathway analysis tools. *Bioinformatics* **25**, 152–158.
- Feist, A. M., Herrgard, M. J., Thiele, I., Reed, J. L., and Palsson, B.Ø. (2009). Reconstruction of biochemical networks in microorganisms. *Nat. Rev. Microbiol.* **7**, 129–143.
- Fischer, E., and Sauer, U. (2003). A novel metabolic cycle catalyzes glucose oxidation and anaplerosis in hungry *Escherichia coli*. *J. Biol. Chem.* **278**, 46446–46451.
- Gagneur, J., and Klamt, S. (2004). Computation of elementary modes: a unifying framework and the new binary approach. *BMC Bioinformatics* **5**, 175, doi: 10.1186/1471-2105-5-175.
- Gianchandani, E. P., Papin, J. A., Price, N. D., Joyce, A. R., and Palsson, B.Ø. (2006). Matrix formalism to describe functional states of transcriptional regulatory systems. *PLoS Comput. Biol.* **2**, e101, doi: 10.1371/journal.pcbi.0020101.

- Gianchandani, E. P., Joyce, A. R., Palsson, B.Ø., and Papin, J. A. (2009). Functional states of the genome-scale *Escherichia coli* transcriptional regulatory system. *PLoS Comput. Biol.* **5**, e1000403, doi: 10.1371/journal.pcbi.1000403.
- Hädicke, O., and Klamt, S. (2010). CASOP: A computational approach for strain optimization aiming at high productivity. *J. Biotechnol.* **147**, 88–101.
- Kaleta, C., de Figueiredo, L. F., Behre, J., and Schuster, S. (2009a). EFMEvolver: Computing elementary flux modes in genome-scale metabolic networks. In “Lecture Notes in Informatics—Proceedings,” (I. Grosse, S. Neumann, S. Posch, F. Schreiber, and P. Stadler, eds.), Vol. P-157, pp. 179–189. Gesellschaft für Informatik, Bonn.
- Kaleta, C., de Figueiredo, L. F., and Schuster, S. (2009b). Can the whole be less than the sum of its parts? Pathway analysis in genome-scale metabolic networks using elementary flux patterns. *Genome Res.* **19**, 1872–1883.
- Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., and Hirakawa, M. (2010). KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **38**, D355–D360, doi: 10.1093/nar/gkp896.
- Kenanov, D., Kaleta, C., Petzold, A., Hoischen, C., Diekmann, S., Siddiqui, R. A., and Schuster, S. (2010). Theoretical study of lipid biosynthesis in wild-type *Escherichia coli* and in a protoplast-type L-form using elementary flux mode analysis. *FEBS J.* **277**, 1023–1034.
- Klamt, S. (2006). Generalized concept of minimal cut sets in biochemical networks. *Biosystems* **83**, 233–247.
- Klamt, S., and Stelling, J. (2002). Combinatorial complexity of pathway analysis in metabolic networks. *Mol. Biol. Rep.* **29**, 233–236.
- Liao, J. C., Hou, S. Y., and Chao, Y. P. (1996). Pathway analysis, engineering, and physiological considerations for redirecting central metabolism. *Biotechnol. Bioeng.* **52**, 129–140.
- Palsson, B.Ø., and Zengler, K. (2010). The challenges of integrating multiomic data sets. *Nat. Chem. Biol.* **6**, 787–789.
- Pfeiffer, T., Sánchez-Valdenebro, I., Nuño, J. C., Montero, F., and Schuster, S. (1999). METATOOL: For studying metabolic networks. *Bioinformatics* **15**, 251–257.
- Poolman, M. G., Fell, D. A., and Raines, C. A. (2003). Elementary modes analysis of photosynthate metabolism in the chloroplast stroma. *Eur. J. Biochem.* **270**, 430–439.
- Price, N. D., Reed, J. L., and Palsson, B.Ø. (2004). Genome-scale models of microbial cells: Evaluating the consequences of constraints. *Nat. Rev. Microbiol.* **2**, 886–897.
- Raghunathan, A., Reed, J., Shin, S., Palsson, B.Ø., and Daefler, S. (2009). Constraint-based analysis of metabolic capacity of *Salmonella typhimurium* during host-pathogen interaction. *BMC Syst. Biol.* **3**, 38, doi: 10.1186/1752-0509-3-38.
- Ruppin, E., Papin, J. A., de Figueiredo, L. F., and Schuster, S. (2010). Metabolic reconstruction, constraint-based analysis and game theory to probe genome-scale metabolic networks. *Curr. Opin. Biotechnol.* **21**, 502–510.
- Schilling, C. H., Letscher, D., and Palsson, B.Ø. (2000). Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theor. Biol.* **203**, 229–248.
- Schürmann, P., and Jacquot, J. P. (2000). Plant thioredoxin systems revisited. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **51**, 371–400.
- Schuster, S., and Hilgetag, C. (1994). On Elementary Flux Modes in biochemical reaction systems at steady state. *J. Biol. Syst.* **2**, 165–182.
- Schuster, S., and Kenanov, D. (2005). Adenine and adenosine salvage pathways in erythrocytes and the role of S-adenosylhomocysteine hydrolase. A theoretical study using elementary flux modes. *FEBS J.* **272**, 5278–5290.

- Schuster, S., Dandekar, T., and Fell, D. A. (1999). Detection of elementary flux modes in biochemical networks: A promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol.* **17**, 53–60.
- Schuster, S., Hilgetag, C., Woods, J. H., and Fell, D. A. (2002a). Reaction routes in biochemical reaction systems: Algebraic properties, validated calculation procedure and example from nucleotide metabolism. *J. Math. Biol.* **45**, 153–181.
- Schuster, S., Pfeiffer, T., Moldenhauer, F., Koch, I., and Dandekar, T. (2002b). Exploring the pathway structure of metabolism: Decomposition into subnetworks and application to *Mycoplasma pneumoniae*. *Bioinformatics* **18**, 351–361.
- Schwartz, J. M., Gauguier, C., Nacher, J. C., de Daruvar, A., and Kanehisa, M. (2007). Observing metabolic functions at the genome scale. *Genome Biol.* **8**, R123, doi: 10.1186/gb-2007-8-6-r123.
- Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., and Gilles, E. D. (2002). Metabolic network structure determines key aspects of functionality and regulation. *Nature* **420**, 190–193.
- Strang, G. (2009). Introduction to Linear Algebra. 4 edn. Wellesley-Cambridge, Wellesley.
- Terzer, M., and Stelling, J. (2008). Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics* **24**, 2229–2235, doi: 10.1093/bioinformatics/btn401.
- Thiele, I., and Palsson, B.Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* **5**, 93–121, doi: 10.1038/nprot.2009.203.
- Trinh, C. T., and Sreenc, F. (2009). Metabolic engineering of *Escherichia coli* for efficient conversion of glycerol to ethanol. *Appl. Environ. Microbiol.* **75**, 6696–6705.
- Trinh, C. T., Carlson, R., Wlaschin, A., and Sreenc, F. (2006). Design, construction and performance of the most efficient biomass producing *E. coli* bacterium. *Metab. Eng.* **8**, 628–638.
- Unrean, P., Trinh, C. T., and Sreenc, F. (2010). Rational design and construction of an efficient *E. coli* for production of diacylglycerol. *Metab. Eng.* **12**, 112–122.
- Urbanczik, R., and Wagner, C. (2005). An improved algorithm for stoichiometric network analysis: Theory and applications. *Bioinformatics* **21**, 1203–1210.
- Weinman, E. O., Srisower, E. H., and Chaikoff, I. L. (1957). Conversion of fatty acids to carbohydrate: Application of isotopes to this problem and role of the Krebs cycle as a synthetic pathway. *Physiol. Rev.* **37**, 252–272.
- Wilhelm, T., Behre, J., and Schuster, S. (2004). Analysis of structural robustness of metabolic networks. *Syst. Biol. (Stevenage)* **1**, 114–120.
- Xi, Y., Chen, Y. P. P., Cao, M., Wang, W., and Wang, F. (2009). Analysis on relationship between extreme pathways and correlated reaction sets. *BMC Bioinformatics* **10**(Suppl 1), S58, doi: 10.1186/1471-2105-10-S1-S58.
- Yeung, M., Thiele, I., and Palsson, B.Ø. (2007). Estimation of the number of extreme pathways for metabolic networks. *BMC Bioinformatics* **8**, 363, doi: 10.1186/1471-2105-8-363.

Predicting the Physiological Role of Circadian Metabolic Regulation in the Green Alga *Chlamydomonas reinhardtii*

Sascha Schäuble^{1*}, Ines Heiland^{1*}, Olga Voytsekh², Maria Mittag², Stefan Schuster¹

1 Department of Bioinformatics, Friedrich Schiller University Jena, Jena, Germany, **2** Institute of General Botany and Plant Physiology, Friedrich Schiller University Jena, Jena, Germany

Abstract

Although the number of reconstructed metabolic networks is steadily growing, experimental data integration into these networks is still challenging. Based on elementary flux mode analysis, we combine sequence information with metabolic pathway analysis and include, as a novel aspect, circadian regulation. While minimizing the need of assumptions, we are able to predict changes in the metabolic state and can hypothesize on the physiological role of circadian control in nitrogen metabolism of the green alga *Chlamydomonas reinhardtii*.

Citation: Schäuble S, Heiland I, Voytsekh O, Mittag M, Schuster S (2011) Predicting the Physiological Role of Circadian Metabolic Regulation in the Green Alga *Chlamydomonas reinhardtii*. PLoS ONE 6(8): e23026. doi:10.1371/journal.pone.0023026

Editor: Miguel A. Blazquez, Instituto de Biología Molecular y Celular de Plantas, Spain

Received: March 16, 2011; **Accepted:** July 9, 2011; **Published:** August 22, 2011

Copyright: © 2011 Schäuble et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The study was funded by the German Federal Ministry of Education and Research (BMBF project 0315260A; <http://www.bmbf.de/>; <http://www.forsys.net/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: heiland.ines@uni-jena.de

These authors contributed equally to this work.

Introduction

Metabolic pathway analysis is a well established and very useful tool in Systems Biology [1,2]. One concept in this field is that of elementary flux modes (EFMs), which represents a minimal set of reactions that can operate at steady state with all reactions proceeding in the thermodynamically feasible direction [3]. The EFM approach has proved its value in diverse biotechnological applications [4]. It has been used to find efficient routes for the production of particular target compounds, such as fatty acids in plants [5], or methionine [6] and cyanophycin [7] in bacteria, to find possible targets for the engineering of metabolic networks through knock-outs or knock-ins [8,9], as well as to assess the impact of enzyme deficiencies [10,11] or the robustness of metabolic networks [12]. Note that in contrast to optimality based approaches like Flux Balance Analysis [13], EFM analysis has the advantage of providing a more comprehensive overview of the existing routes through a given network by providing a complete data set of possible fluxes rather than solely an optimality restricted set. A disadvantage arises from the problem of combinatorial explosion [14]. Therefore, it is impossible to compute all EFMs in genome-scale models up to now, although advances have been made recently coping with large networks [15].

Beside a growing number of methods for the analysis of metabolic networks, connecting experimental data to reconstructed models remains a major task to systems biology [16–20]. However, this potential should not be underestimated, as immense data are produced by modern techniques, such as high throughput sequencing, as well as microarrays and proteomics. Moreover, inherent information in DNA sequences, like recognition motifs, can be utilised as well and ultimately applied to network analysis,

linking genomics, proteomics and metabolomics. This offers an access to regulation processes that possibly lead to altered metabolic fluxes and consequently influence the entire metabolism of an organism.

To demonstrate the usefulness of our method with a case study, we describe the analysis of a reconstructed metabolic network of nitrogen uptake in the green algae *Chlamydomonas reinhardtii*, a model process for green crop plants. Assimilating nitrogen is a key step of metabolism required by phototrophic organisms in order to grow and survive in natural habitats [21]. Nitrogen metabolism in this green algae is circadian-clock regulated, via an mRNA binding factor named CHLAMY1, a heteromer that consists of two subunits, C1 and C3, the latter being well conserved in humans [22]. This regulator is known to bind UG-repeats that comprise at least seven non-interrupted UG-repetitions and are located in the 3' UTR of various mRNAs including nitrite reductase and argininosuccinate lyase [22–24]. It has been shown experimentally that introduction of UG-repeats into the 3' UTR of reporter constructs results in circadian expression [25]. The binding activity is controlled by the circadian clock, as it increases at the end of the day and decreases again at the end of the night. As activity levels of nitrite reductase, whose mRNA bears a UG-repeat, and of reporters that are under control of the UG-repeats are highest at the beginning of the day, it is assumed that CHLAMY1 binding prevents translation during the night [25,26].

Here, we combine genome based sequence and metabolic pathway analyses by computing EFMs. This allows us to evaluate the changes in nitrogen assimilation and amino acid anabolism that are caused by CHLAMY1 binding and thus, determine the physiological role of this circadian RNA-binding factor. We study amino acid biosynthesis of alanine, glycine, asparagine, lysine and

arginine, which permits physiological interpretation and comparison to known data from other organisms. These amino acids were chosen as they are either overrepresented in *C. reinhardtii* or contain a high nitrogen content in their side chain and, thus, are particularly suitable for nitrogen storage.

As we will show, the application of optimality principles that solely focus on analysing maximum yields like in Flux Balance Analysis [13], only offers a limited view on a given system and is therefore not suitable for our approach as the complete capability of the network has to be taken into account.

Results

As it is not feasible to analyse the complete metabolism of *C. reinhardtii* using elementary flux mode analysis, we first had to confine our model. *C. reinhardtii* is able to grow either autotrophically, heterotrophically or mixotrophically. As we simulate only metabolism during the night here, we have chosen acetate and glucose-6-phosphate (G6P) as carbon sources. G6P is provided by starch degradation. The degradation is not explicitly included into the model.

To model the nitrogen uptake, we analysed the biosynthesis of five different amino acids. First, we selected amino acids that have the highest nitrogen to carbon ratio, those are lysine, asparagine and arginine. Furthermore, we analysed the amino acid composition of all predicted proteins in *C. reinhardtii* and identified glycine and alanine as most abundant and highly overrepresented amino acids compared to other organisms (Fig. 1). Additionally, glutamate, glutamine and aspartate are present in the model as intermediates.

Taken together, our reconstructed model of nitrogen metabolism of *C. reinhardtii* comprises 105 reactions and 95 metabolites. An overview is given in Fig. 2, while a complete list of reactions can be found in the Supplementary Tables S1 and S2. The sequence analysis revealed that six enzymes are entirely encoded

by mRNAs that contain $UG_{\geq 7}$ -repeats in their 3' UTRs and are hence presumably under control of CHLAMY1 (Fig. 2).

The computation of EFMs gave rise to 404252 EFMs for glycine, 684036 EFMs for alanine, 177294 EFMs for asparagine, 406560 EFMs for lysine and 1352352 EFMs for arginine biosynthesis, when G6P as well as acetate were assumed to be available. Three example EFMs are depicted in Fig. 3. The shown EFMs producing asparagine and lysine are the most efficient ones with respect to the yield of amino acids under study per mole carbon source. As for arginine, a less efficient mode is shown to reduce overlap with the other depicted modes and to show another variant, running via the pentose phosphate pathway.

Maximum carbon yields

To compare the biosynthetic yield of different amino acids, we calculated a so called carbon yield. As described in the Analysis section it was calculated based on the stoichiometric equations of EFMs. It represents the number of carbon atoms in the target amino acid divided by the number of carbon atoms in the carbon source. As beside G6P and acetate, CO_2 was the only carbon source that was set external, a carbon yield lower than 1 corresponds to a release of CO_2 during biosynthesis. In contrast, a carbon yield greater than one corresponds to a non photosynthetic incorporation of CO_2 .

We compared maximum carbon yields of EFMs for the unperturbed system and the extreme case, where the mRNAs of enzymes under control of CHLAMY1 are completely downregulated. For this analysis, we first computed all EFMs that convert one of the given carbon sources (G6P or acetate) into glycine, alanine, asparagine, lysine or arginine (Fig. 4). During a second run we removed those enzymes whose translation is potentially downregulated during the night by CHLAMY1. As argininosuccinate lyase (ASL) is encoded by an $UG_{\geq 7}$ -repeat-containing mRNA and subsequently modelled inactive, there are no EFMs for arginine synthesis left under these conditions. Furthermore,

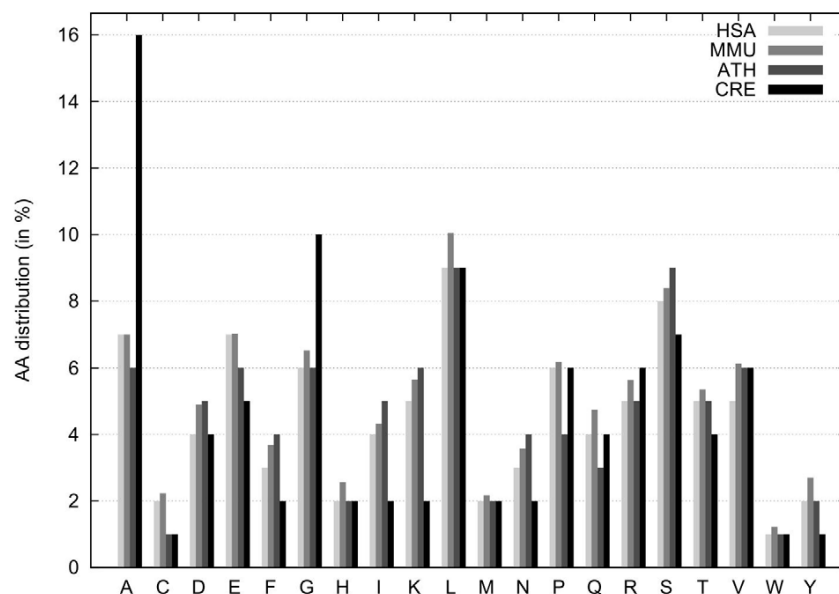


Figure 1. Distribution of amino acids among different species. Percentage share of amino acids are given for each amino acid using one letter code. The amino acid compositions of selected organisms were derived from complete genome ORF prediction from different databases (see Analysis section). HSA: *Homo sapiens*, MMU: *Mus musculus*, ATH: *Arabidopsis thaliana*, CRE: *Chlamydomonas reinhardtii*. doi:10.1371/journal.pone.0023026.g001

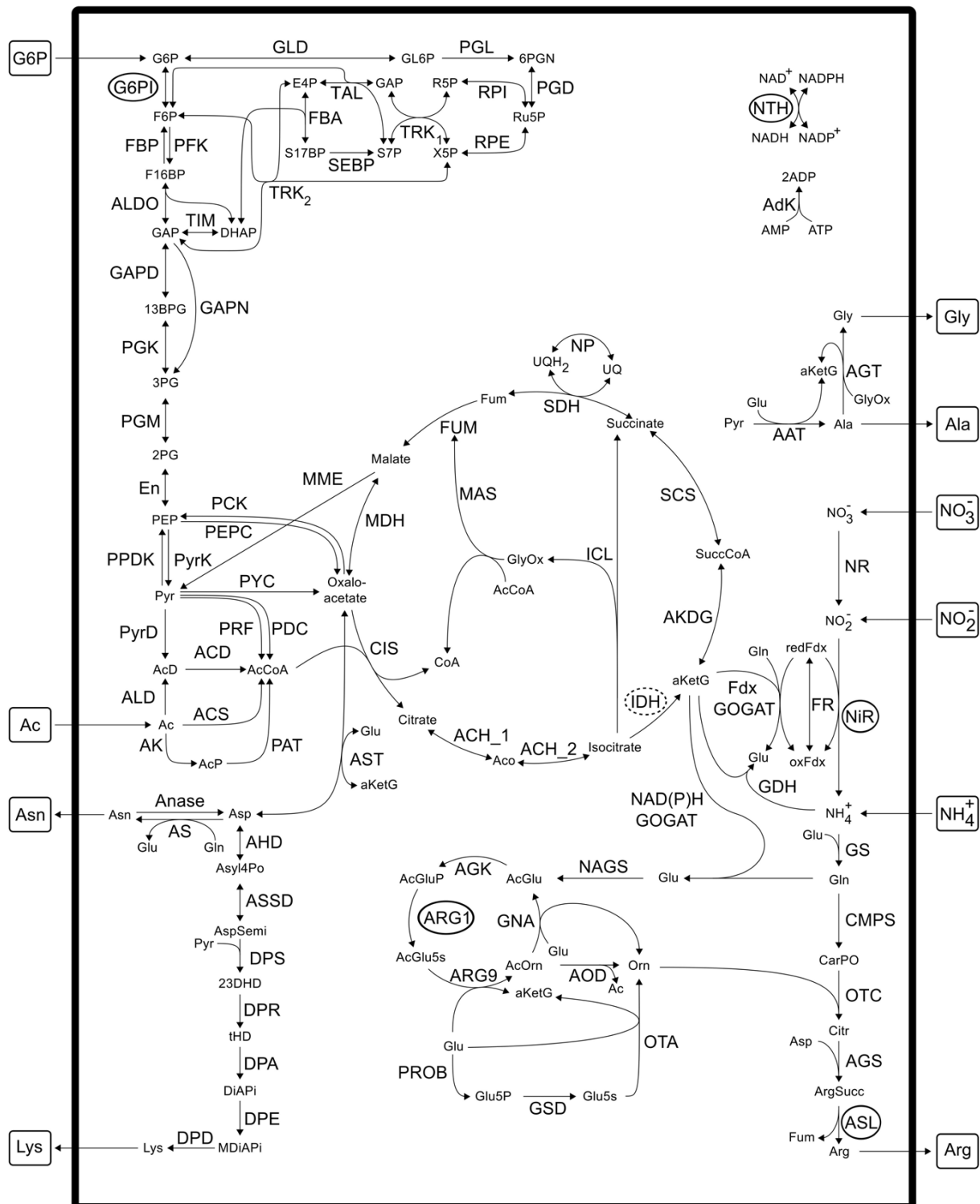


Figure 2. Overview of the reconstructed network of nitrogen metabolism in *C. reinhardtii*. Co-factors such as ATP and NAD(P)H creation or consumption, or CO₂, phosphate and water as well as the reactions of pyrophosphatase and the electron transport chain are not shown. For a list of all abbreviations, modelled reactions and species, see Supplementary Tables S1 and S2. External metabolites are framed and enzymes, whose mRNAs are downregulated by CHLAMY1 are circled. As only the NADPH dependent variant of isocitrate dehydrogenase (IDH) is affected by CHLAMY1, it is marked with a dashed circle.

doi:10.1371/journal.pone.0023026.g002

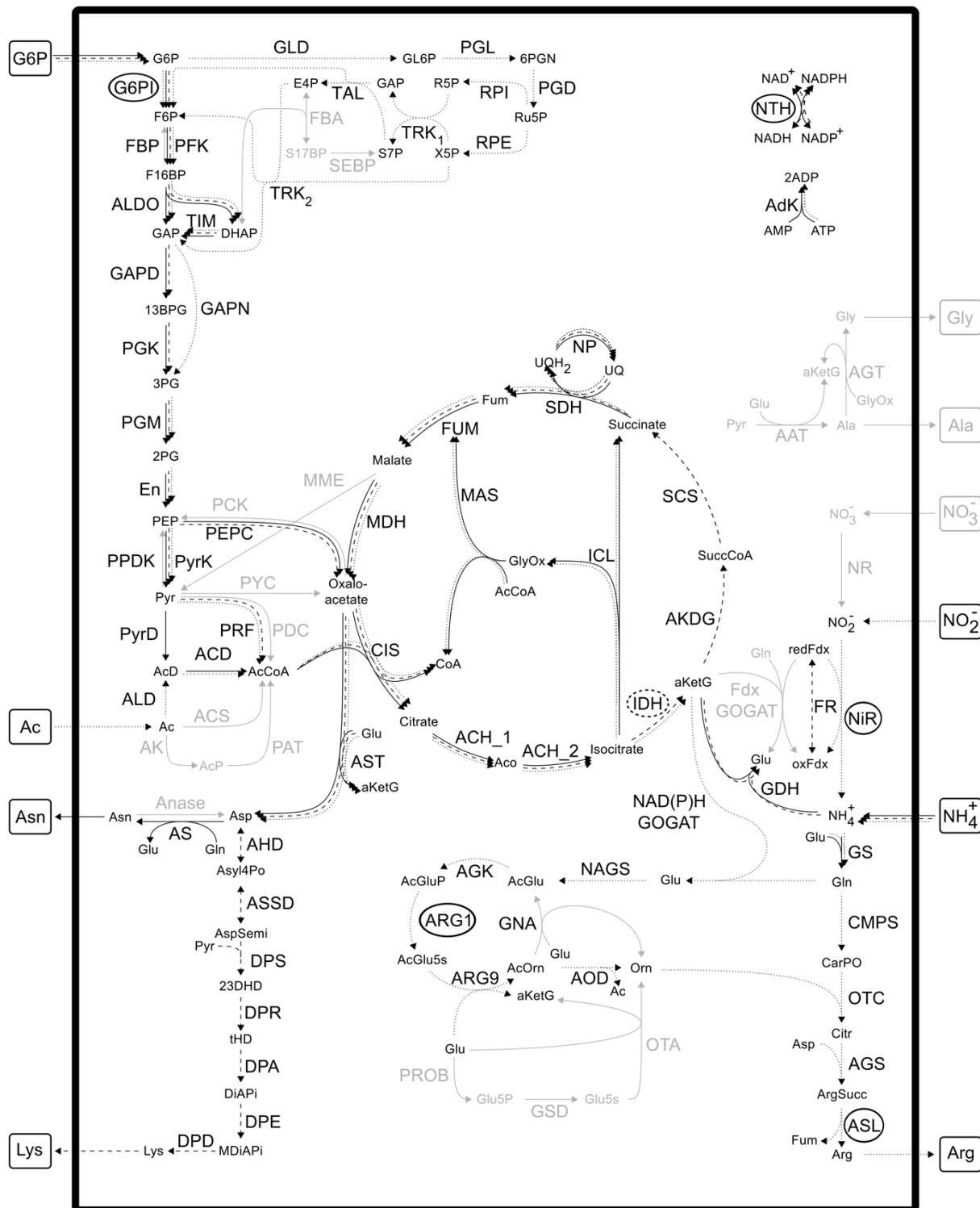


Figure 3. Three example elementary flux modes (EFMs). Solid arrows, most efficient EFM producing asparagine; dashed arrows, most efficient EFM producing lysine; dotted arrows, one selected EFM producing arginine via the pentose phosphate pathway.
doi:10.1371/journal.pone.0023026.g003

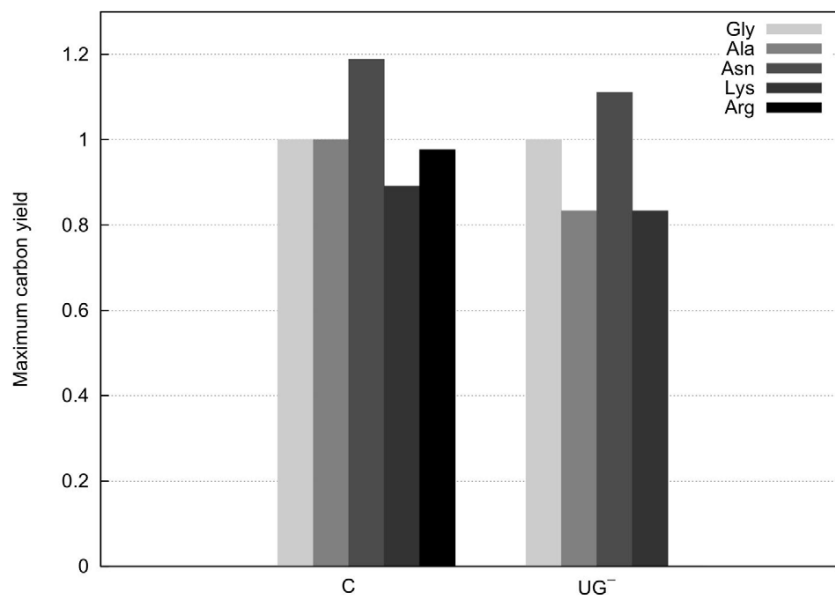


Figure 4. Comparison of maximum carbon yields that are obtained by EFM analysis. The carbon yield was calculated based on stoichiometric equations as described in the Analysis section. It represents the number of carbon atoms in the target amino acid divided by the number of carbon atoms in the carbon source. The comparison is based on ammonium uptake and all possible carbon sources (G6P and acetate). Yields correspond to two different conditions: C, the complete set of enzymes are active at normal rate; UG⁻, all CHLAMY1 regulated mRNAs and thus, related enzymes are completely inactive. In this all-or-nothing modelling approach, growth on nitrate or nitrite, as well as arginine biosynthesis, is impossible if CHLAMY1 regulation is considered, since nitrite reductase (NiR) and argininosuccinate lyase (ASL) are essential for these processes (see text and Fig. 2).

doi:10.1371/journal.pone.0023026.g004

since nitrite reductase has the same property, we do no longer find EFMs with nitrate and nitrite consumption if we assume complete downregulation of CHLAMY1 regulated mRNAs and thus, corresponding enzymes. Hence, in this case all EFMs use ammonium as sole nitrogen source.

Beside glycine, the maximum yields for biosynthesis of all amino acids are reduced, if complete downregulation by CHLAMY1 was assumed. However, analysing maximum yields only uses a very small portion of the information about the network's metabolic capabilities. In contrast, EFMs offer a more detailed view on the metabolic capacity. This significant advantage will be exploited below.

Yield distribution

To make use of the full potential of flux distribution, we first took all EFMs and respective yields into account, rather than analysing solely optimised fluxes with respect to carbon yields. Again, as ASL is a key step in arginine biosynthesis that is inactivated by CHLAMY1, we did not conduct any further analysis of arginine metabolism.

From Fig. 5 it can be observed that CHLAMY1 predominantly downregulates pathways that have lower yields. Although, the maximum yields decrease the mean carbon yields for all amino acids increase after removing EFMs affected by CHLAMY1 (Fig. 6).

Weighted influence

Note that until now all EFMs were discarded that contain at least one reaction that is under influence of CHLAMY1. However, such drastic downregulation is questionable and asks for a more realistic modelling.

In the following calculations we therefore circumvented the need to inactivate fluxes, regulated by CHLAMY1, which is usually enforced by EFM analysis. We now assume downregulation of the corresponding fluxes to 10% due to CHLAMY1 binding, rather than complete inactivation. We reduced the impact of inhibited EFMs by weighting EFMs differently, depending on whether they are under CHLAMY1 control or not (Fig. 7). The extent of downregulation by CHLAMY1 is chosen arbitrarily, as corresponding quantitative data is not available. However, reduction factors deviating slightly from 10% do not change the result qualitatively here. Note that downregulation leads to a reduced increase of the interquartile range compared to inactivation (see also Fig. 5). This is due to the large portion of CHLAMY1 controlled fluxes numbering 388832 (96.19%) for glycine, 674436 (98.6%) for alanine, 173543 (97.88%) for asparagine and 394404 (97.01%) for lysine biosynthesis, respectively (Fig. 8). However, this provides a more realistic view on the metabolic state than complete downregulation.

Additionally, we calculated the mean of molar yields by using the formula for weighted means given by Eq. (3). This enables us to weight every derived yield and hence, also the underlying flux, represented by its respective EFM. We applied a weight of 10% to EFMs affected by CHLAMY1 and a unity weight to all remaining fluxes. The resulting mean yield is considerably increased upon CHLAMY1 binding (Fig. 9).

Discussion

In this study, we have outlined a method that interconnects sequence based knowledge with metabolic pathway analysis. The method has been illustrated by nitrogen metabolism of *C. reinhardtii*, which is under the control of the circadian clock. The

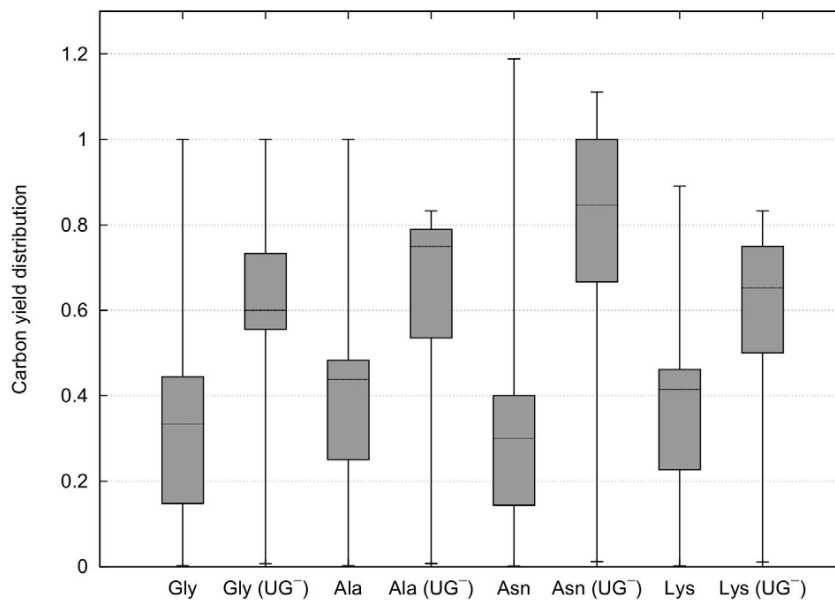


Figure 5. Carbon yield distribution considering complete inactivation. Box plots (with whiskers ranging from minimum to maximum and thick solid line indicating the median) of carbon yield distribution for glycine (Gly), alanine (Ala), asparagine (Asn) and lysine (Lys) associated pathways based on all available carbon sources. Complete inactivation of CHLAMY1 affected reactions is considered here. Knockout of ASL completely inhibits arginine biosynthesis and is hence, not shown (see also Fig. 2). If not marked with UG⁻, boxplots show complete EFM distribution. Otherwise, they show distribution for all EFMs that are not affected by CHLAMY1 downregulation.
doi:10.1371/journal.pone.0023026.g005

calculated elementary flux modes provide a data set that is well-suited for quantifying and understanding the complex architecture of this network. The large number of modes (e. g. 1352352 for

arginine) point to a considerable redundancy of this network. Intriguingly, our results show that downregulation of circadian controlled enzymes improves carbon distribution and thus,

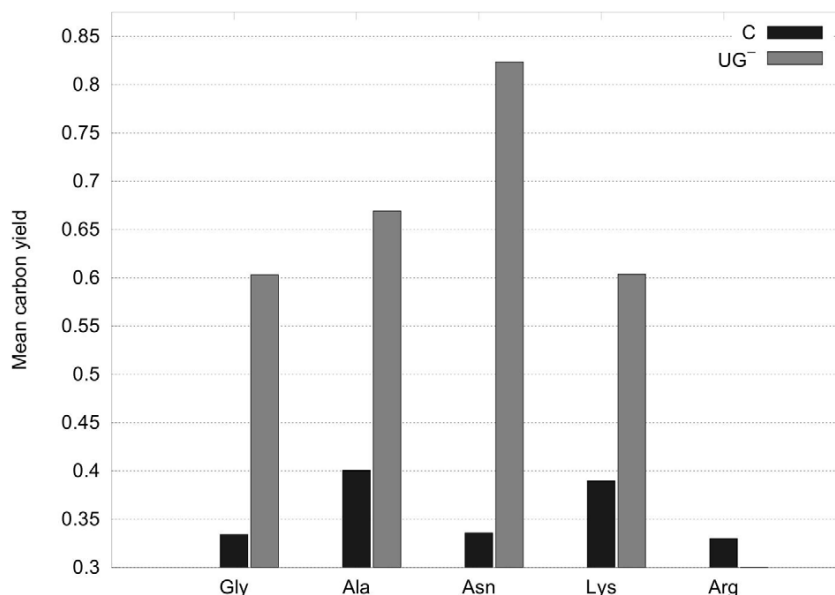


Figure 6. Mean yields assuming complete downregulation by CHLAMY1. In contrast to Fig. 4, the carbon yields of either all EFMs (C) or all those EFMs that do not have CHLAMY1 regulated mRNAs (UG⁻) were calculated here. The sum of these yields divided by the number of corresponding EFMs results in the mean yield shown. An increase of the mean yield after downregulation by CHLAMY1 can be observed for all amino acids, except for arginine, as in this case CHLAMY1 downregulates expression of ASL, which is crucial for the arginine pathway (see also Fig. 2).
doi:10.1371/journal.pone.0023026.g006

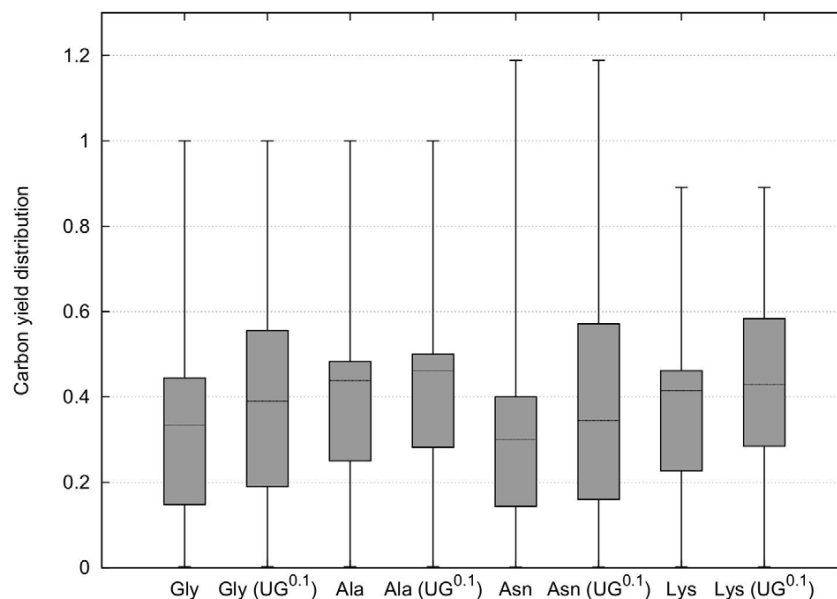


Figure 7. Carbon yield distribution considering partial downregulation. Weighted mean value of carbon yield distribution for the same pathways as in Fig. 5 based on all available carbon and nitrogen sources, considering downregulation of CHLAMY1 affected enzymes to 10%. Either all enzymes are active at normal rate or CHLAMY1 is assumed to downregulate mRNAs with UG_{≥7}-repeat-motif to 10% activity (UG^{0.1}) leading subsequently to reduced yield contribution of the affected EFMs. doi:10.1371/journal.pone.0023026.g007

decreases energy consumption. The somewhat counter-intuitive result that knocking out or downregulating several enzymes may lead to an increase in average yield arises, because poor pathways are deleted or downregulated, so that more efficient pathways become more dominant. A similar phenomenon was observed earlier in the context of strain optimisation [9,27,28]. Our approach focuses on analysing altered fluxes due to regulatory influences in general. We compare two or more physiological situations (e. g. day- and night-time) rather than manipulated setups. Other examples may be provided by hibernation vs. summertime stage or different developmental stages such as embryonic vs. adult. The regulatory information can be provided in a wide variety of forms, including transcriptional regulatory events as time-dependent constraints [29].

Additionally, we have considered information derived from sequence data. Since it is known that many regulatory proteins bind to specific motifs in the mRNA or to promoters, such information is extremely useful in modelling regulation of metabolism. Moreover, measuring fluxes in detail is already a demanding task for simple model organisms, like *Escherichia coli* or *Saccharomyces cerevisiae*, but might be virtually infeasible for higher organisms, when regulatory complexity becomes more sophisticated. Thus, our approach proves to be an easy-to-use, helpful method to determine the type and impact of influences of regulatory factors.

Considering only the maximum carbon yields of pathways summarised in Fig. 4, indicates that *C. reinhardtii* remains able to synthesise glycine, alanine, asparagine and lysine but with reduced theoretical effectiveness while not being able to synthesise arginine if one assumes complete downregulation by CHLAMY1 at night-time. However, as there are more than three million possible routes within the network producing the target amino acids and the main portion of all EFMs (above 96% for all amino acids, see also Fig. 8) is affected by CHLAMY1 action, solely focusing on

maximum carbon yields provides a limited view and would lead to misinterpretations. Furthermore, the calculation of the maximal yield is sensitive to the size of the model and the carbon sources chosen. If we use glyceraldehyde-3-phosphate (GAP) and acetate as carbon source and thus, remove glycolysis and the pentose phosphate pathway from the model, the maximum yield does not change between sets of EFMs with and without CHLAMY1 affected reactions (see Fig. 10).

To study the spectrum of metabolic capabilities, we analysed the whole yield distribution. The results, shown in Fig. 5, reveal that CHLAMY1 influences the mRNA expression of enzymes mainly taking part in EFMs that realise low yields. Thus, translational downregulation by CHLAMY1 during the night leads to an increased median yield for the considered amino acid production whereas the maximum yield decreases.

During night-time, photosynthesis is impossible and, hence, energy is largely limited. A prohibition of energy-consuming reactions that usually contribute to low carbon yields during the night has already been observed experimentally for *Arabidopsis thaliana* [30]. The decrease in maximum carbon yield observed in our analysis is mainly due to the fact that G6PI is regulated by CHLAMY1 and thus, G6P is forced to enter the pentose phosphate pathway (PPP). This might be necessary as the PPP is required for the synthesis of nucleotides. As DNA-replication occurs preferentially during the night, this regulatory compromise can be considered as an optimised outcome of evolution.

Taken together, our results are in good agreement with experimental observations and evolutionary considerations. In contrast to the dependency of the decrease of the maximum yields on the model size and carbon source chosen, the increase of the yield distribution can be found for both G6P and GAP as carbon source (see Fig. 5 and Fig. 10, respectively).

Beside ASL and NiR, CHLAMY1 regulated enzymes are identified based on UG-repeats found in the annotated 3' UTR of

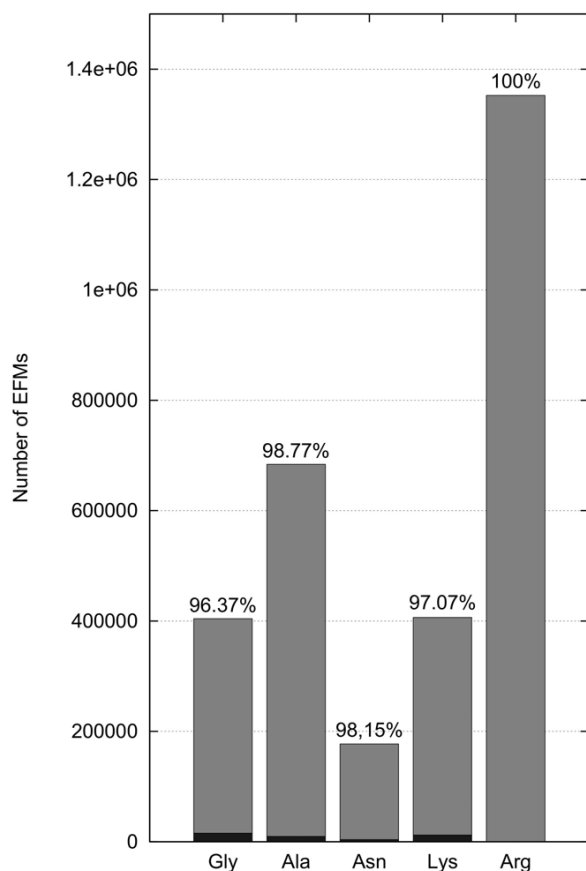


Figure 8. Number of EFMs with and without circadian regulation. More than 96% of all elementary flux modes (gray) are influenced by CHLAMY1. Remaining elementary flux modes, assuming complete downregulation by CHLAMY1, are shown in black.
doi:10.1371/journal.pone.0023026.g008

the respective genes. For ASL and NiR CHLAMY1 binding has been shown experimentally and the introduction of the respective 3' UTR sequences into luciferase constructs lead to a robust circadian enzyme activity [25]. Furthermore, NiR activity has been shown to cycle in circadian manner [26]. To analyse how the prediction of CHLAMY1 regulated enzymes based on sequence analysis might influence our results we calculated the yield distribution with only ASL and NiR downregulated by CHLAMY1 for comparison. As Fig. 11 shows, NiR and ASL are the enzymes mainly contributing to an increased yield. As G6PI is still active in this case the maximum yield is equal to that of the complete model. This again demonstrates that the calculation of the maximum yield alone is relatively sensitive to changes in the model and may lead to misinterpretations.

The analysis presented here was restricted to the influence of CHLAMY1 regulation on metabolism. There might be other processes influencing the circadian regulation of nitrogen metabolism like transcriptional and posttranslational regulation. It has been described that the transcription of some enzymes included in our model are regulated in a circadian manner [31]. However, all enzymes described in the aforementioned approach have isoenzymes that are not under the control of the circadian clock. As we did not distinguish between different isoforms as long as they use

the same cofactors, an inclusion of transcriptional regulation would not affect our simulations. Furthermore, we did not include any compartmentalisation in our model, as due to the resulting complexity the calculation of all EFMs would not be feasible.

Here, we have assumed that all fluxes contribute with equal probability to an overall flux, as done earlier in the case of incomplete knowledge [11]. This assumption probably does not describe reality properly. However, it allows one to analyse the robustness and full flexibility against altered environmental conditions. Moreover, it enables us to predict qualitative changes of the metabolic system under investigation. Furthermore, it has been noticed that approaches based on optimality principles are dependent on the applied constraints [13,32–34] and matching them to experimental results meets with various difficulties [35,36]. As we have shown in this study, weighting EFMs affected by regulating factors differently from unaffected EFMs, preserves the EFM inherent yield, while changing the overall yield distribution (Fig. 7). Additionally, computing a weighted arithmetic mean of all carbon yields provides valuable information about effects of the regulating factors, while circumventing artificial all-or-none simulations. The simplicity of this approach provides the advantage that no parameters, like reaction rates, are required and no additional assumption have to be made.

Further analysis of the calculated EFMs shows that only approximately 1/50th of the original set of EFMs is still fully active after CHLAMY1 binding. Therefore, the metabolic flux through the system is considerably reduced during the night-time, which is in line with the reduction of carbon and energy consumption when photosynthesis is inactive. This holds independently of the carbon source chosen. Particularly EFMs with a low yield are suppressed, so that the average yield increases. If CHLAMY1 binding is reduced at the end of the night resulting in the expression of target enzymes at the beginning of the day when photosynthetic energy is again available, the metabolic capability and robustness of nitrogen metabolism is greatly increased and allows fast incorporation of nitrogen into the organism. As energy is no longer limiting, there is no need to restrict to those reactions with high yields and low energy consumption. Therefore, CHLAMY1 binding during the night appears to ensure energy conservation while still allowing nitrogen fixation. Due to the stabilisation of mRNA by CHLAMY1 and release at the end of the night [23,25,26], it furthermore enables a high metabolic capacity as soon as enough energy is available.

Fig. 7 reveals that downregulation rather than inactivation of CHLAMY1 affected reactions, still leads to an increase in global carbon yields, although the increase is remarkably lower. This is due to the large portion of CHLAMY1 influenced fluxes.

In general, using weighted influences instead of the simplified all-or-none approach, can be used to study the impact of two regulators leading to different residual activity of enzymes. Furthermore, it could also be used to interpret microarray or other expression data. Here, fold changes could be used as weighting factors to simulate metabolic changes of a given system. Hence, it provides a useful tool to connect the growing amount of high throughput expression data to pathway analysis.

Analysis

Calculating amino acid composition

The amino acid compositions of selected organisms were derived from complete genome open reading frame (ORF) prediction data in fasta file format. The fasta files from *Homo sapiens*, *Mus musculus*, and *Arabidopsis thaliana* were obtained from the UniProt database [37], while the fasta file for *Chlamydomonas*

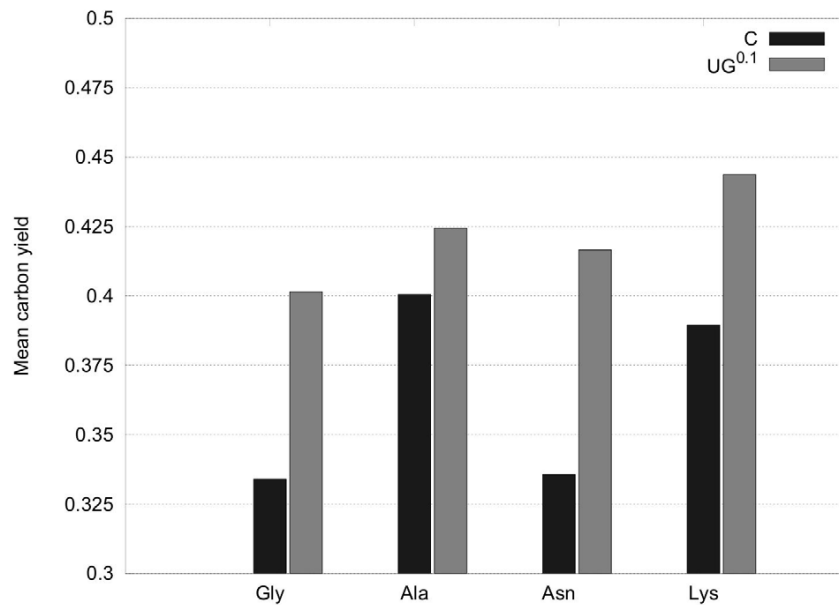


Figure 9. Mean yields assuming partial downregulation of genes by CHLAMY1. All EFM's have been used to calculate a weighted mean according to Eq. 3. An increase of the weighted mean yield can again be observed for all amino acids as in Fig. 6. However, the increase is less pronounced.

doi:10.1371/journal.pone.0023026.g009

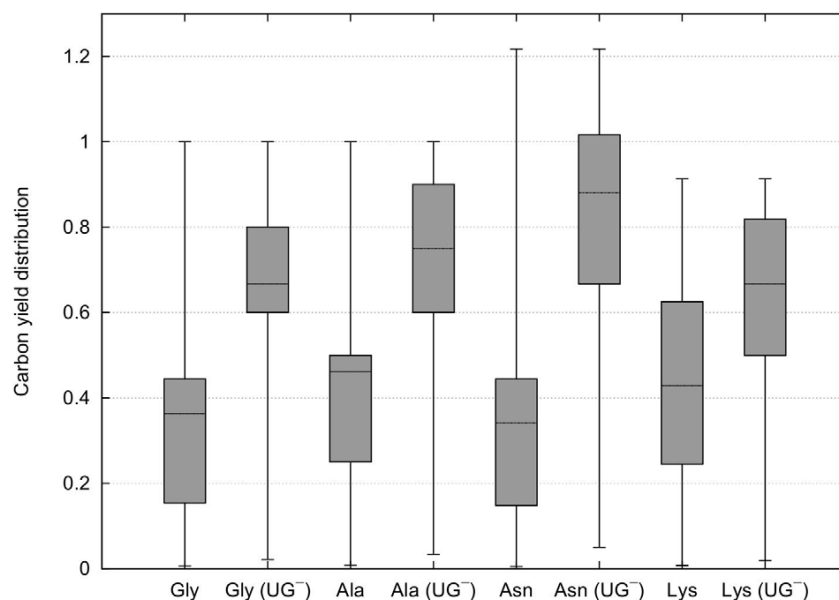


Figure 10. Yield distribution for GAP and acetate as carbon source. Carbon yield distribution was calculated assuming complete inactivation. In contrast to Fig. 5 GAP and acetate were used as carbon source and the maximum yields (upper whiskers) do not change between EFM sets including all enzyme or those that are not affected by CHLAMY1 downregulation (UG⁻). Box plots (with whiskers ranging from minimum to maximum and thick solid line indicating the median) of carbon yield distribution for glycine (Gly), alanine (Ala), asparagine (Asn) and lysine (Lys) associated pathways based on all available carbon sources.

doi:10.1371/journal.pone.0023026.g010

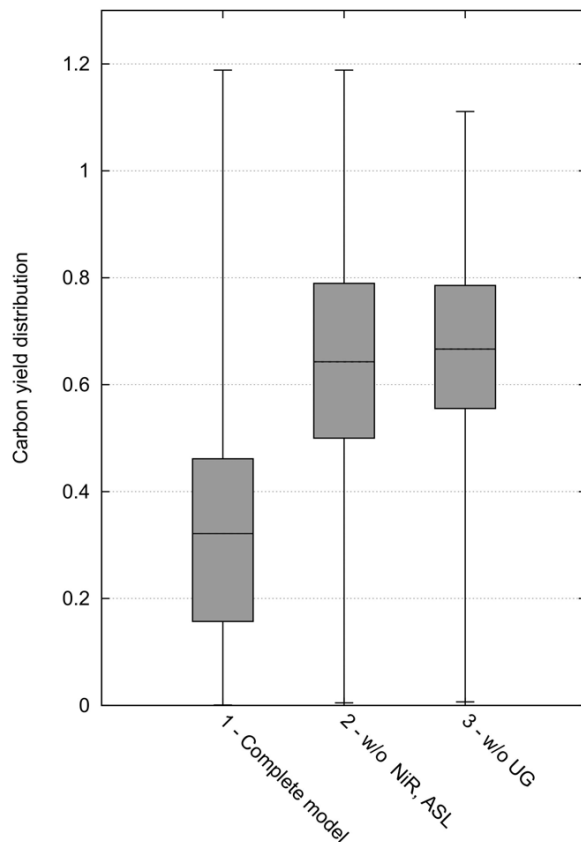


Figure 11. Combined carbon yield distribution for complete downregulation by CHLAMY1. For the depicted boxplots the yield distribution for all amino acids including arginine have been combined into one yield distribution. Box plots (with whiskers ranging from minimum to maximum and thick solid line indicating the median) of carbon yield distribution are based on all available carbon and nitrogen sources. Complete inactivation of CHLAMY1 affected enzymes is considered here.
doi:10.1371/journal.pone.0023026.g011

reinhardtii was fetched from a database provided by the Joint Genome Institute [38]. These files were scanned for total amino acid distribution and the results summarised in Fig. 1.

Pathway reconstruction

A metabolic network comprising the nitrogen metabolism in *C. reinhardtii* (Fig. 2) was reconstructed using the KEGG [39] and ChlamyCyc [40] databases, the biochemical pathways textbook [41] as well as bibliomic data. All reactions were manually curated, which included mass balancing if required and verification of reaction reversibility based on existing biochemical knowledge. If the irreversibility was not conclusive, we set the corresponding reaction reversible.

The carbohydrate metabolism under study includes glycolysis, gluconeogenesis, the pentose phosphate pathway, acetate uptake, the citrate cycle and the glyoxylate shunt. The nitrogen uptake model was reconstructed using data from [42–45]. Moreover, the biosynthetic pathways of glycine, alanine, asparagine, lysine and arginine, which provide the target metabolites of the model, are taken into account by comparing charts [41] with the above-mentioned databases and biological literature. The accessible

carbon sources are acetate, simulating heterotrophic growth, and glucose-6-phosphate (G6P), resulting from starch breakdown during the night. Moreover, molecular nitrogen is provided by nitrate, nitrite or ammonium uptake. Consequently, those substances as well as G6P, acetate and the five above-mentioned amino acids are modelled as external metabolites, that is, their concentrations are considered to be buffered. In contrast, we modelled all energy and redox carriers, such as ATP, NAD(P)H and ferredoxin, as internal. The network in SBML A SBML version of the network is provided in the Supplements.

Sequence analysis

The mRNA sequences that are associated to the enzymes included in the model were analysed for perfect $UG_{\geq 7}$ -repeats (UG UG UG) in annotated 3' UTR [22] of models from the Joint Genome Institute database version 4.0 of *C. reinhardtii* [38].

Special emphasis had to be put on isoenzymes, as in several cases mRNAs encoding enzymes contained $UG_{\geq 7}$ -repeats, while others associated to enzymes catalysing the same reaction did not. As we did not regard localisation of enzymes and did not distinguish between isoenzymes as long as they use the same cofactors, the corresponding reactions were simulated not to be under control of the circadian clock via CHLAMY1 for the EFM analysis.

Computation of elementary flux modes

EFMs were computed with efmtool [46] inside the MATLAB environment, version 2008b (The MathWorks, Natick, MA, USA). Details of elementary flux mode calculation are described elsewhere [46,47].

Calculation of yields

In order to compare the effectiveness of different modelled amino acid pathways, we computed carbon yields according to:

$$\text{carbon yield} = \frac{\#C(aa)}{\#C(G6P + \text{acetate})}, \quad (1)$$

where $\#C(aa)$ and $\#C(G6P + \text{acetate})$ refer to the numbers of carbon atoms in the considered target amino acid (*aa*) and in the substrates *G6P* and *acetate*, respectively. The number of carbon atoms were obtained from the overall chemical equation of each elementary mode. The mean yields \bar{y} of all yields y_i were calculated according to standard formula for mean calculation:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad (2)$$

Weighted yields

To calculate the effect of CHLAMY1 downregulation rather than full inactivation of the influenced enzymes, yields from EFMs were weighted differently in the resulting yield distribution. We arbitrarily assumed downregulation to 10% as experimental measures for the degree of downregulation are not available. For the calculation of the yield distribution and visualisation in boxplot graphics the yields unaffected by CHLAMY1 were counted ten times, whereas yields affected by that regulator were only taken into account once.

Additionally, we calculated weighted mean yields for a simplified visualisation of the downregulating effect of CHLAMY1. To do so we computed the weighted arithmetic

mean yield \bar{y}_w according to the following equation:

$$\bar{y}_w = \frac{\sum_i w_i * y_i}{\sum_i w_i}, \quad (3)$$

with w_i being the weight of the derived carbon yield y_i . These weighting factors can be defined in various ways. Here, we use the definition given above based on the fractional extent of downregulation.

Supporting Information

Table S1 Overview of modelled metabolites and corresponding abbreviations. (PDF)

Table S2 Overview of modelled enzymes and corresponding EC-numbers, abbreviations as well as JGI database IDs (cre v4.0;

References

- Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BO (2009) Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol* 7: 129–143.
- Ruppin E, Papin JA, de Figueiredo LF, Schuster S (2010) Metabolic reconstruction, constraint-based analysis and game theory to probe genome-scale metabolic networks. *Curr Opin Biotechnol* 21: 502–510.
- Schuster S, Hilgetag C (1994) On elementary ux modes in biochemical reaction systems at steady state. *J Biol Syst* 2: 165–182.
- Trinh CT, Wlaschin A, Sreenc F (2009) Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Appl Microbiol Biotechnol* 81: 813–826.
- Schwender J, Goffman F, Ohlrogge JB, Shachar-Hill Y (2004) Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. *Nature* 432: 779–782.
- Krömer JO, Wittmann C, Schröder H, Heinzle E (2006) Metabolic pathway analysis for rational design of L-methionine production by *Escherichia coli* and *Corynebacterium glutamicum*. *Metab Eng* 8: 353–369.
- Diniz SC, Voss I, Steinbüchel A (2006) Optimization of cyanophycin production in recombinant strains of *Pseudomonas putida* and *Ralstonia eutropha* employing elementary mode analysis and statistical experimental design. *Biotechnol Bioeng* 93: 698–717.
- Carlson R, Fell D, Sreenc F (2002) Metabolic pathway analysis of a recombinant yeast for rational strain development. *Biotechnol Bioeng* 79: 121–134.
- Trinh CT, Sreenc F (2009) Metabolic engineering of *Escherichia coli* for efficient conversion of glycerol to ethanol. *Appl Environ Microbiol* 75: 6696–6705.
- Schuster S, Kenanov D (2005) Adenine and adenosine salvage pathways in erythrocytes and the role of S-adenosylhomocysteine hydrolase. A theoretical study using elementary ux modes. *FEBS J* 272: 5278–5290.
- Kenanov D, Kaleta C, Petzold A, Hoischen C, Diekmann S, et al. (2010) Theoretical study of lipid biosynthesis in wild-type *Escherichia coli* and in a protoplast-type L-form using elementary ux mode analysis. *FEBS J* 277: 1023–1034.
- Behre J, Wilhelm T, von Kamp A, Ruppin E, Schuster S (2008) Structural robustness of metabolic networks with respect to multiple knockouts. *J Theor Biol* 252: 433–441.
- Varma A, Palsson BO (1994) Stoichiometric ux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl Environ Microbiol* 60: 3724–3731.
- Acua V, Marchetti-Spaccamela A, Sagot MF, Stougie L (2010) A note on the complexity of finding and enumerating elementary modes. *Biosystems* 99: 210–214.
- Kaleta C, de Figueiredo LF, Schuster S (2009) Can the whole be less than the sum of its parts? Pathway analysis in genome-scale metabolic networks using elementary ux patterns. *Genome Res* 19: 1872–1883.
- Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED (2002) Metabolic network structure determines key aspects of functionality and regulation. *Nature* 420: 190–193.
- Covert MW, Knight EM, Reed JL, Herrgård MJ, Palsson BO (2004) Integrating high-throughput and computational data elucidates bacterial networks. *Nature* 429: 92–96.
- Schwarz R, Liang C, Kaleta C, Kühnel M, Ho_mann E, et al. (2007) Integrated network reconstruction, visualization and analysis using YANAsquare. *BMC Bioinformatics* 8: 313.
- Schwartz JM, Gauguier C, Nacher JC, de Daruvar A, Kanehisa M (2007) Observing metabolic functions at the genome scale. *Genome Biol* 8: R123.
- Shlomi T, Cabili MN, Herrgård MJ, Palsson BO, Ruppin E (2008) Network-based prediction of human tissue-specific metabolism. *Nat Biotechnol* 26: 1003–1010.
- Daniel-Vedele F, Filleur S, Caboche M (1998) Nitrate transport: a key step in nitrate assimilation. *Curr Opin Plant Biol* 1: 235–239.
- Zhao B, Schneid C, Iliev D, Schmidt EM, Wagner V, et al. (2004) The circadian RNA-binding protein CHLAMY 1 represents a novel type heteromer of RNA recognition motif and lysine homology domain-containing subunits. *Eukaryot Cell* 3: 815–825.
- Mittag M (1996) Conserved circadian elements in phylogenetically diverse algae. *Proc Natl Acad Sci U S A* 93: 14401–14404.
- Waltenberger H, Schneid C, Grosch JO, Bareiss A, Mittag M (2001) Identification of target mRNAs for the clock-controlled RNA-binding protein Chlamy 1 from *Chlamydomonas reinhardtii*. *Mol Genet Genomics* 265: 180–188.
- Kiaulehn S, Voytsekh O, Fuhrmann M, Mittag M (2007) The Presence of UG-repeat sequences in the 3'-UTRs of reporter luciferase mRNAs mediates circadian expression and can determine acrophase in *Chlamydomonas reinhardtii*. *J Biol Rhythms* 22: 275–277.
- Iliev D, Voytsekh O, Schmidt EM, Fiedler M, Nykytenko A, et al. (2006) A heteromeric RNA-binding protein is involved in maintaining acrophase and period of the circadian clock. *Plant Physiol* 142: 797–806.
- Wlaschin AP, Trinh CT, Carlson R, Sreenc F (2006) The fractional contributions of elementary modes to the metabolism of *Escherichia coli* and their estimation from reaction entropies. *Metab Eng* 8: 338–352.
- Hädicke O, Klamt S (2010) CASOP: a computational approach for strain optimization aiming at high productivity. *J Biotechnol* 147: 88–101.
- Covert MW, Schilling CH, Palsson BO (2001) Regulation of gene expression in ux balance models of metabolism. *J Theor Biol* 213: 73–88.
- Piques M, Schulze WX, Höhne M, Usadel B, Gibon Y, et al. (2009) Ribosome and transcript copy numbers, polysome occupancy and enzyme dynamics in *Arabidopsis*. *Mol Syst Biol* 5: 314.
- Kucho KI, Okamoto K, Tabata S, Fukuzawa H, Ishiura M (2005) Identification of novel clockcontrolled genes by cDNA macroarray analysis in *Chlamydomonas reinhardtii*. *Plant Mol Biol* 57: 889–906.
- Edwards JS, Ibarra RU, Palsson BO (2001) In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat Biotechnol* 19: 125–130.
- Ibarra RU, Edwards JS, Palsson BO (2002) *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* 420: 186–189.
- Famili I, Forster J, Nielsen J, Palsson BO (2003) *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proc Natl Acad Sci U S A* 100: 13134–13139.
- Schuster S, Pfeiffer T, Fell DA (2008) Is maximization of molar yield in metabolic networks favoured by evolution? *J Theor Biol* 252: 497–504.
- Feist AM, Palsson BO (2010) The biomass objective function. *Curr Opin Microbiol*.
- UniProt Consortium (2010) The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res* 38: D142–D148.
- Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, et al. (2007) The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* 318: 245–250.
- Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, et al. (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res* 36: D480–D484.
- May P, Christian JO, Kempa S, Walther D (2009) ChlamyCyc: an integrative systems biology database and web-portal for *Chlamydomonas reinhardtii*. *BMC Genomics* 10: 209.
- Michal G, ed. (1999) *Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology* John Wiley & Sons, 1st edition 1999 edition.

Predicting the Role of Circadian Regulation

42. Quesada A, Galván A, Schnell RA, Lefebvre PA, Fernández E (1993) Five nitrate assimilation-related loci are clustered in *Chlamydomonas reinhardtii*. *Mol Gen Genet* 240: 387–394.
43. González-Ballester D, Camargo A, Fernández E (2004) Ammonium transporter genes in *Chlamydomonas*: the nitrate-specific regulatory gene *Nit2* is involved in *Amt1*;1 expression. *Plant Mol Biol* 56: 863–878.
44. González-Ballester D, de Montaigu A, Higuera JJ, Galván A, Fernández E (2005) Functional genomics of the regulation of the nitrate assimilation pathway in *Chlamydomonas*. *Plant Physiol* 137: 522–533.
45. Fernández E, Galván A (2007) Inorganic nitrogen assimilation in *Chlamydomonas*. *J Exp Bot* 58: 2279–2287.
46. Terzer M, Stelling J (2008) Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics* 24: 2229–2235.
47. Schuster S, Fell DA, Dandekar T (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat Biotechnol* 18: 326–332.

Research Article

Metabolic costs of amino acid and protein production in *Escherichia coli*

Christoph Kaleta¹, Sascha Schäuble¹, Ursula Rinas^{2,3} and Stefan Schuster⁴¹ Research Group Theoretical Systems Biology, Friedrich Schiller University Jena, Jena, Germany² Helmholtz Centre for Infection Research, Braunschweig, Germany³ Institute of Technical Chemistry - Life Science, Leibniz University of Hannover, Hannover, Germany⁴ Department of Bioinformatics, Friedrich Schiller University Jena, Jena, Germany

Escherichia coli is the most popular microorganism for the production of recombinant proteins and is gaining increasing importance for the production of low-molecular weight compounds such as amino acids. The metabolic cost associated with the production of amino acids and (recombinant) proteins from glucose, glycerol and acetate was determined using three different computational techniques to identify those amino acids that put the highest burden on the biosynthetic machinery of *E. coli*. Comparing the costs of individual amino acids, we find that methionine is the most expensive amino acid in terms of consumed mol of ATP per molecule produced, while leucine is the most expensive amino acid when taking into account the cellular abundances of amino acids. Moreover, we show that the biosynthesis of a large number of amino acids from glucose and particularly from glycerol provides a surplus of energy, which can be used to balance the high energetic cost of amino acid polymerization.

Received	30 NOV 2012
Revised	05 APR 2013
Accepted	28 MAY 2013
Accepted article online	07 JUN 2013

Supporting information
available online

Keywords: Amino acid biosynthetic cost · Amino acid production · *Escherichia coli* metabolism · Protein production cost · Recombinant proteins

1 Introduction

In a technological setup, many biological products are currently being synthesized using bacteria such as *Escherichia coli*. *E. coli* is the most common bacterial host employed for the production of recombinant proteins that are required for pharmaceutical and technical applications but also for research purposes [1–3] and is one of the workhorses of synthetic biology [4, 5]. In addition, *E. coli* emerges as an important producer of amino acids utilized as food additives or as components of cosmetics and

pharmaceuticals [6–8]. Thus, a deeper understanding of the metabolic costs of amino acid and protein production and the characteristics of the corresponding pathways is crucial for the rational design of producer strains.

Using defined medium with glucose or especially with glycerol as carbon substrate, *E. coli* can be grown to high cell densities when cells are kept under balanced growth conditions with reduced carbon supply [9–12]. When grown under unbalanced conditions with carbon substrate excess, acetate formation can occur, which if accumulating to toxic concentrations, can prevent further growth and product formation [13–15]. However, acetate at moderate concentrations can also be utilized as carbon source [15]. Thus, glucose, glycerol, and acetate are major starting carbon compounds, which are transformed into amino acids and proteins.

While pathways for synthesizing individual amino acids are biochemically well-defined in textbooks [16], they usually cover only a small portion of reactions that are required to convert a source compound such as glucose into a certain product. However, a pathway has to be con-

Correspondence: Jun.-Prof. Dr. Christoph Kaleta, Research Group Theoretical Systems Biology, Friedrich Schiller University Jena, Leutragraben 1, D-07743 Jena, Germany
E-mail: Christoph.Kaleta@uni-jena.de

Abbreviations: Acetyl-CoA, acetyl coenzyme A; ATP, adenosine triphosphate; FADH, flavin adenine dinucleotide hydrate; LP, linear programming; Mthf, methylenetetrahydrofolate; NAD(P)H, nicotinamide adenine dinucleotide (phosphate) hydrate

sidered in the context of the entire network to account also for precursor production or balancing of co-factors and, thus, to allow for an unbiased characterization and comparison with other pathways.

In this work, we analyze amino acid and protein synthesis by determining and characterizing flux distributions of all required reactions using three different methods and evaluating three different measures for cost. We start with a “manual” computation of amino acid biosynthetic cost from central metabolic precursors, as done by Stouthamer et al. [17] and Akashi and Gojobori [18]. We update and correct the values given by Stouthamer in the light of new knowledge of metabolic pathways in *E. coli*.

We proceed by using two different linear programming (LP) methods, which allow a computation of flux distributions with maximal yield (product-to-substrate ratio) [19] in a genome-scale metabolic model of *E. coli* [20]. In the first method, the “LP standard” method, we assume that flux distributions have to balance energy currency metabolites such as adenosine triphosphate (ATP) and nicotinamide adenine dinucleotide (phosphate) hydrate [NAD(P)H]. We relax this requirement in the second “LP unlimited energy” method formulation where it is assumed that the cells have an excess of energy [ATP (H^+)] and free electrons available. In the following, we will distinguish between these three methods by using the terms “manual”, “LP standard” and “LP unlimited energy”, respectively. Computing the measures of ATP consump-

tion, carbon yield, and per kilodalton cost for flux distributions and comparing them across different amino acids allows to draw important conclusions on the characteristics of amino-acid biosynthetic pathways. The different cost measures that we compute reflect different aspects of the biosynthetic routes as well as the carbon sources. While yield is strongly influenced by the molecular mass of a carbon source, ATP consumption is influenced by the energetic content of a carbon source and carbon yield by the position at which a carbon source enters central metabolism. Linear programming has been previously used to compute maximal yields of different amino acids in *E. coli* using reduced models available at that time [21, 22]. We have extended these studies using a genome-scale metabolic model leading to more accurate data from linear programming and determined the metabolic cost of protein synthesis for which we additionally need to consider the cost of polymerizing amino acids into proteins. These calculations can serve as a starting point to design improved media for the production of amino acids and proteins since they indicate the metabolic cost of the production of individual amino acids and entire proteins for the cell.

2 Materials and methods

2.1 Data

As metabolic model, we used the genome-scale reconstruction of *E. coli* metabolism iAF1260 [20]. For information on how we modified the metabolic network for our calculations see Supporting information, Text S1. For the manual determination of flux distributions for amino acid production we used pathways as displayed in the metabolic maps of *E. coli* of the BiGG database [23] as reference. Data on the sequence of proteins and mass of individual amino acids were taken from EcoCyc [24]. For details on the computation of the synthesis cost of all amino acids present in the cell see Supporting information, Text S2.

2.2 Manual computation of amino acid costs

Precursors of individual amino acids were assumed to be those as described before [25] (Supporting information, Table S1 and Fig. 1). Costs in terms of NADPH, NADH, flavin adenine dinucleotide hydrate (FADH), ATP, and methylenetetrahydrofolate (Mthf) were determined for each pathway. It was assumed that the energy content of NADPH is equal to that of NADH and twice that of FADH and ATP. In the case of alternative routes for biosynthesis of an amino acid, the cost of the pathway with the lowest number of ATP consumed was used as reference. Furthermore, all intermediates of biosynthetic pathways except precursors and energetic co-factors were assumed to be balanced. To calculate overall ATP costs for each

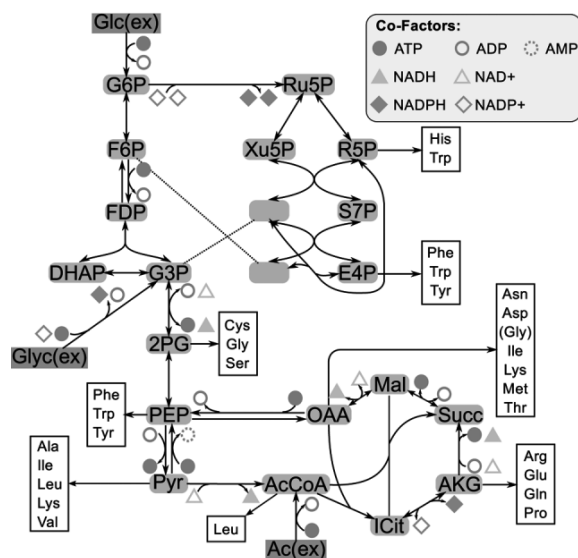


Figure 1. Simplified representation of central metabolism considered for the production of precursors for amino acid biosynthesis. Uptake routes for carbon sources are also depicted. Note that the reaction converting Succ to Mal reduces 1 mol of FAD^+ , which is equivalent to converting 1 mol of ADP into ATP in *E. coli*. Metabolites connected with dashed lines are identical. Metabolites considered as carbon sources are drawn in boxes. For a list of abbreviations, see Supporting information, Table S5.

amino acid, the ATP produced during formation of each precursor from the respective amount of carbon source was added to the overall ATP cost of the amino acid. For these calculations glycolysis, tricarboxylic acid (TCA) cycle, and pentose phosphate pathway are assumed to synthesize precursors (Fig. 1). The corresponding data are provided in Supporting information, Table S2.

2.2.1 Computation of yields

Yields were computed by determining, for each amino acid, the amount of carbon source that is required to produce the precursors of these amino acids (Supporting information, Table S2). In the case that the biosynthetic pathway consumes ATP, it was assumed that ATP is balanced by oxidizing the respective quantities of carbon source. If ATP is produced, it was assumed that the energy is dissipated or utilized through other means (e.g., protein synthesis) and thus does not affect yield. In *E. coli*, complete oxidation of 1 mol of glucose leads to a theoretical maximum yield of 26 mol of ATP, 1 mol of glycerol yields a maximum of 15 mol of ATP and 1 mol of acetate yields a maximum of 7 mol of ATP. Additionally, it was assumed that the carbon source is used to supply Mlthf in cases where the biosynthesis of the amino acid requires this compound. Using linear programming, we determined that 1 mol of glucose can be used to convert 4.55 mol of tetrahydrofolate into Mlthf. For glycerol and acetate this number is 2.48 and 1.16, respectively. Total glucose consumption for synthesis of each amino acid was determined by summing up the amount of carbon source required to produce the precursors, the amount of carbon source oxidized to produce ATP, and the amount of carbon source required to produce tetrahydrofolate. Yields were then computed by dividing one by the amount of carbon sources consumed. Hence, the resulting yield is given in mol/mol.

2.3 Computational methods

Linear programming was used to compute optimal flux distributions. Given the stoichiometric matrix of the metabolic network, in which rows correspond to metabolites and columns to reactions, we assumed that all intermediates are balanced (steady-state condition) and that reversible reactions are only used in thermodynamically feasible directions. For computational reasons we split reversible reactions into irreversible forward and backward steps. For details on how we computed yield-optimal flux distributions with and without assuming an unlimited energy supply see Supporting information, Text S3.

3 Results

For our computations, we used a genome-scale model of *E. coli* metabolism [20]. We assumed aerobic conditions

and unlimited availability of ammonia, sulfate, water (including H^+ , OH^-), and carbon dioxide. For the interconversion of energy equivalents in the form of ATP, NADH, NADPH and FADH we used ATP as reference assuming a Phosphate/Oxygen (P/O) ratio of 2.0. Thus, 2 mol of ATP are energetically equivalent to 1 mol of NADH, 1 mol of NADPH and 2 mol of FADH. Since this value depends on culture conditions, the data provided for the manual yield computations in Supporting information, Table S2 allow to determine yields also for other P/O ratios.

3.1 Methods to characterize amino acid biosynthetic flux distributions

There exist various approaches to characterize flux distributions. Two important aspects are the yield, a measure that describes the product-to-substrate ratio, and energetic cost, such as the amount of ATP consumed or produced by a flux distribution.

Yield can be expressed by different quantities. A measure that is not biased by the mass of the final product is carbon yield. The carbon yield of a flux distribution relates the amount of carbon atoms present in the source compounds of a flux distribution to the carbon atoms in the final product. Thus, a carbon yield equal to 1 C-mol/C-mol indicates that the same numbers of carbon atoms are present in the source compounds and in the product. Hence, a carbon yield below 1 C-mol/C-mol indicates that carbon atoms have been lost to a side-product. A similar measure is cost per kilodalton, that is, the amount of source compounds that is required to synthesize 1 kDa of pathway product. In contrast to carbon yield, this measure does also take into account elements other than carbon.

Apart from measures to characterize flux distributions, methods to determine them are another important issue. In the classical work of Stouthamer [17] it was assumed that glycolysis, pentose phosphate pathway, and citric acid cycle are used to produce a set of predefined precursors for amino acid biosynthesis. From these precursors biosynthetic pathways leading to amino acids are relatively linear and there seldom exist alternative routes. These pathways were then characterized in terms of ATP, NADPH and NADH consumed and produced. In an extension of this work, Akashi and Gojobori [18] and Heizer et al. [5] additionally took into account the energy produced during synthesis of the precursors and determined the overall amount of carbon source required to synthesize each amino acid.

Besides this manual method to determine flux distributions, another commonly used approach is linear programming [22, 26, 27]. This method allows to determine a flux distribution that is optimal with respect to an objective function [4, 19, 28–30]. Such calculations allow the determination of the maximal amount of product that can be synthesized from a set of source compounds. This is

Table 1. Characteristics of flux distributions producing the 20 amino acids from glucose, glycerol, and acetate

AA ^{a)}	Glucose				Glycerol				Acetate			
	ATP cons. ^{b)}	Yields ^{c)} (mol/mol)			ATP cons.	Yields (mol/mol)			ATP cons.	Yields (mol/mol)		
		Manual	LP Std	LP UE		Manual	LP Std	LP UE		Manual	LP Std	LP UE
Ala	−1	2.00	2.00	2.00	−3	1.00	1.00	1.00	−1	0.50	0.50	0.50
Arg	0	1.00	0.89	1.33	−4	0.50	0.51	0.67	5	0.27	0.25	0.33
Asn	2	1.73	1.74	2.00	0	1.00	1.00	1.00	1	0.47	0.47	0.50
Asp	0	2.00	1.86	2.00	−2	1.00	1.00	1.00	−1	0.50	0.50	0.50
Cys	8	1.24	1.03	2.00	6	0.71	0.60	1.00	8	0.32	0.27	0.50
Glu	−7	1.00	1.15	1.33	−11	0.50	0.60	0.67	−2	0.33	0.33	0.33
Gln	−6	1.00	1.19	1.33	−10	0.50	0.61	0.67	−1	0.33	0.33	0.33
Gly	−2	2.00	2.73	4.00	−4	1.00	1.54	2.00	−2	0.50	0.71	1.00
His	3	0.75	0.89	1.20	4.33	0.42	0.49	0.60	7.67	0.19	0.22	0.30
Ile	7	0.79	0.75	1.00	3	0.45	0.44	0.50	6	0.21	0.20	0.25
Leu	−9	0.67	0.75	0.80	−15	0.33	0.38	0.40	−5	0.20	0.20	0.20
Lys	5	0.84	0.80	1.00	1	0.48	0.47	0.50	4	0.22	0.21	0.25
Met	18	0.71	0.62	2.00	16	0.40	0.36	1.00	17	0.19	0.16	0.50
Phe	0	0.57	0.56	0.60	−4.33	0.30	0.30	0.30	2.33	0.14	0.14	0.15
Pro	−2	1.00	1.01	1.33	−6	0.50	0.57	0.67	3	0.29	0.29	0.33
Ser	−2	2.00	2.00	2.00	−4	1.00	1.00	1.00	−2	0.50	0.50	0.50
Thr	6	1.37	1.30	2.00	4	0.78	0.75	1.00	5	0.37	0.35	0.50
Trp	−1	0.44	0.47	0.50	−2	0.25	0.25	0.25	6	0.11	0.12	0.13
Tyr	−2	0.57	0.58	0.60	−6.33	0.30	0.30	0.30	0.33	0.15	0.15	0.15
Val	−2	1.00	1.00	1.00	−6	0.50	0.50	0.50	−2	0.25	0.25	0.25

a) Amino acids.

b) ATP cons. indicates the number of ATP consumed according to the manual calculations for each carbon source as indicated on the top of the rows (assuming 1 NADH = 1 NADPH = 2 FADH = 2 ATP).

c) Yields corresponds to yields obtained from different computation schemes for flux distributions: Manual corresponds to manual computation, LP Std corresponds to LP using the standard formulation (LP standard) and LP UE corresponds to linear programming assuming unlimited energy supply (LP unlimited energy). For information about the ATP consumption for LP standard and LP unlimited energy see Supporting information, Tables S3–S5.

particularly useful in bioengineering to predict the best yield that can be expected upon manipulating biosynthetic pathways. In this LP standard approach, it is commonly assumed that flux distributions balance all intermediates including energy-currency co-factors. Additionally, these calculations can be performed with a network that contains a reaction that converts energy-currency co-factors back into their active form, thus simulating unlimited energy supply (see Sect. 2 and Supporting information, Text S3). In vivo, such an artificial electron supply can be introduced, for instance, by a light-driven proton pump [16]. Moreover, these calculations offer the possibility to determine the energy-independent performance limits of the metabolic network and allow to analyze pathways independent of energy-supplying reactions.

3.2 Costs of amino acid biosynthesis

3.2.1 Manual computation of ATP costs

A simplified representation of the metabolic network considered for production of central metabolic intermediates of amino acid biosynthesis is depicted in Fig. 1. We computed ATP consumption of amino acid biosynthesis for

the synthesis from central metabolic precursors (Supporting information, Table S1) and the overall costs from different carbon sources (Table 1). Starting from the precursors, glycine and serine are the cheapest amino acids since their biosynthesis is ATP neutral. Methionine, whose biosynthesis consumes 20 mol of ATP, is the most expensive amino acid. A table similar to Supporting information, Table S1, which is often used as reference for amino acid biosynthesis cost, has been published [25], but contains several errors. These are due to biosynthetic pathways not known at the time of publication and an erroneous balancing of carbon atoms in several cases. Furthermore, we consider the consumption and production of Mlthf, which serves as a principal donor of C1-bodies for several amino acid biosynthetic pathways. This is of particular importance for the overproduction of amino acids whose synthesis requires this compound [31]. Additionally taking into account the amount of ATP produced during synthesis of precursors gives rise to costs as displayed in Table 1. In this table 10 out of the 20 amino acids are listed with a negative ATP consumption when using glucose as carbon substrate and thus, a net ATP production along their corresponding pathways. This is most pronounced for leucine with a net production of 9 mol of

ATP. The most expensive amino acid is methionine, which consumes 18 mol of ATP, which arises primarily from incorporating sulfur. Using glycerol as carbon source, the biosynthesis of 13 amino acids entails a net production of ATP. Compared to glucose as carbon source, the effect of ATP production through amino acid synthesis is even more pronounced with leucine being the cheapest amino acid with a net production of 15 mol of ATP. When synthesizing amino acids from acetate, only 8 amino acids entail a net production of ATP. Thus, the metabolic cost associated with the production of amino acids depends strongly on the carbon source. This issue will be discussed in more detail in Sect. 3.3.

We furthermore computed the yields of the flux distributions determined with the manual approach (Table 1). For growth on glucose, tryptophan has the lowest yield of 0.444 mol/mol, while alanine, aspartate, glycine and serine allow for the highest yield of 2.0 mol/mol.

3.2.2 Calculating costs by linear programming

We computed the maximum yields of individual amino acid biosynthetic flux distributions starting from different carbon sources using linear programming under the assumption that (i) pathways have to balance energy containing co-factors (LP standard) and (ii) that reduction potential in form of NADH is available in excess (LP unlimited energy). In consequence, also NADPH and ATP are available in unlimited amounts.

3.2.2.1 LP standard formulation

The yields for the LP standard case are displayed in Table 1. They contained only slight deviations or were

identical in most cases to the yields computed with the manual method in Sect. 3.2.1. However, in particular the yields of amino acids that implied a net production of ATP are higher than in the manual computations. In this case, we assumed that the organism uses standard routes through central metabolism to produce the precursors and uses the route with the smallest overall ATP consumption to produce the amino acid from the precursors. Thus, we used two basic assumptions for the identification of the corresponding flux distribution: (i) Standard routes of central metabolism are used and (ii) the yield of the amino acids is maximized.

These two assumptions are conflicting to some extent since the standard routes in central metabolism often produce energy in form of ATP, which in turn necessarily reduces the amount of synthesizable product. This is particularly apparent from the production of acetyl coenzyme A (acetyl-CoA) that is required in leucine biosynthesis. On the standard route through glycolysis, acetyl-CoA can be produced from 0.5 mol of glucose with an accompanied release of 1 mol of carbon dioxide. Using an alternative route via the pentose phosphate pathway, nucleotide synthesis and deoxynucleotide degradation (Fig. 2A) 5/6 mol of glucose are converted into 1 mol of glyceraldehyde 3-phosphate (G3P) and acetyl-CoA. Here, the carbon dioxide releasing step is avoided and only 1/3 mol of glucose are required to produce 1 mol of acetyl-CoA, but require an overall 2 mol of ATP. When converting glucose into acetyl-CoA along glycolysis 5 mol of ATP are produced in total instead. Thus, in the case of leucine biosynthesis the yield from the manual computations is 0.667 mol/mol with 9 mol of ATP produced while the opti-

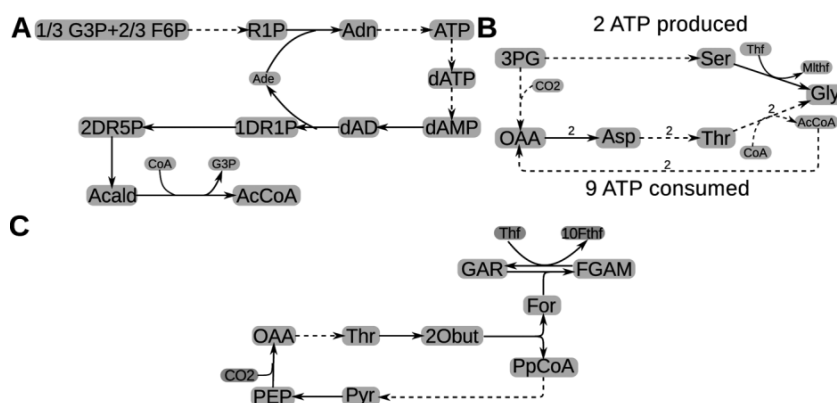


Figure 2. (A) Pathway for production of acetyl-CoA that bypasses carbon dioxide production: 1 mol of glycerol-3P (G3P) and 2/3 mol of fructose-6P (F6P) are converted into 1 mol of AcCoA and G3P each. Dashed arrows correspond to lumped reactions. (B) Simplified representation of glycine biosynthetic and related pathways in *E. coli* from 3-phosphoglycerate (3PG) and oxaloacetate (OAA). Small numbers above the arrows indicate relative fluxes larger than one. The upper flux distribution produces 2 mol of ATP with a yield of 1 mol of glycine from 1 mol of 3PG. The lower pathway consumes 9 mol of ATP with a glycine-over-3PG yield of two. (C) Simplified representation of a pathway that allows fixation of carbon dioxide by tetrahydrofolate (Thf). Please note that 10-formyltetrahydrofolate (10Fthf) can be converted into 5-Mthf that is required for methionine biosynthesis. Additional abbreviations: 1DR1P, 2-deoxy-D-ribose-1-phosphate; 2DR5P, 2-deoxy-D-ribose-5-phosphate; 2Obut, 2-oxobutanoate; Acald, acetaldehyde; Adn, adenosine; dAD, deoxyadenosine; dAT(M)P, deoxy-AT(M)P; FGAM, N2-formyl-N1-(5-phospho-D-ribosyl)glycinamide; For, formate; GAR, N1-(5-phospho-D-ribosyl)-glycinamide; PpCoA, propanoyl-CoA; R1P, ribose-1-phosphate.

mal yield is 0.752 mol/mol and uses the route depicted in Fig. 2A to synthesize acetyl-CoA.

Furthermore, in some cases there exist several routes to produce amino acids, where a pathway with a less favorable ATP balance can result in a higher yield. One example is glycine biosynthesis (Fig. 2B). There exist three pathways for glycine biosynthesis in *E. coli*, two from oxaloacetate via threonine biosynthesis and one from 3-phosphoglycerate via serine biosynthesis. Starting from 1 mol of 3-phosphoglycerate, 1 mol of glycine is produced along with a zero net consumption of ATP (Supporting information, Table S1). Additionally, this pathway produces 1 mol of Mthf, which can be converted back into tetrahydrofolate while reducing 1 mol of NADP⁺ and, thus, providing 2 mol of ATP, which is the overall ATP production along this pathway. The other pathway via threonine has a net consumption of 1 mol of oxaloacetate and 9 mol of ATP. Hence, the pathway via serine has the higher ATP production but the lower yield. In contrast, the pathway via threonine has a higher yield but consumes ATP. In consequence, the optimal pathway combines the ATP production along the pathway via serine with the ATP consumption of the pathway via threonine in order to achieve a higher yield than with the ATP-producing pathway alone. In vivo, glycine biosynthesis via serine is the major source of glycine since a strain with a knockout in the enzyme converting serine to glycine, serine hydroxymethyltransferase (GlyA), is glycine auxotrophic [2].

3.2.2.2 Linear programming assuming unlimited energy supply (LP unlimited energy)

The yields of yield-optimal flux distributions assuming unlimited energy supply are shown in Table 1. In a few cases, these yields are similar to LP standard-computed flux distributions. Glycine biosynthesis is particularly

interesting, since the yield increases from 2.7 to 4.0 mol/mol, as the pathway via threonine can be used exclusively if energy is available in excess. However, 4.125 mol of ATP per mol glycine produced are required to achieve this yield (Supporting information, Table S3).

The most drastic increase can be observed for methionine biosynthesis, where the yield increases more than threefold from 0.621 to 2.0 mol/mol. Thereby, the cost increases to 36.25 mol of ATP per mol of methionine produced (Supporting information, Table S3). The reasons for this drastic increase are due to two carbon atoms of methionine that can be provided from carbon dioxide. On the biosynthetic pathway of methionine, 1 mol of 3-phosphoglycerate is converted into 1 mol of methionine. Additionally, 1 mol of L-homoserine is converted into pyruvate and 1 mol of 5-Mthf is converted into tetrahydrofolate, thereby providing two additional carbon atoms. During conversion of pyruvate back into L-homoserine, one carbon dioxide is fixed. Furthermore, a pathway that allows fixation of carbon dioxide by tetrahydrofolate and uses reactions of threonine degradation and purine biosynthesis can be used to convert tetrahydrofolate into 5-Mthf (Fig. 2C). In consequence, 1 mol of glucose can be used to synthesize 2 mol of methionine. However, as the high energetic cost suggests, this pathway might not be feasible under physiological conditions. Nevertheless, these numbers are important as they show the maximal theoretical yield from the carbon source even though additional energy (electrons) would be required from another source.

3.2.3 Carbon yields of amino acid biosynthesis

Next, we determined the carbon yields of the flux distributions for the different computation schemes and carbon sources (Fig. 3). In case of the flux distributions com-

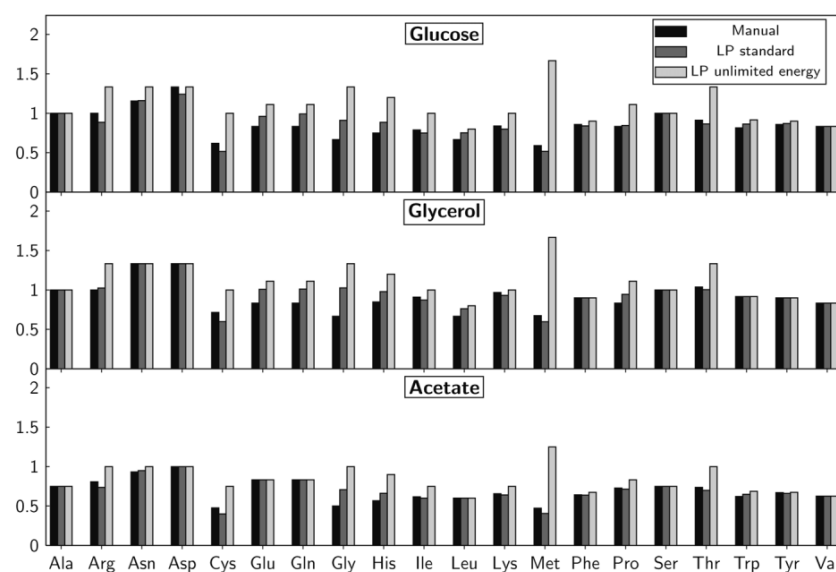


Figure 3. Fraction of carbon atoms retained after biosynthesis of different amino acids for different cost computation schemes and carbon sources (given in C-mol/C-mol).

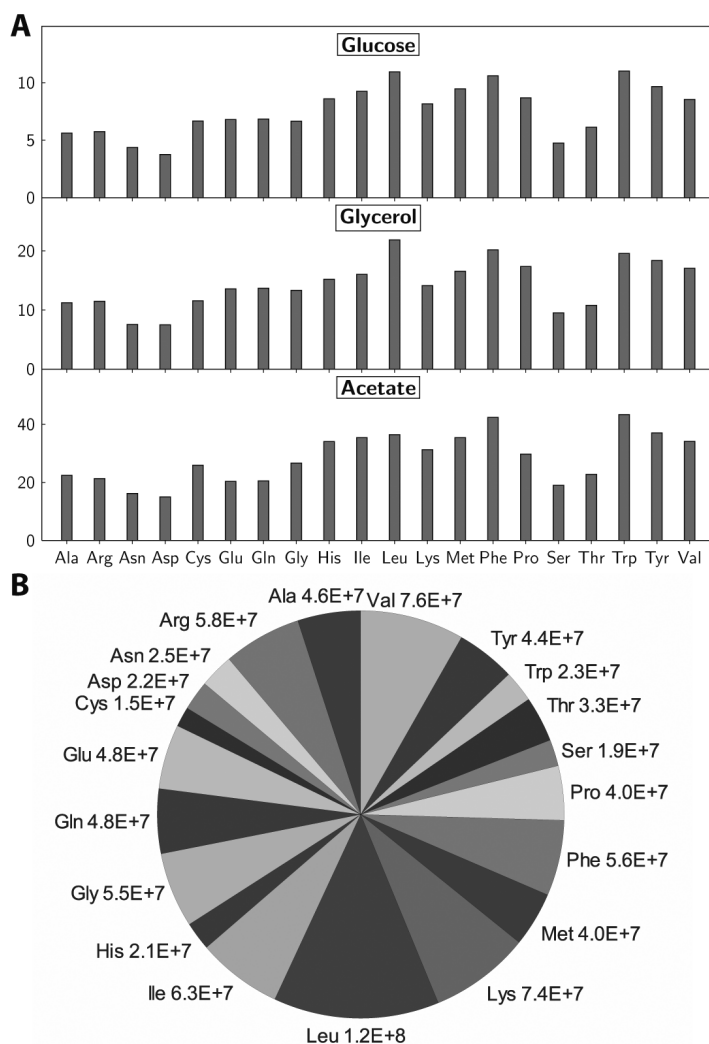


Figure 4. (A) Number of carbon source molecules required for the production of 1 kDa of each amino acid for various carbon sources. For the computation, yields from the manual computations were used as reference. (B) Number of glucose molecules consumed for the production of the amount of each amino acid present in one *E. coli* cell (comprising protein bound and free amino acids). Yields from the manual computations were used.

puted with the manual method using glucose as carbon source, carbon yields are mostly below 1 C-mol/C-mol with the exception of alanine, serine, asparagine and aspartate. For the latter two amino acids, carbon yields are even above 1 C-mol/C-mol, showing that the synthesis of these amino acids involves fixation of carbon dioxide. The rather low carbon yield of cysteine and methionine of around 0.5 C-mol/C-mol is due to the high energy investment for their biosyntheses (Table 1). Using the LP standard formulation to compute maximal yields only slightly changes the carbon yields in comparison to the manual computations (Fig. 3). Assuming an unlimited energy supply, we found that the carbon yields are above 1 C-mol/C-mol in ten cases, equal to 1 C-mol/C-mol in five cases and below 1 C-mol/C-mol in another five cases (Fig. 3). Thus, the biosynthesis of half of the amino acids can, theoretically, be used to fix carbon dioxide. For glyc-

erol as carbon source the yields tend to be slightly higher than in the case of glucose utilization. For acetate, the carbon yields show a strong decrease with only one being larger than one if assuming unlimited energy supply. This can be explained by gluconeogenesis requiring the action of either phosphoenolpyruvate carboxykinase or malic enzyme to produce phosphoenolpyruvate or pyruvate, respectively, from acetate. In both cases, one carbon atom is lost to carbon dioxide.

3.2.4 Costs per kilodalton and overall amino acid synthesis

One bias present in the number of ATPs required for the synthesis of individual amino acids as well as the yields is that the masses of the amino acids are neglected. Thus, the yields computed via the manual method assuming glucose as carbon source indicate that the production of

tryptophan (yield of 0.44 mol/mol) is almost five times as expensive as the production of alanine (yield of 2.0 mol/mol). However, alanine only contains 3 carbon atoms whereas tryptophan contains 11. To reduce this bias we computed for each amino acid the number of carbon source molecules that are required to synthesize an equivalent of 1 kDa of the amino acid including the costs of oxidizing carbon sources to provide energy for amino acid synthesis (Fig. 4).

In the context of these numbers, tryptophan is only twice as expensive as alanine. Moreover, these costs show that phenylalanine and methionine are almost as expensive as tryptophan if synthesized from glucose and more expensive if synthesized from glycerol.

Furthermore, we computed the overall number of glucose molecules invested into the production of the amount of every amino acid present in one *E. coli* cell, based on the yields from the manual calculations and the biomass composition of *E. coli* reported in [20] (Fig. 4B). These numbers show that while tryptophan is the most expensive amino acid in terms of per molecule synthesis cost, it only represents 1.1% of all proteic amino acids on a molar basis and its production requires only 2.5% (2.3×10^7 glucose molecules) of the glucose consumed for the production of all amino acids. In contrast, the biosynthesis of leucine that comprises 8.4% of all proteic amino acids demands 1.2×10^8 glucose molecules, which corresponds to 13% of the glucose consumed for total amino acid production.

3.3 Computation of protein synthesis costs

In a next step, we combined the computation of the biosynthesis costs of individual amino acids with the computation of the biosynthesis cost of entire proteins

(Sect. 2). Besides amino acids, additional energy is required to translate the mRNA from DNA and to polymerize amino acids. We assumed that each amino acid requires a nucleotide triplet that can be synthesized with a cost of 6 ATP [17]. Furthermore, we assumed that on average 30 proteins are translated per mRNA [25]. Thus, the transcription costs add up to 6 ATP/30 copies per mRNA = 0.2 ATP per amino acid. The costs of translation are 4 ATP per amino acid [32]. In consequence, the cost of polymerization is approximately 4.2 ATP per amino acid residue. Since LP-based costs tend to involve non-canonical routes for the synthesis of individual amino acids, we used the costs calculated with the manual method.

3.3.1 ATP investment for protein production

In order to analyze the impact of biosynthesis and polymerization cost on metabolism, we first computed the average ATP investment per residue required for the biosynthesis of every protein of *E. coli*, including polymerization costs. A histogram of these values over all proteins of *E. coli* assuming glucose as carbon source shows that the synthesis and polymerization of the amino acids to generate most proteins consume ATP with a median of 3.7 mol of ATP cost per residue (Fig. 5). Since the polymerization costs of 4.2 mol of ATP per residue are included, there is on average a net production of $3.7 - 4.2 = -0.5$ mol of ATP per residue if accounting only for the costs of amino acid syntheses.

This result is even more pronounced using glycerol as carbon source (Fig. 5A). Here, 0.44 mol of ATP are consumed on average per residue. Thus, based on glycerol, for most proteins the polymerization cost of amino acids can be replenished from the energy produced during amino acid biosynthesis alone. This particular characteristic of protein production from glycerol is, as discussed in

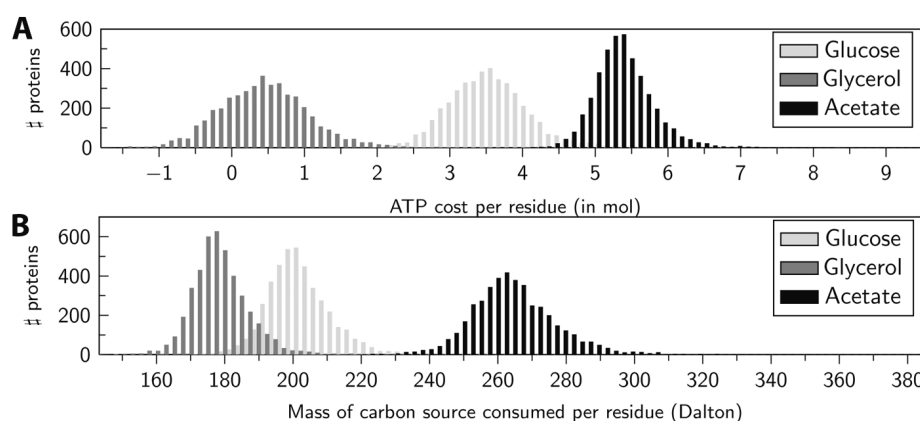


Figure 5. Cost of protein production. (A) Histogram of the average ATP balance (per amino acid) in the synthesis of amino acids for all proteins in *E. coli* including polymerization costs. For the corresponding numbers for each protein of *E. coli* see Supporting information, Data S1. (B) Histogram of the cost of the biosynthesis of all *E. coli* proteins (per residue and mol of protein) from various carbon sources. Costs are given as the average mass of carbon source (in Da) that is required per residue of the protein. Yields of individual amino acids from the manual computations were used. These numbers include polymerization costs. For the corresponding numbers of each protein of *E. coli* see Supporting information, Data S1.

Section 3.2, due to the fact that the synthesis of 65% of amino acids from glycerol provides a surplus of energy.

A completely different picture can be obtained for acetate (Fig. 5A), where the synthesis of amino acids contained in proteins has a median of 5.19 mol of ATP consumed per residue. Hence, a larger fraction of carbon source has to be additionally oxidized in order to synthesize amino acids and to provide ATP for polymerization in comparison to glucose and glycerol.

3.3.2 Carbon source required to produce proteins

We determined the amount of carbon source that is required to produce each protein. In order to reduce bias due to the different sizes of proteins, we divided the amount of carbon source required for the production of each protein by the respective number of amino acid residues contained in the protein. Moreover, we accounted for the different molecular masses of the carbon sources by determining, for each carbon source, the mass of the carbon source molecules required to synthesize each amino acid residue.

Analyzing the histogram of the per residue costs of all proteins, we found monomodal distributions that resemble each other for the three carbon sources (Fig. 5B). For glucose, we found a median of 204 Da of glucose consumed per amino acid residue per protein. For glycerol and acetate, the respective medians are at 178 and 259 Da of carbon source, respectively. Thus, also considering the mass of the invested carbon source, glycerol is most effective for protein production.

4 Discussion

Here, we have characterized amino acid biosynthetic pathways in *E. coli* starting from different carbon sources using different types of costs. The knowledge of ATP yields is of great relevance in biochemistry in general. Many pathways (not only amino acid-producing pathways) are characterized and compared according to their ATP yield. This has practical implications in biotechnology and synthetic biology because products of interest can only be produced when the net energy balance allows this. Moreover, energy-efficient pathways are obviously of special interest. Furthermore, these numbers can serve as an important guideline for optimizing growth media for recombinant protein production. Depending on the physiological condition of the organism during the production process, such as intracellular ATP levels or levels of amino acid precursors, as well as experimental conditions, such as concentration of the carbon source, different cost measures can be most informative.

Classically, the ATP consumption along amino acid biosynthetic pathways as well as the product yield have been considered the most important factors [17, 18]. However, as we have shown, flux distributions with a high pro-

duction of ATP often have a reduced yield with respect to the carbon substrate. We demonstrate that yield-optimal flux distributions of individual amino acids can contain non-canonical routes, which can potentially increase product yields compared to those routes that are normally used by the cell. Moreover, our analysis of protein producing flux distributions showed that the polymerization costs of a large number of proteins could be partially or, in the case of glycerol, almost fully compensated by energy produced during biosynthesis of amino acids. These results are confirmed to some extent by the findings that glycerol is often better suited as carbon source than glucose, in particular after recombinant protein production has been initiated [33, 34].

In order to reduce the bias due to the different molecular mass of individual amino acids we have computed carbon yields and costs per kilodalton. These computations showed that with the carbon sources considered here, carbon atoms are lost to carbon dioxide in most amino acid synthesizing flux distributions. However, in theory there exist flux distributions that allow for fixation of carbon atoms from carbon dioxide for half of the amino acids. The calculations showed that tryptophan is, despite being the largest amino acid, not the most expensive amino acid if accounting for the mass of amino acids. Instead, phenylalanine as well as methionine are equally expensive, if synthesized from glucose. Assuming canonical routes for the production of amino acid precursors, methionine is the most expensive amino acid. Its biosynthesis from glucose costs 18 mol of ATP, which is of particular importance for the design of production processes and strains over-producing this amino acid, as it is one of the most important commercial amino acids [7, 31, 35]. In contrast, considering the quantities in which amino acids are present in *E. coli*, we found that the cell invests the largest amount of carbon source molecules into the production of leucine. Our analysis gives a detailed overview on the metabolic costs of amino acid and protein synthesis from different points of view. These results can serve as an important guideline for the improvement of production media. In particular, the utilization of glycerol as carbon source and the supplementation of expensive amino acids as indicated by our analysis may help to considerably improve production processes.

Financial support from the German Ministry of Education and Research (BMBF) within the Jena Centre of Bioinformatics, the FORSYS-Partner program (grant FKZ 0315285D and E) and the Jena Centre for Systems Biology of Ageing, JenAge (grant FKZ0315581D) is gratefully acknowledged.

The authors declare no conflict of interest.

5 References

- [1] Makrides, S. C., Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol. Rev.* 1996, 60, 512–538.
- [2] Baneyx, F., Mujacic, M., Recombinant protein folding and misfolding in *Escherichia coli*. *Nat. Biotechnol.* 2004, 22, 1399–1408.
- [3] Carlson, R. P., Metabolic systems cost-benefit analysis for interpreting network structure and regulation. *Bioinformatics* 2007, 23, 1258–1264.
- [4] Kayser, A., Weber, J., Hecht, V., Rinas, U., Metabolic flux analysis of *Escherichia coli* in glucose-limited continuous culture. I. Growth-rate-dependent metabolic efficiency at steady state. *Microbiology* 2005, 151, 693–706.
- [5] Heizer, E. M., Raiford, D. W., Raymer, M. L., Doom, T. E. et al., Amino acid cost and codon-usage biases in 6 prokaryotic genomes: A whole-genome analysis. *Mol. Biol. Evol.* 2006, 23, 1670–1680.
- [6] Park, J. H., Lee, S. Y., Fermentative production of branched chain amino acids: A focus on metabolic engineering. *Appl. Microbiol. Biotechnol.* 2010, 85, 491–506.
- [7] Park, J. H., Lee, S. Y., Metabolic pathways and fermentative production of L-aspartate family amino acids. *Biotechnol. J.* 2010, 5, 560–577.
- [8] Wendisch, V. F., Bott, M., Eikmanns, B. J., Metabolic engineering of *Escherichia coli* and *Corynebacterium glutamicum* for biotechnological production of organic acids and amino acids. *Curr. Opin. Microbiol.* 2006, 9, 268–274.
- [9] Korz, D. J., Rinas, U., Hellmuth, K., Sanders, E. A., Deckwer, W. D., Simple fed-batch technique for high cell density cultivation of *Escherichia coli*. *J. Biotechnol.* 1995, 39, 59–65.
- [10] Lee, S. Y., High cell-density culture of *Escherichia coli*. *Trends Biotechnol.* 1996, 14, 98–105.
- [11] Shiloach, J., Fass, R., Growing *E. coli* to high cell density – a historical perspective on method development. *Biotechnol. Adv.* 2005, 23, 345–357.
- [12] Rinas, U., Hellmuth, K., Kang, R., Seeger, A., Schlieker, H., Entry of *Escherichia coli* into stationary phase is indicated by endogenous and exogenous accumulation of nucleobases. *Appl. Environ. Microbiol.* 1995, 61, 4147–4151.
- [13] Luli, G. W., Strohl, W. R., Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations. *Appl. Environ. Microbiol.* 1990, 56, 1004–1011.
- [14] Shiloach, J., Rinas, U., Glucose and acetate metabolism in *E. coli* – system level analysis and biotechnological applications in protein production processes, in: Lee, S. Y. (Ed.), *Systems Biology and Biotechnology of E. coli*, Springer-Verlag, Berlin, Heidelberg, New York 2009, pp. 377–400.
- [15] Wolfe, A. J., The acetate switch. *Microbiol. Mol. Biol. Rev.* 2005, 69, 12–50.
- [16] Stryer, L., Berg, J. M., Tymoczko, J. L., *Biochemistry*, 7th Edn., W H Freeman & Co 2010.
- [17] Stouthamer, A. H., A theoretical study on the amount of ATP required for synthesis of microbial cell material. *Antonie Van Leeuwenhoek* 1973, 39, 545–565.
- [18] Akashi, H., Gojobori, T., Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* 2002, 99, 3695–3700.
- [19] Varma, A., Palsson, B. O., Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl. Environ. Microbiol.* 1994, 60, 3724–3731.
- [20] Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M. et al., A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* 2007, 3, 121.
- [21] Varma, A., Boesch, B. W., Palsson, B. O., Biochemical production capabilities of *Escherichia coli*. *Biotechnol. Bioeng.* 1993, 42, 59–73.
- [22] Burgard, A. P., Maranas, C. D., Probing the performance limits of the *Escherichia coli* metabolic network subject to gene additions or deletions. *Biotechnol. Bioeng.* 2001, 74, 364–375.
- [23] Schellenberger, J., Park, J. O., Conrad, T. M., Palsson, B. O., BiGG: A biochemical genetic and genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics* 2010, 11, 213.
- [24] Keseler, I. M., Collado-Vides, J., Santos-Zavaleta, A., Peralta-Gil, M. et al., EcoCyc: A comprehensive database of *Escherichia coli* biology. *Nucleic Acids Res.* 2011, 39, D583–D590.
- [25] Ingraham, J. L., Maaloe, O., Neidhardt, F. C., *Growth of the bacterial cell*, Sinauer Associates, Sunderland, Mass. 1983.
- [26] Varma, A., Palsson, B. O., Metabolic capabilities of *Escherichia coli*: I. Synthesis of biosynthetic precursors and cofactors. *J. Theor. Biol.* 1993, 165, 477–502.
- [27] Sajitz-Hermstein, M., Nikoloski, Z., A novel approach for determining environment-specific protein costs: The case of *Arabidopsis thaliana*. *Bioinformatics* 2010, 26, i582–i588.
- [28] Fell, D. A., Small, J. R., Fat synthesis in adipose tissue. An examination of stoichiometric constraints. *Biochem. J.* 1986, 238, 781–786.
- [29] Shlomi, T., Benyamini, T., Gottlieb, E., Sharan, R., Ruppin, E., Genome-scale metabolic modeling elucidates the role of proliferative adaptation in causing the Warburg effect. *PLoS Comput. Biol.* 2011, 7, e1002018.
- [30] Feist, A. M., Palsson, B. O., The biomass objective function. *Curr. Opin. Microbiol.* 2010, 13, 344–349.
- [31] Krömer, J. O., Wittmann, C., Schroder, H., Heinzle, E., Metabolic pathway analysis for rational design of L-methionine production by *Escherichia coli* and *Corynebacterium glutamicum*. *Metab. Eng.* 2006, 8, 353–369.
- [32] Lehninger, A., Nelson, D. L., Cox, M. M., *Lehninger Principles of Biochemistry*, 5th Edn., W. H. Freeman 2008.
- [33] Luo, Q., Shen, Y. L., Wei, D. Z., Cao, W., Optimization of culture on the overproduction of TRAIL in high-cell-density culture by recombinant *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 2006, 71, 184–191.
- [34] Berger, C., Montag, C., Berndt, S., Huster, D., Optimization of *Escherichia coli* cultivation methods for high yield neuropeptide Y receptor type 2 production. *Protein Expr. Purif.* 2011, 76, 25–35.
- [35] Ikeda, M., Amino acid production processes. *Adv. Biochem. Eng. Biotechnol.* 2003, 79, 1–35.

4. Discussion

Studying the dynamics of biological systems is a compelling albeit challenging endeavor. A primary goal is to understand the inherent dynamics and their driving causes, since it explains in more detail the underlying mechanism of a studied phenomena. This is feasible for small, well defined and definite problems, which allow for simulations of time dependent transition courses of model species. Larger scaled problems, however, commonly prohibit such an approach due to their sheer complexity. In these cases, assuming that a system is operating at steady state still enables a comprehensive analysis, as numerous studies have shown (cf. Papin et al. (2004); Trinh et al. (2008); Lewis et al. (2010, 2012); Unrean and Srienc (2012); Khandelwal et al. (2013); Harcombe et al. (2013)). As has been mentioned before, all included model species or state variables are assumed constant over time at steady state, thus are neither accumulating nor depleting.

This thesis is dedicated to five studies in systems biology comprising dynamic and constraint-based models, the latter being used when the corresponding project proved to be of a scale, where kinetic modeling was not feasible.

4.1. ODE based modeling to study substrate competition and cellular senescence

In chapter 2, two studies are presented that were assessed with an ODE based approach. In the first study we investigated the effect on metabolic network simulations, when common kinetic rate laws, such as the Michaelis-Menten kinetics are extended to include the possibility of substrate competition. This theoretical work provides valuable insights for the interpretations of experiments. In addition, we list modified rate laws for more sophisticated bi-bi reaction mechanisms.

In the second study, we conducted a classical systems biology project, where we integrated experimental data of primary human fibroblasts into a hypothetical model

of cellular senescence and proceeded with an iterative model refinement process that gave rise to new experiments. In short, the combination of our model and experimental data revealed that different cell lines of human fibroblasts are not equally robust against external stresses. Furthermore, our model allowed us to discriminate between different cell types at every simulation time-point and the transition rates between them. By quantitatively fitting growth curves we can evaluate marker specificity to resemble a certain cell type, which can be of help for further marker studies in cell populations. Forthcoming, both studies are discussed in more detail.

4.1.1. Effect of substrate competition potentially influences simulation results

Substrate competition has already been reported to influence various biochemical mechanisms, such as the degradation of polymeric carbohydrates (Kartal et al. 2011), metabolic transport (Marquez and Van Bambeke 2011; Pardridge and Choi 1986) or gene regulation (Lee et al. 1995; Jöres and Wagner 2003). Moreover, the effect of substrate competition can be substantial, as several studies have shown (Kartal et al. 2011; Kim et al. 2011; De Vos et al. 2011; Dingerkus et al. 2012; Baks et al. 2006; Pomerening 2008; Venkatraman et al. 2011). Considering an irreversible Michaelis-Menten mechanism, Chou and Talaly derived an appealing simple rate law modification, which considers the competition of an arbitrary number of substrates for the binding site of the associated enzyme (Chou and Talaly 1977). Note that substrate competition in irreversible reactions is in principle equivalent to a competitive inhibition mechanism of a repressor. Two or more substrates can hinder each other to attach to the enzymatic binding site, thus mimicking the effect of a competitive inhibitor. However, we have not found readily applicable rate laws which include aspects of competition in a reversible kinetic rate law, when the product is taken into account as well. Under the assumption that only one substrate at a time can be catalyzed, we filled this gap by deriving rate laws for reversible Michaelis-Menten kinetics. Additionally, we proposed modified rate laws accounting for substrate competition in bimolecular mechanisms, such as the bi-bi ordered or ping-pong mechanism.

In order to evaluate, whether kinetic rate laws that include the possibility of substrate competition have an actual effect on metabolic network simulation, we constructed and simulated three artificial study cases: A, competition between complete

distinct pathways, B, competition in branched pathways and C, multistep reactions with alternating identical enzymes (see section 2).

To assess effects of substrate competition on both, time dependent and steady state concentrations, we conducted two *in silico* experiments. In the first experiment we fixed the external input substrate concentration and studied the transitional concentrations leading towards a steady state. In a second experiment, we studied the steady state concentrations deviating from simulations where substrate concentration is neglected, considering increasing external input substrate concentrations.

Our results show that only under certain circumstances substrate competition can be neglected. One example is provided by two distinct metabolic pathways, where one enzyme catalyzes one reaction in each pathway. Here, substrate competition can be neglected, if one is only interested in the steady state concentration of the final product. On the other hand, one might be particularly interested in the concentration of an intermediate metabolite. Already in this simply reaction scheme, we have shown that substrate competition can have a substantial effect on involved intermediate metabolites (see also Fig. 1, Schäuble et al. (2013), chapter 2).

The effect of substrate competition is even more pronounced, when considering multistep pathways, where one enzyme catalyzes multiple reactions (case C). A prominent example of multistep reaction systems, is given by polymer chain elongation, among others. Starting with only the initial substrate present, our time course simulations show that the concentration of most metabolites is underestimated. In contrast, the transient concentration of the final product and of some intermediates is overestimated, when substrate competition is ignored (Fig. 3 in Schäuble et al. (2013), chapter 2). The steady state simulations confirm that this difference in intermediate metabolite concentrations increases for increasing concentrations of the initial external metabolite. In contrast, the steady state concentrations of the first and last considered metabolite do not change upon substrate competition, due to the irreversible uptake or release of these metabolites, which prevent a change in the simulated steady state concentrations. The underestimation of most of the metabolites is noteworthy, nevertheless, especially in reversible systems. Interestingly, it has also been noted before that metabolic control analysis is affected as well in a multistep reaction system (Cascante et al. 1990).

Our results for study cases A and B show less obvious, yet still notable results. In case A, where we studied substrate competition in two distinct pathways, the time

course shows medium temporary changes for the product of the simulated pathway. Additionally, the concentrations of the intermediate metabolites are underestimated throughout the simulated time frame and stay underestimated at the steady state, when competition is ignored (Fig. 1, Schäuble et al. (2013), chapter 2). Again, the concentration of the final product does not change due to its irreversible excretion. The more complex case B shows modest relative differences in both time course and steady state simulations. Notably, the concentration of the final products differ and may be under- or overestimated, when neglecting substrate competition, whereas the difference is more pronounced in an irreversible reaction scheme (Fig. 2, Schäuble et al. (2013), chapter 2). When considering reversible reactions, already modest differences occur at a first transition phase for the last metabolites. After this phase the differences virtually vanish, but increase again, once the external concentration of the initial substrate is increased notably beyond the K_m of the enzyme under investigation.

Summarizing, our results show that already in exemplifying artificial minimal pathways substrate competition can have a significant impact on both time course as well as steady state concentrations. In consequence, ignoring substrate competition can distort pathway simulation results to a varying degree. The differences in simulation results will even increase when taking into account that in many pathways substrate competition can occur in a combination of the here showed study cases. For instance, the enzyme aromatic-L-amino-acid decarboxylase (DDC, EC: 4.1.1.28) present in the tryptophan pathway contains a combination of our modeled cases A and B. This enzyme catalyzes both, reactions in different branches of the pathway, as well as reactions on other pathways, such as the tyrosine or histidine pathway. An exact determination of the involved metabolite concentration may thus be difficult to acquire, when substrate competition is ignored.

It might be argued that cells of more complex organisms prevent substrate competition effects by compartmentalization or because of a tissue specific use of certain pathways (McKenna 2012). In the modeling realm, however, compartmentalized pathways are often not considered, since information about compartmentalization is frequently missing. To solve this dilemma more experimental clarification is required, an endeavor which is often not feasible due to technical reasons. Here, including substrate competition or not is hard to decide. We have shown that often occurring pathway constellations infer notable differences in substrate concentrations

when including competition. Thus, our presented results clearly suggest to consider substrate competition, since it potentially increases the quality of the estimation of a pathway's dynamics.

4.1.2. Modeling cellular senescence allows for a quantitative description of growth curves of human fibroblasts and evaluation of cell state indicating biomarkers

Since Leonard Hayflick's observation that isolated human cells show a limited capacity to replicate (Hayflick and Moorhead 1961), cellular senescence relishes a wide interest in the field. To contribute to its clarification, we presented a classic systems biology approach in a collaboration with the group of Prof. Stephan Diekmann from the Fritz-Lippmann Institute in Jena. We evaluated experimental growth and biomarker data of primary human fibroblasts in order to derive a dynamic model, which sheds light on the dynamics of cellular senescence. Our modeling results suggested further experiments in turn, not only to refine model parameters, but also to tackle new questions that arose from the modeling itself.

Key to our modeling approach was the differentiation of three cell types: proliferating (P), cell cycle arrested (C) and senescent (S) cells. Via cellular checkpoints, cells detect DNA damage and depending on the magnitude of the damage either activate repair mechanisms, while arresting the cell cycle or induce an orderly cell destruction via activation of apoptotic pathways. When facing harsh stress, however, some cell types can neither be destroyed, nor are able to re-enter the cell cycle, but lock irreversibly in a senescent cell state (Rodier and Campisi 2011; Toussaint et al. 2002, 2011). We hypothesized that before cells enter an irreversible senescent cell state, they attempt recovery while being cell cycle arrested, since mild stress can potentially be overcome and repaired. Thus, next to a constant proliferation rate r , which determines how fast P cells can reproduce in the absence of damage, we introduced three transition rates f_1 , f_2 and f_3 . These rates dictate the transition between the three considered cell types. While f_1 and f_2 describe the reversible conversion between P and C cells, f_3 accounts for the irreversible C to S transition (see also Fig. 2 Schäuble et al. (2012), chapter 2).

So far our model is able to resemble infinite growth or no growth, as an analytical analysis of our model has shown (see supplementary material, Schäuble et al. (2012),

chapter A). To account also for replicative senescence, or in other words for the transition from growth to proliferation stop, we added an equation F to our model. F serves as a stress aggregating and processing unit and was required to allow the cell population to tolerate mild stress, but to stop growth in a fast switch-like behavior when the stress dose surpasses a critical level. We realized F as a single nonlinear function, which has been similarly used in Ludwig et al. (1978). The chosen form of F , which only influences the transition rate from P to C cells, allowed us to acquire a simple, yet meaningful equation. It still features, nonetheless, the desired behavior of robustness and bistability, when stress is severe. As long as the stress level is still moderate, F allows the cells to repair the damage and re-enter proliferation. In its second stable state, however, the bistable property of F prohibits the instant switch-back, once the stress source disappeared, thus mimicking stress remembrance. F 's nonlinear form allows therefore for a sigmoidal switch, which infers a fast slow down of cell proliferation and finally, senescence, when stress doses are critically high for a sufficient amount of time. Notably, specific network structures, like double negative feedback loops, can potentially resemble the desired bistability effects in the context of a biological pathway. Nevertheless, we chose F as described for the sake of simplicity and its advantageous analytical properties.

We distinguished the stress form itself into three categories: pulsed, increasing over time, or permanent. The pulsed form might origin from a sudden intensive stress, such as a radiation impulse, or sudden oxygen deficiency, whereas increasing stress over time resembles e.g. mutation accumulation or telomere shortening. A permanent high stress dose can correspond to a constant high concentration of reactive oxygen species, among others. To be able to simulate these stress types within our model, we introduced a stress term of the form $\gamma = \alpha + \beta t$, which influences the behavior of the stress response function F . Whereas α depicts a time-independent stress influence, β denotes a time-dependent contribution.

We fitted our model to different cellular data of three different primary human fibroblasts. Firstly, a pulsed stressor was introduced by irradiating cells once for a short time, and once for a longer time-period (simulating mild and severe stress). The results were compared to cells of the same fibroblast type, which have not been irradiated. We successfully fitted one parameter set for all three scenarios (no stress, short and long stress pulse) per fibroblast dataset, thus, having only the amount of stress defining the fate of the cell population. Secondly, we fitted our model to cells

growing into cellular senescence induced primarily by oxidative stress and telomere shortening. Here, we obtained a convincing quantitative fit of the respective growth curves.

By comparing fitted parameters and inferred growth curves, we detected different response sensitivities for WI-38, BJ and MRC-5 fibroblasts, which is in agreement with published cell line properties (Benanti and Galloway 2004; Itahana et al. 2003). WI-38 cells are potentially more susceptible to stress, especially when compared to MRC-5 data, since the former cell type showed a more pronounced reaction to smaller γ values. Additionally, WI-38's stress susceptibility is apparent from a low P to C and a high C to S transition rate, whereas BJ cells show a high flux between P and C cells, but a low transition to S cells. The advanced resistance of BJ cells to stress might be caused by a constantly high maintenance flux, as is resembled by its high flux between P and C cells. This picture is coherent with the faster transition of WI-38 cells into senescence at lower population doublings than BJ or MRC-5 cells.

We were not able to show a consistent picture of marker up-regulation for C cells (p21 in BJ and MRC-5 cells) and S cells (SA- β -Gal in MRC-5 cells), however, which might be caused by the rather unspecific nature of the markers p21 and SA- β -Gal to resemble C and S cell fractions, respectively. We cannot rule out that these markers do not have further side effects, which might affect the transition rates or growth itself. Thus, our work highlights the need for further validation of these biomarkers or a study of more specific biomarkers that reflect cell cycle arrested or senescent cells. For instance, annexin A5 might serve as an alternative senescence marker as it accumulates at the nuclear envelope during replicative as well as artificially induced senescence (Klement et al. 2012).

Next to our modeling approach, specific senescence related issues have been analyzed, such as the influence of telomere shortening (Tan 1999; Rubelj and Vondracek 1999; Sozou and Kirkwood 2001; Golubev et al. 2003; Portugal et al. 2008). The stochastic model published by Sozou and Kirkwood (2001) includes telomere shortening, oxidative stress and proposes a link to mitochondrial damage as well as nuclear somatic mutations. This model is in well agreement with variable cell doubling potential, published already decades ago by Smith and Whitney (1980). The influence of telomere length on cellular senescence was modeled by Golubev et al. (2003), who propose that telomere shortening alone does not account for senescence. In-

stead, further effects such as stress via accumulation of reactive oxygen species have to be taken into account as well, which is in agreement with Sozou and Kirkwood (2001). Furthermore, Portugal et al. (2008) presented a pure theoretical stochastic model where the cell division probability decreases with decreasing telomere length. A well described theoretical framework for cell cycle regulation and division was presented by Csikász-Nagy et al. (2006); Tyson and Novak (2008), comprising a protein interaction network of the regulation of DNA synthesis and mitosis, but not cellular senescence. Lawless et al. (2010) published a simple model to assess marker specificity in cellular senescence. The authors did not consider an intermediate cell cycle arrested cell state C, however, whereas we could show that upon mild stress, *in vitro* cell populations can regain their original proliferation rate, once the damage from a mild stress source has been repaired (see also Fig. 4, Schäuble et al. (2012), chapter 2). By stochastically modeling single cells, Passos et al. (2010) showed that DNA damage induces a feedback loop of reactive oxygen species production, which sustains cellular senescence. Integrating our data into their model also indicates the cellular response to low as well as high stress doses. Nevertheless, it does not allow for a quantitative comparison to the growth curve of the respective population, a strength we see in our model.

In summary, our model enables the determination of specific cell fractions at every simulation time point and a comparison to marker specific up-regulations in the investigated cell population or at a single cell level. Despite its value, our endeavor to describe cellular senescence can be further extended to describe additional cellular states such as quiescent or terminally differentiated cells. Moreover, cells might enter apoptotic pathways upon certain stress events. To include these aspects as well, a further set of markers need to be quantified, which can again be well evaluated by our modular modeling approach.

4.2. Constraint-based methods enable the study of large-scale models

The prior studies demonstrated convincingly that the dynamics of a system can be well studied by dynamic models. Constructing and analyzing these models by the means of ordinary differential equations enables a time dependent and, if existing, a steady state analysis, thus giving valuable insight into the dynamics of a system.

The combination of experimental data with a given theoretical model, nevertheless, usually increases the significance of a model. Besides, deriving dynamic models comes with a price, since the type of the included reaction velocities needs to be identified and requires the determination of model parameters. Even if the underlying kinetic rate law is known, commonly the precise parameter value is difficult to assess in a given biological context. Therefore, using kinetic modeling techniques to investigate the dynamics of a system bears the disadvantage that the parameter space becomes easily overwhelmingly vast. In small encapsulated models, conducting necessary experiments might still be achievable, but already moderately scaled models would require a tremendous, if not infeasible, workload. This accounts especially for mechanistic modeling approaches, where various parameters that define the regulation and saturation of enzymes need to be included and critically influence a model's behavior. Hence, the prior biological knowledge of underlying kinetics is unsatisfactory, whereas the computational power required for simulation is often unavailable. In consequence, constraint-based methods emerged as a promising approach to study a specific biological phenomena or organism (Papin et al. 2004). Forthcoming, I discuss three studies, which are comprised in chapter 3, where I utilized the advantages of EFM analysis as well as FBA analysis in order to tackle problems that were prohibited to approach with an ODE based analysis.

In the first study, we introduced the concept of EFMs and applied it to a photosynthate model to analyze starch and sucrose usage. In the second study, we investigated the role of the circadian regulated heteromer CHLAMY1 in nitrogen metabolism of the green algae *C. reinhardtii*. In the third and last study, we provide a broad overview of metabolic costs of amino acid and protein production in *E. coli* based on an FBA analysis, which is of particular importance for an optimal growth media choice in biotechnology and metabolic bioengineering.

4.2.1. Elementary Flux Modes provide a large application range and propose infeasibility of sucrose and starch production in plants at night

Exploring the fundamental properties of metabolic networks, significantly improves our understanding of the mechanisms intrinsic in living systems (Price et al. 2004; Ruppin et al. 2010). Based on network stoichiometry and concentration constraints,

a number of methods have been developed to determine important routes in metabolic networks, known as constraint-based methods (Price et al. 2004). One important representative of this method class is the concept of EFMs (Schuster 1994; Schuster et al. 1999), which allow for a decomposition of a given reaction network into feasible minimal reaction pathways.

In Schäuble et al. (2011b), presented in chapter 3, we described in detail the mathematical background for EFM computation and analysis. Furthermore, we exemplified the usefulness of the method by analyzing a simple example network of the TCA cycle including the glyoxylate shunt, which gave rise to all four possible EFMs, when considering phosphoenolpyruvate (PEP) as external metabolite source. Although it has been shown that at least *E. coli* and *Mycobacterium tuberculosis* use the complete TCA cycle in parallel to the PEP-glyoxylate cycle (Fischer and Sauer 2003; Beste et al. 2011), our analysis does not include a single EFM, which resembles this behavior. In fact, the definition of EFMs to be minimal does not allow for such an EFM, since the simultaneous use of the complete TCA cycle is a superposition of two pathways and can thus be represented by the overlay of two EFMs (see Fig. 22.2, Schäuble et al. (2011b), chapter 3).

To show the usefulness of EFM analysis to determine important network properties, we additionally presented an accurate reconstruction of the photosynthate metabolism and its EFM based analysis. The model comprises reactions from the Calvin cycle and includes the photosynthate metabolism of the chloroplast stroma. The Calvin cycle is primarily regulated by the thioredoxin system (Schürmann and Jacquot 2000). In contrast to Poolman et al. (2003), who analyzed a similar model with respect to triose phosphate excretion, we focused on the systems starch and sucrose production capability.

The complete system gave rise to 42 EFMs, whereas only 12 remained, when considering the down-regulation of reactions from the oxidative pentose phosphate pathway and transaldolase at daytime (Schürmann and Jacquot 2000). Sucrose can be produced solely from carbon dioxides or from both, carbon dioxides and starch. The EFM yield analysis shows that the production via carbon dioxides only is energetically more efficient. However, plants might also rely on starch, when light sources are not available. Since the 30 remaining EFMs also rely on the activity of Rubisco or sedoheptulose-bisphosphatase, these EFMs are not operable at night, because these reactions are down-regulated at night (Schürmann and Jacquot 2000).

In consequence, neither starch nor sucrose production is possible at night in the context of our model. Even if an ATP source would be available, still starch and sucrose build-up is not possible due to the down-regulation of key reactions during night.

Naturally, EFM analysis is not limited to the analysis of photosynthesis alone, but has a wide range of applicability. In our work, we discuss several such possibilities. Among others, EFM analysis has been used to analyze an alternative path to oxidize glucose, (Liao et al. 1996; Schuster et al. 1999) or the response of yeast to external stresses (Schwartz et al. 2007). Knock out simulations are possible as well and let to the study of a mutated *E. coli* strain that lacks a cell wall (Kenanov et al. 2010) or to enzyme deficiency studies in human erythrocytes (Çakir et al. 2004; Schuster and Kenanov 2005). Alternatively, functional deficiencies can be explored by computing and analyzing minimal cut sets, an EFM related concept (Klamt 2006). EFM analysis is also important in the context of bioengineering, where the study and design of microorganism enables the overproduction of biotechnological relevant metabolites (Carlson et al. 2002; Trinh et al. 2006; Trinh and Sreenc 2009; Unrean et al. 2010).

We emphasize that in general EFM analysis strongly depends on a models quality. Even though the number of available models is steadily increasing, often it is unavoidable to conduct a network reconstruction from scratch. The amount of time required to reconstruct a high quality model can be enormous, however, since databases on metabolism regularly contain contradicting information (Feist et al. 2009; Rupp et al. 2010; Thiele and Palsson 2010). Examples include but are not limited to diverse information about reaction reversibility states or the protonation status of metabolites.

Since the number of EFMs increases exponentially with network size (Klamt and Stelling 2002), the computation and analysis of all EFMs is limited to moderately sized networks even when considering modern highly parallelized computation methods (Terzer and Stelling (2008), cf. Fig. 1.5). For instance, the number of extreme pathways, a subset of EFMs, in a genome-scale metabolic network of the human metabolism (Duarte et al. 2007) has been estimated to be about up to 10^{29} (Yeung et al. 2007). Thus, regardless of the chosen method to assess EFMs, computing the complete set of all EFMs is unfeasible for most genome-scale models. Enumerating a subset of feasible EFMs or calculating all EFMs of a subsystem, which are

still valid with respect to the complete network has been shown to circumvent the model-size limitation of EFM applicability. Without the necessity of computing all EFMs the first approach allows to calculate the shortest EFMs (de Figueiredo et al. 2009) or a large sample set of EFMs using a specific reaction (Kaleta et al. 2009b). Alternatively, elementary flux patterns can be computed, which allow for the identification of possible EFMs within a subsystem which is still valid in the entire network (Kaleta et al. 2009a).

Taken together EFM analysis has proven to be a promising tool to break down metabolic networks into minimal feasible pathways. When considering the increasing number of available genome-scale models, however, the classic approach to compute all EFMs of a given network is impracticable. Even though not all EFMs are computable anymore, a particular interesting subset of EFMs is still assessable. Thus, the advantage of EFM analysis ranging from pure theoretical to biotechnological relevant applications, persists in the present and in the future.

4.2.2. EFM analysis predicts a shut down of less effective nitrogen assimilation pathways in *C. reinhardtii* via the circadian controlled factor CHLAMY1 at night

The green algae *C. reinhardtii* has proven a valuable model organism for crop plants, whereby nitrogen metabolism is essential for growth and survival (Harris 2008). To contribute a new perspective on algae science, we investigated the impact of circadian control on *C. reinhardtii*'s nitrogen metabolism. We conducted an EFM based pathway analysis, and included information about the circadian regulator CHLAMY1. Due to a lack of knowledge about involved kinetic parameters, a kinetics based analysis of the dynamics of the nitrogen uptake and assimilation pathways was not feasible. Assessing the theoretical capacity of this metabolic network proved to be valuable, nevertheless, as will be discussed further on.

As a novel aspect we interconnected sequence data into our analysis pipeline of interpreting the calculated EFM set. The heteromer CHLAMY1 is known to recognize and down-regulate mRNA transcripts containing $UG_{\geq 7}$ repeats at their 5' UTR at night (Iliev et al. 2006; Kiaulehn et al. 2007). To include CHLAMY1 based regulation into our model, we investigated *C. reinhardtii*'s sequence information. This allowed for the identification of enzymes that are solely encrypted by mRNAs con-

taining $UG_{\geq 7}$ repeats. Our approach circumvents the complexity of flux analyses, which might be infeasible for higher organisms and proved to be a simple yet efficient and viable method to determine the degree of influence of regulatory factors. In total, six enzymes of the nitrogen pathway are affected by CHLAMY1 in the context of our model. Although more transcripts with a CHLAMY1 recognizable $UG_{\geq 7}$ repeat exist, we only simulated a down-regulation, when no isoenzyme can compensate the affected catalytic deficiencies.

The reconstructed network gave rise to about three million EFMs in total, which comprise the incorporation of nitrogen into the amino acids glycine, alanine, asparagine, lysine and arginine based on the carbon sources Glucose-6-phosphate (G6P) or acetate. These amino acids had been chosen as nitrogen sinks, since they comprise an efficient and above average nitrogen to carbon ratio. We only considered a non-compartmentalized model, since the complete knowledge about the transport reactions connecting the compartments was not available. Moreover, the complexity of the resulting EFM set would have produced a tremendous number of EFMs, which are difficult to assess and interpret.

The sheer number of EFMs suggests a high redundancy and potentially robustness present in the network. Interestingly, our analysis shows that CHLAMY1 shuts down the majority of valid EFMs, narrowing down the solution space for possible nonzero fluxes (see also Fig. 8, Schäuble et al. (2011a), chapter 3). At first glance that might seem intriguing. The down-regulation of inefficient pathways upon CHLAMY1 action at night, however, results in a decrease in the overall energy consumption of nitrogen uptake processes. Similar mechanisms of forcing certain pathways to be left feasible have also been observed in strain optimization (Wlaschin et al. 2006; Trinh and Sreenc 2009; Hädicke and Klamt 2010). Here, we focused on different physiological situations (day and night phases) instead of strain manipulation. Further examples may be present in hibernating organisms, or by comparing developmental versus old tissues.

The analysis of the maximum carbon yields of the investigated pathways shows that the green algae can still incorporate nitrogen, although with reduced theoretical efficiency, when G6P or acetate is available at night. For instance, the arginine synthesizing pathway is completely knocked out, when considering complete down-regulation by CHLAMY1. Here, CHLAMY1 down-regulates the final synthesizing step into arginine, which is catalyzed by the enzyme argininosuccinate lyase (ASL,

see also Fig. 2, Schäuble et al. (2011a), chapter 3). If the external carbon source G6P is interchanged with glyceraldehyde-3-phosphate (GAP), which effectively removes the glycolysis and pentose phosphate pathway from the model, the maximum carbon yields do not change at all, regardless of CHLAMY1 action. Thus, a constraint based modeling approach aiming solely at optimal fluxes such as an FBA analysis is not suitable to detect possible quantitative differences. In contrast, our EFM analysis revealed that about 96 % of all EFMs are affected by CHLAMY1 at night. This result suggests that only a small portion of nitrogen incorporating fluxes remain feasible upon CHLAMY1 down-regulation (Fig. 8, Schäuble et al. (2011a), chapter 3). Moreover, it strengthens the idea that CHLAMY1 acts as a global regulator of a flux distribution rather than affecting only optimal fluxes. Indeed, the down-regulation of mRNAs recognized by CHLAMY1 during the night infers an increase in the median yield of amino acid production at the cost of losing most optimal fluxes (Fig. 5, Schäuble et al. (2011a), chapter 3). As has been shown in various microbacteria, relaxing the need of operating only optimal pathways provides a trade-off between flexibility and optimal fluxes in fluctuating environments (Schuetz et al. 2012). Thus, sacrificing optimal, but cost intensive nitrogen assimilating fluxes at night potentially allows for an adjusted workload of energy and nitrogen metabolism throughout the daily light dark cycle. This in turn enables an efficient uptake of nitrogen residues throughout the day and particularly at night. Since light is unavailable to photosynthetic active organisms at night, particularly energy consuming reactions and thus fluxes are prohibitive, which has also been experimentally confirmed in *Arabidopsis thaliana* (Piques et al. 2009). For *C. reinhardtii*, the decrease in maximum carbon yield is mainly due to the influence of CHLAMY1 on phosphoglucose isomerase (G6PI). The down-regulation of G6PI forces G6P to be converted via the pentose phosphate pathway, which is required by nucleotide synthesis. Since DNA replication is primarily occurring at night, this might well be affected by an evolutionary process to provide sufficient building block concentrations for DNA synthesis.

Summarizing, the concept of evolutionary adaption or optimization to fluctuating environments as well as experimental observations agree nicely with our theoretical investigation. Rather than altering optimal fluxes, we found a noteworthy change in yield distribution for both, G6P and GAP as carbon sources. We have also shown that the main contribution to the altered flux distribution is brought out by shutting

down nitrite reductase (NiR) and ASL. It has been shown experimentally that both factors are recognizable by CHLAMY1 (Kiaulehn et al. 2007; Iliev et al. 2006). In this project we restricted our investigation to one regulator. Even though we have shown that CHLAMY1 greatly influences flux distribution, it cannot be ruled out that other processes such as (post-)transcriptional or (post-)translational regulation influences nitrogen uptake as well. It has been shown e. g. that the transcription of several enzymes that are included in our model, are regulated in a circadian manner (Kuchó et al. 2005). In our approach, we left reactions unaffected whenever there is an isoenzyme present that is capable of conducting the identical reaction. Thus, including such transcriptional information into our model has no effect on simulation results, since for all affected reactions of our model, isoenzymes exist that are not circadian clock controlled. We are aware of the fact that our analysis was based on the assumption that all fluxes have an equal likelihood to occur, as has been done earlier by Kenanov et al. (2010) in the case of incomplete knowledge. Although this is a notable simplification, it allows for an efficient analysis of the models intrinsic robustness and adjustments during diurnal changes. Besides, approaching metabolic system with optimality principles bears a noteworthy dependency on applied constraints (Edwards et al. 2001; Ibarra et al. 2002; Famili et al. 2003) and meets with various difficulties such as scaling flux results to experimental results (Schuster et al. 2008; Feist and Palsson 2010).

One might also argue that treating regulation in a boolean on/off manner provides only a limited analysis. Therefore, we additionally calculated weighted arithmetic means of the carbon yield distribution, which allowed us to fine-tune the extend of the regulatory influence. The observation of changed yield distribution upon CHLAMY1 regulation persists, nevertheless, although with less clarity. Regardless of the chosen carbon source, our analysis had shown that during the night, the metabolic flux capability is reduced to about 2 % of all possible fluxes. Notably, fluxes with particularly low yields are down-regulated, which economizes energy and carbon sources. Therefore, down-regulation instead of inactivation leads to less pronounced yield alterations due to the large amount of affected EFMs. Starting with the day and CHLAMY1 inactivation, all fluxes are feasible again. Hence, the dependency on energy efficient routes becomes less critical due to the extra amount of available energy through photosynthesis. Since CHLAMY1 binds and stabilizes recognized mRNAs during the night, but releases the respective transcripts

with daylight, it ensures a high metabolic capacity as soon as additional energy is available. Thus, it enables a most efficient nitrogen fixation during night as well as day time.

Taken together, we have shown that one regulatory factor can have a tremendous impact on metabolic flux distributions. In consequence, it allows for an active metabolic system that corresponds optimal to the corresponding environmental condition. Moreover, we have proposed using weighted influences instead of active/inactive regulation, which might be useful for a number of applications, such as studying different impacts of multiple regulators or weighting EFMs according to fold changes given by microarray data. Finally, our findings with respect to *C. reinhardtii* might have implications for further studies and biotechnological design of e.g. biofuel production. Here, nitrogen starvation of the green algae is currently heavily studied (Bono et al. 2013; Ito et al. 2013; James et al. 2013).

4.2.3. Computing costs of metabolic compounds with FBA guides the search for suitable, efficient growth media for *E. coli*

By introducing the concept of EFMs and exemplifying a valuable application to a model organism, we demonstrated the usefulness of approaching larger scaled models by the means of constraint-based modeling. However, EFM analysis is only one possibility of a constraint-based modeling approach. Although not appropriate for the prior study, FBA has proven its value in many other studies, especially when assessing information about optimal growth behavior (Varma and Palsson 1994; Burgard and Maranas 2001; Segrè et al. 2002; Feist and Palsson 2010; Khandelwal et al. 2013; Chakrabarti et al. 2013).

Optimality or achieving the highest efficiency plays an important role in metabolic engineering as well. Here, *E. coli* became a widely used industrial organism for biosynthetic production such as recombinant proteins or low-weight amino acid compounds. Due to the notable dependency on a given growth media, determining the metabolic costs associated to this production is of high biotechnological interest. We described the required biosynthetic pathways for amino acid production in *E. coli* based on different carbon sources (see also Kaleta et al. (2013), chapter 3). Furthermore, we presented different types of costs, since the desired optimal solution may

differ given an alternative experimental or industrial setup. Primarily, the physiological conditions of a given organismic host, such as the amount of available ATP or amino acid precursors, determine the chosen cost type. Note that optimizing the ATP yield is a preferred goal, since energy balance is a prerequisite to assess biotechnological products of interest. Additionally, the experimental environment may contain different carbon source concentrations. Thus, different cost measures should be appreciated. In fact, the product yield itself, but also the ATP consumption of amino acid production pathways has hitherto been accepted to be primarily optimized (Stouthamer 1973; Akashi and Gojobori 2002). However, as our results show, an efficient ATP production often causes the product yield to be less optimal in the context of the consumed carbon sources. We have demonstrated that the biochemical routes that enable yield optimal fluxes can be non-trivial and differ from routes that are normally favored by cells. Interestingly, our analysis of flux distributions for protein syntheses are often accompanied by energy producing pathways, which can partially compensate the energy demand of amino acid polymerization. In the case of glycerol as carbon source the compensation is even nearly matched up to protein production costs. Our observations are in line with prior studies, where glycerol has been shown to be favorable compared to glucose as carbon source, especially once recombinant protein production is stably running (Luo et al. 2006; Berger et al. 2011).

We noted that our results are biased by the different molecular complexity and mass of different amino acids. To complete our findings with cost calculations circumventing this bias, we additionally computed carbon yields as well as costs on a kilodalton basis. These calculations reveal that in amino acid producing flux distributions most of the lost carbons are due to a carbon dioxide release accompanying the synthesizing process. In contrast, for half of the amino acids exist theoretical flux distributions that provide carbon fixation from carbon dioxide. Moreover, our calculations, which considered the mass of amino acids as well, showed that phenylalanine as well as methionine are the most expensive amino acids, if glucose is the carbon source. Presuming short canonical production routes for amino acid precursors, the synthesis of methionine from glucose consumes 18 mol of ATP. Since methionine is one of the most important synthetically produced amino acids (Ikeda 2003; Krömer et al. 2006; Park and Lee 2010b), this number might be of high relevance. Interestingly, our calculations showed that the largest amino acid

tryptophan is not the most expensive amino acid to produce, if the mass of amino acids is taken into account as well. Instead, when considering the concentrations of amino acids present in *E. coli*, the microbacteria invests most of its carbon sources into synthesizing leucine.

Taken together, our calculations based on different carbon sources and considering different cost measures, provide a comprehensive overview on metabolic production costs of amino acids as well as proteins. We expect our numbers to serve as a reference to guide the choice of optimal growth media for biotechnological relevant *E. coli* strains. In general, our calculations suggest that preferring glycerol as well as supplementing particularly expensive amino acids may notably improve the efficiency of synthetic production of biological compounds.

5. Conclusion

This thesis is dedicated to the study of complex dynamic processes in systems biology. I approached given study objects either with ODE based modeling techniques, or with constraint-based methods. The latter approach was used, once specific information about enzyme kinetics were too complex to assess and thus, not available, or if the computational demand for modeling and simulation was overwhelming and surpassed a feasible solution space. Both approaches enabled me to model and analyze system properties in an efficient manner, giving rise to valuable insights at the cost of reasonable and simplifying assumptions. Such assumptions comprise e.g. negligible space occurrences of model entities or mandatory balanced masses of intermediary metabolites in metabolic networks.

Even though well approved and valuable, studies in the life sciences are not limited to the here considered approaches. An ODE based approach is a promising method, when the modeled species number is high and thus, fluctuations and noise are comparably low. In contrast, investigations dedicated to a small number of entities may require a different approach. Here, the stochastic impact is noteworthy and should be taken into account as well. This can be achieved e.g. by modeling a system with the well known Gillespie algorithm to study the probability of species occurrences (Gillespie 1976) or by appreciating noise management in molecular systems (Bruggerman et al. 2009). Moreover, the systems studied in chapter 2 were assumed to be homogeneously distributed. Although well justified in these studies, different scales of biological phenomena as well as the number of comprised species may well infer a heterogeneously distributed population. Additionally to a potential noise effect when modeled species numbers are low, here, spatial dynamics need to be taken into account as well. To deal with spatiality, further methods are required, such as partial differential equations, or individual-based modeling. These are commonly used in ecology and systems biology (Grimm et al. 2005; Martin et al. 2013).

Approaching larger scaled systems poses particularly the challenge to keep a problem solvable. Most commonly, studying dynamic models based e.g. on kinetics is

prohibited due to its computational demand in conjugation with the regular lack of knowledge of velocity affecting parameters. As I have shown in chapter 3, approaching this class of biological problems with constraint-based methods provides valuable insights into model properties, while keeping the problem in a solvable region. In order to cope with emerging technological advances, such as next generation sequencing techniques and subsequent model generation up to genome scale, several extension to EFM and FBA have been proposed. Notably, EFM based methods were developed to enable its application to genome scale models (Kaleta et al. 2009a,b; de Figueiredo et al. 2009). Concerning FBA, the method has been extended to account e. g. for multiple optimal fluxes (Mahadevan et al. 2002) or potential cross feeding in microbial communities (Khandelwal et al. 2013).

Among others, these efforts exemplify the state-of-the-art of both methods. They are a promising alternative, when dynamic modeling efforts fail, even if it comes with an acceptable decrease in the predictive power of the dynamics of a model. Particularly the recent advances in generating context specific models based on expression data in an automatic manner (Jerby et al. 2010; Zur et al. 2010; Jensen and Papin 2010; Wang et al. 2012) will potentially deliver further applications of the described methods in the future.

Finally, if we think again back to Newton and his study of planet movements, it is regularly neither wise nor appropriate to model and study as many details as possible in order to draw conclusion about the underlying mechanistic principles. Without the knowledge of Einsteins's theory of relativity, Newton was still able to propose a realistic model of gravity, which essentially influenced further research not only in astro physics. In contrast, appreciating the scale of detail of an investigated phenomena should always guide the choice in the utilized tool or method, to allow for an efficient yet valuable and meaningful analysis.

Bibliography

- J. C. Acosta, A. O’Loghlen, A. Banito, M. V. Guijarro, A. Augert, S. Raguz, M. Fumagalli, M. Da Costa, C. Brown, N. Popov, Y. Takatsu, J. Melamed, F. d’Adda di Fagagna, D. Bernard, E. Hernando, and J. Gil. Chemokine signaling via the CXCR2 receptor reinforces senescence. *Cell*, 133(6):1006–1018, 2008.
- H. Akashi and T. Gojobori. Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA*, 99(6):3695–700, 2002.
- D. J. Baker, T. Wijshake, T. Tchkonja, N. K. LeBrasseur, B. G. Childs, B. van de Sluis, J. L. Kirkland, and J. M. van Deursen. Clearance of p16Ink4a-positive senescent cells delays ageing-associated disorders. *Nature*, 479(7372):232–236, 2011.
- T. Baks, A. E. M. Janssen, and R. M. Boom. A kinetic model to explain the maximum in alpha-amylase activity measurements in the presence of small carbohydrates. *Biotechnology and bioengineering*, 94(3):431–40, 2006.
- F. Baneyx and M. Mujacic. Recombinant protein folding and misfolding in *Escherichia coli*. *Nat Biotechnol*, 22(11):1399–1408, 2004.
- J. Behre, T. Wilhelm, A. von Kamp, E. Ruppin, and S. Schuster. Structural robustness of metabolic networks with respect to multiple knockouts. *J Theor Biol*, 252(3):433–441, 2008.
- I. Ben-Porath and R. A. Weinberg. When cells get stressed: an integrative view of cellular senescence. *J Clin Invest*, 113(1):8–13, 2004.
- I. Ben-Porath and R. A. Weinberg. The signals and pathways activating cellular senescence. *Int J Biochem Cell Biol*, 37(5):961–976, 2005.
- J. A. Benanti and D. A. Galloway. Normal human fibroblasts are resistant to RAS-induced senescence. *Mol Cell Biol*, 24(7):2842–2852, 2004.

- C. Berger, C. Montag, S. Berndt, and D. Huster. Optimization of *Escherichia coli* cultivation methods for high yield neuropeptide Y receptor type 2 production. *Protein Expr Purif*, 76(1):25–35, 2011.
- D. J. V. Beste, B. Bonde, N. Hawkins, J. L. Ward, M. H. Beale, S. Noack, K. Nöh, N. J. Kruger, R. G. Ratcliffe, and J. McFadden. ^{13}C Metabolic Flux Analysis Identifies an Unusual Route for Pyruvate Dissimilation in Mycobacteria which Requires Isocitrate Lyase and Carbon Dioxide Fixation. *PLoS Pathog*, 7(7):e1002091, 2011.
- M. S. Bono, B. A. Ahner, and B. J. Kirby. Detection of algal lipid accumulation due to nitrogen limitation via dielectric spectroscopy of *Chlamydomonas reinhardtii* suspensions in a coaxial transmission line sample cell. *Bioresour Technol*, 143: 623–631, 2013.
- N. R. Boyle and J. A. Morgan. Flux balance analysis of primary metabolism in *Chlamydomonas reinhardtii*. *BMC Syst Biol*, 3:4, 2009.
- G. E. Briggs and J. B. Haldane. A note on the kinetics of enzyme action. *Biochem J*, 19(2):338–339, 1925.
- F. J. Bruggeman, N. Blüthgen, and H. V. Westerhoff. Noise management by molecular networks. *PLoS Comput Biol*, 5(9):e1000506, 2009.
- A. P. Burgard and C. D. Maranas. Probing the performance limits of the *Escherichia coli* metabolic network subject to gene additions or deletions. *Biotechnol Bioeng*, 74(5):364–375, 2001.
- J. Campisi and J. Sedivy. How does proliferative homeostasis change with age? What causes it and how does it contribute to aging? *J Gerontol A Biol Sci Med Sci*, 64(2):164–166, 2009.
- R. Carlson, D. Fell, and F. Sreenc. Metabolic pathway analysis of a recombinant yeast for rational strain development. *Biotechnol Bioeng*, 79(2):121–134, 2002.
- R. P. Carlson. Metabolic systems cost-benefit analysis for interpreting network structure and regulation. *Bioinformatics*, 23(10):1258–1264, 2007.

- M. Cascante, E. I. Canela, and R. Franco. Control analysis of systems having two steps catalyzed by the same protein molecule in unbranched chains. *Eur J Biochem*, 192(2):369–371, 1990.
- T. Çakir, C. S. Tacer, and K. O. Ülgen. Metabolic pathway analysis of enzyme-deficient human red blood cells. *Biosystems*, 78(1-3):49–67, 2004.
- A. Chakrabarti, L. Miskovic, K. C. Soh, and V. Hatzimanikatis. Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiology constraints. *Biotechnol J*, 2013.
- V. Chaturvedi, J. Z. Qin, M. F. Denning, D. Choubey, M. O. Diaz, and B. J. Nickoloff. Apoptosis in proliferating, senescent, and immortalized keratinocytes. *J Biol Chem*, 274(33):23358–23367, 1999.
- T. C. Chou and P. Talaly. A simple generalized equation for the analysis of multiple inhibitions of michaelis-menten kinetic systems. *J Biol Chem*, 252(18):6438–6442, 1977.
- B. L. Clarke. Stoichiometric network analysis. *Cell Biophys*, 12:237–253, 1988.
- W. W. Cleland. The kinetics of enzyme-catalyzed reactions with two or more substrates or products. I. Nomenclature and rate equations. *Biochim Biophys Acta*, 67:104–137, 1963.
- I. H. G. S. Consortium. Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011):931–945, 2004.
- J.-P. Coppé, P.-Y. Desprez, A. Krtolica, and J. Campisi. The senescence-associated secretory phenotype: the dark side of tumor suppression. *Annu Rev Pathol*, 5: 99–118, 2010.
- W. Cosme-Blanco and S. Chang. Dual roles of telomere dysfunction in initiation and suppression of tumorigenesis. *Exp Cell Res*, 314(9):1973–1979, 2008.
- N. M. Crawford. Nitrate: nutrient and signal for plant growth. *Plant Cell*, 7(7): 859–868, 1995.
- N. M. Crawford and H. N. Arst. The molecular genetics of nitrate assimilation in fungi and plants. *Annu Rev Genet*, 27:115–146, 1993.

- V. J. Cristofalo, A. Lorenzini, R. G. Allen, C. Torres, and M. Tresini. Replicative senescence: a critical review. *Mech Ageing Dev*, 125(10-11):827–848, 2004.
- A. Csikász-Nagy, D. Battogtokh, K. C. Chen, B. Novák, and J. J. Tyson. Analysis of a generic model of eukaryotic cell-cycle regulation. *Biophys J*, 90(12):4361–4379, 2006.
- F. d’Adda di Fagagna, P. M. Reaper, L. Clay-Farrace, H. Fiegler, P. Carr, T. Von Zglinicki, G. Saretzki, N. P. Carter, and S. P. Jackson. A DNA damage checkpoint response in telomere-initiated senescence. *Nature*, 426(6963):194–198, 2003.
- F. Daniel-Vedele, S. Filleur, and M. Caboche. Nitrate transport: a key step in nitrate assimilation. *Curr Opin Plant Biol*, 1:235–239, 1998.
- G. B. Dantzig, A. Orden, and P. Wolfe. The Generalized Simplex Method for Minimizing a Linear Form Under Linear Inequality Restraints. *Pacific Journal Math*, 5:183–195, 1955.
- H. Dathe, K. Prager, and M. Mittag. Novel interaction of two clock-relevant RNA-binding proteins C3 and XRN1 in *Chlamydomonas reinhardtii*. *FEBS Lett*, 586(22):3969–3973, 2012.
- L. F. de Figueiredo, A. Podhorski, A. Rubio, C. Kaleta, J. E. Beasley, S. Schuster, and F. J. Planes. Computing the shortest elementary flux modes in genome-scale metabolic networks. *Bioinformatics*, 25(23):3158–3165, 2009.
- D. De Vos, F. J. Bruggeman, H. V. Westerhoff, and B. M. Bakker. How molecular competition influences fluxes in gene expression networks. *PLoS one*, 6(12):e28494, 2011.
- F. Debacq-Chainiaux, C. Borlon, T. Pascal, V. Royer, F. Eliaers, N. Ninane, G. Carrard, B. Friguet, F. de Longueville, S. Boffe, J. Remacle, and O. Toussaint. Repeated exposure of human skin fibroblasts to UVB at subcytotoxic level triggers premature senescence through the TGF-beta1 signaling pathway. *J Cell Sci*, 118(Pt 4):743–758, 2005.
- A. Di Leonardo, S. P. Linke, K. Clarkin, and G. M. Wahl. DNA damage triggers a prolonged p53-dependent G1 arrest and long-term induction of Cip1 in normal human fibroblasts. *Genes Dev*, 8(21):2540–2551, 1994.

- G. P. Dimri, X. Lee, G. Basile, M. Acosta, G. Scott, C. Roskelley, E. E. Medrano, M. Linskens, I. Rubelj, and O. Pereira-Smith. A biomarker that identifies senescent human cells in culture and in aging skin in vivo. *Proc Natl Acad Sci U S A*, 92(20):9363–9367, 1995.
- G. P. Dimri, K. Itahana, M. Acosta, and J. Campisi. Regulation of a senescence checkpoint response by the E2F1 transcription factor and p14(ARF) tumor suppressor. *Mol Cell Biol*, 20(1):273–285, 2000.
- V. L. S. Dingerkus, T. J. Gaber, K. Helmbold, S. Bubenzer, A. Eisert, C. L. Sánchez, and F. D. Zepf. Acute tryptophan depletion in accordance with body weight: influx of amino acids across the blood-brain barrier. *J Neural Transm*, 119(9):1037–45, 2012.
- N. C. Duarte, S. A. Becker, N. Jamshidi, I. Thiele, M. L. Mo, T. D. Vo, R. Srivas, and B. Ø. Palsson. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A*, 104(6):1777–1782, 2007.
- J. S. Edwards and B. Ø. Palsson. Systems Properties of the *Haemophilus influenzae* Rd Metabolic Genotype. *J Biol Chem*, 274(25):17410–17416, 1999.
- J. S. Edwards, R. U. Ibarra, and B. Ø. Palsson. In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat Biotechnol*, 19:125–130, 2001.
- I. Famili, J. Forster, J. Nielsen, and B. Ø. Palsson. *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proc Natl Acad Sci U S A*, 100(23):13134–13139, 2003.
- R. G. Faragher, I. R. Kill, J. A. Hunter, F. M. Pope, C. Tannock, and S. Shall. The gene responsible for Werner syndrome may be a cell division "counting" gene. *Proc Natl Acad Sci U S A*, 90(24):12030–12034, 1993.
- A. M. Feist and B. Ø. Palsson. The biomass objective function. *Curr Opin Microbiol*, 2010.

- A. M. Feist, C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis, and B. Ø. Palsson. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol*, 3:121, 2007.
- A. M. Feist, M. J. Herrgård, I. Thiele, J. L. Reed, and B. Ø. Palsson. Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol*, 7(2):129–143, 2009.
- D. A. Fell and J. R. Small. Fat synthesis in adipose tissue. an examination of stoichiometric constraints. *Biochem J*, 238(3):781–786, 1986.
- E. Fernández and A. Galván. Inorganic nitrogen assimilation in *Chlamydomonas*. *J Exp Bot*, 58:2279–2287, 2007.
- E. Fischer and U. Sauer. A novel metabolic cycle catalyzes glucose oxidation and anaplerosis in hungry *Escherichia coli*. *J Biol Chem*, 278(47):46446–46451, 2003.
- J. Gagneur and S. Klamt. Computation of elementary modes: a unifying framework and the new binary approach. *BMC Bioinformatics*, 5:175, 2004.
- J. Gebauer, S. Schuster, L. F. de Figueiredo, and C. Kaleta. Detecting and investigating substrate cycles in a genome-scale human metabolic network. *FEBS J*, 279(17):3192–3202, 2012.
- D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, 1976.
- V. Gire, P. Roux, D. Wynford-Thomas, J.-M. Brondello, and V. Dulic. DNA damage checkpoint kinase Chk2 triggers replicative senescence. *EMBO J*, 23(13):2554–2563, 2004.
- A. Golubev, S. Khrustalev, and A. Butov. An in silico investigation into the causes of telomere length heterogeneity and its implications for the Hayflick limit. *J Theor Biol*, 225(2):153–170, 2003.
- A. A. Goodarzi, A. T. Noon, D. Deckbar, Y. Ziv, Y. Shiloh, M. Löbrich, and P. A. Jeggo. ATM signaling facilitates repair of DNA double-strand breaks associated with heterochromatin. *Mol Cell*, 31(2):167–177, 2008.

- V. Grimm, E. Revilla, U. Berger, F. Jeltsch, W. M. Mooij, S. F. Railsback, H.-H. Thulke, J. Weiner, T. Wiegand, and D. L. DeAngelis. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *Science*, 310(5750):987–991, 2005.
- O. Hädicke and S. Klamt. CASOP: A computational approach for strain optimization aiming at high productivity. *J Biotechnol*, 147(2):88–101, 2010.
- A. Hall. *Philosophers at War: The Quarrel between Newton and Leibniz*. Cambridge University Press, 1980.
- W. R. Harcombe, N. F. Delaney, N. Leiby, N. Klitgord, and C. J. Marx. The ability of flux balance analysis to predict evolution of central metabolism scales with the initial distance to the optimum. *PLoS Comput Biol*, 9(6):e1003091, 2013.
- E. H. Harris, editor. *The Chlamydomonas Sourcebook*. Academic Press, 2008.
- A. M. Havelka, M. Berndtsson, M. H. Olofsson, M. C. Shoshan, and S. Linder. Mechanisms of action of DNA-damaging anticancer drugs in treatment of carcinomas: is acute apoptosis an "off-target" effect? *Mini Rev Med Chem*, 7(10):1035–1039, 2007.
- L. Hayflick and P. S. Moorhead. The serial cultivation of human diploid cell strains. *Exp Cell Res*, 25:585–621, 1961.
- R. Heinrich and T. A. Rapoport. A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. *Eur J Biochem*, 42(1):89–95, 1974.
- R. Heinrich and S. Schuster. *The Regulation of Cellular Systems*. Chapman and Hall, New York, 1996.
- E. M. Heizer, D. W. Raiford, M. L. Raymer, T. E. Doom, R. V. Miller, and D. E. Krane. Amino acid cost and codon-usage biases in 6 prokaryotic genomes: a whole-genome analysis. *Mol Biol Evol*, 23(9):1670–1680, 2006.
- V. Henri. Théorie générale de l'action de quelques diastases. *Compt. Rend. Hebd. Acad. Sci. Paris*, 135:916–919, 1902.
- V. Henri. *Lois générales de l'action des diastases*. Herman, 1903.

- U. Herbig, W. A. Jobling, B. P. C. Chen, D. J. Chen, and J. M. Sedivy. Telomere shortening triggers senescence of human cells through a pathway involving ATM, p53, and p21(CIP1), but not p16(INK4a). *Mol Cell*, 14(4):501–513, 2004.
- R. Hooke and T. A. Jeeves. "Direct Search" Solution of Numerical and Statistical Problems. *Journal of the ACM*, 2:212–229, 1961.
- E. S. Hwang. Replicative senescence and senescence-like state induced in cancer-derived cells. *Mech Ageing Dev*, 123(12):1681–1694, 2002.
- R. U. Ibarra, J. S. Edwards, and B. Ø. Palsson. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature*, 420:186–189, 2002.
- T. Ideker, T. Galitski, and L. Hood. A new approach to decoding life: systems biology. *Annu Rev Genomics Hum Genet*, 2:343–372, 2001.
- M. Ikeda. Amino acid production processes. *Adv. Biochem. Eng. Biotechnol.*, 79:1–35, 2003.
- D. Iliev, O. Voytsekh, E.-M. Schmidt, M. Fiedler, A. Nykytenko, and M. Mittag. A heteromeric RNA-binding protein is involved in maintaining acrophase and period of the circadian clock. *Plant Physiol*, 142:797–806, 2006.
- K. Itahana, Y. Zou, Y. Itahana, J.-L. Martinez, C. Beausejour, J. J. L. Jacobs, M. Van Lohuizen, V. Band, J. Campisi, and G. P. Dimri. Control of the replicative life span of human fibroblasts by p16 and the polycomb protein Bmi-1. *Mol Cell Biol*, 23(1):389–401, 2003.
- T. Ito, M. Sugimoto, Y. Toya, Y. Ano, N. Kurano, T. Soga, and M. Tomita. Time-resolved metabolomics of a novel trebouxiohycean alga using (13)CO2 feeding. *J Biosci Bioeng*, 2013.
- G. O. James, C. H. Hocart, W. Hillier, G. D. Price, and M. A. Djordjevic. Temperature modulation of fatty acid profiles for biofuel production in nitrogen deprived chlamydomonas reinhardtii. *Bioresour Technol*, 127:441–447, 2013.
- P. A. Jensen and J. A. Papin. Functional Integration of a Metabolic Network Model and Expression Data without Arbitrary Thresholding. *Bioinformatics*, 2010.

- L. Jerby, T. Shlomi, and E. Ruppin. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Mol Syst Biol*, 6:401, 2010.
- G. Jetschke. *Mathematik der Selbstorganisation*. Verlag Harri Deutsch, Frankfurt am Main, 2009.
- J. C. Jeyapalan, M. Ferreira, J. M. Sedivy, and U. Herbig. Accumulation of senescent cells in mitotic tissue of aging primates. *Mech Ageing Dev*, 128(1):36–44, 2007.
- L. Jöres and R. Wagner. Essential steps in the ppGpp-dependent regulation of bacterial ribosomal RNA promoters can be explained by substrate competition. *The Journal of biological chemistry*, 278(19):16834–43, 2003.
- H. Kacser and J. A. Burns. The control of flux. *Symp Soc Exp Biol*, 27:65–104, 1973.
- C. Kaleta, L. F. de Figueiredo, J. Behre, and S. Schuster. EFMEvolver: Computing elementary flux modes in genome-scale metabolic networks. In I. Grosse, S. Neumann, S. Posch, F. Schreiber, and P. Stadler, editors, *Lecture Notes in Informatics - Proceedings*, volume P-157, pages 179–189, Bonn, 2009a. Gesellschaft für Informatik.
- C. Kaleta, L. F. de Figueiredo, and S. Schuster. Can the whole be less than the sum of its parts? Pathway Analysis in Genome-Scale Metabolic Networks using Elementary Flux Patterns. *Genome Res*, 19(10):1872–1883, 2009b.
- C. Kaleta, S. Schäuble, U. Rinas, and S. Schuster. Metabolic costs of amino acid and protein production in *Escherichia coli*. *Biotechnol J*, 8(9):1105–1114, 2013.
- N. Karmarkar. A New Polynomial-Time Algorithm for Linear Programming. *Combinatorica*, 4:373–395, 1984.
- O. Kartal, S. Mahlow, A. Skupin, and O. Ebenhöf. Carbohydrate-active enzymes exemplify entropic principles in metabolism. *Mol Sys Biol*, 7:542, 2011.
- A. Kayser, J. Weber, V. Hecht, and U. Rinas. Metabolic flux analysis of *Escherichia coli* in glucose-limited continuous culture. I. Growth-rate-dependent metabolic efficiency at steady state. *Microbiology*, 151(Pt 3):693–706, 2005.

- D. Kenanov, C. Kaleta, A. Petzold, C. Hoischen, S. Diekmann, R. A. Siddiqui, and S. Schuster. Theoretical study of lipid biosynthesis in wild-type *Escherichia coli* and in a protoplast-type L-form using elementary flux mode analysis. *FEBS J*, 277(4):1023–1034, 2010.
- R. A. Khandelwal, B. G. Olivier, W. F. M. Röling, B. Teusink, and F. J. Bruggeman. Community flux balance analysis for microbial consortia at balanced growth. *PLoS One*, 8(5):e64567, 2013.
- S. Kiaulehn, O. Voytsekh, M. Fuhrmann, and M. Mittag. The Presence of UG-repeat sequences in the 3'-UTRs of reporter luciferase mRNAs mediates circadian expression and can determine acrophase in *Chlamydomonas reinhardtii*. *J Biol Rhythms*, 22:275–277, 2007.
- I. R. Kill, R. G. Faragher, K. Lawrence, and S. Shall. The expression of proliferation-dependent antigens during the lifespan of normal and progeroid human fibroblasts in culture. *J Cell Sci*, 107 (Pt 2):571–579, 1994.
- T. Y. Kim, S. B. Sohn, Y. B. Kim, W. J. Kim, and S. Y. Lee. Recent advances in reconstruction and applications of genome-scale metabolic models. *Curr Opin Biotechnol*, 23(4):617–623, 2012.
- Y. Kim, M. J. Andreu, B. Lim, K. Chung, M. Terayama, G. Jiménez, H. Berg, Celeste a Lu, and S. Y. Shvartsman. Gene regulation by MAPK substrate competition. *Developmental cell*, 20(6):880–7, 2011.
- E. L. King and C. Altman. A schematic method of deriving the rate laws for enzyme-catalyzed reactions. *J Phys Chem*, 60:1375–1378, 1956.
- H. Kitano. Systems biology: a brief overview. *Science*, 295(5560):1662–1664, 2002.
- S. Klamt, J. Gagneur, and A. von Kamp. Algorithmic approaches for computing elementary modes in large biochemical reaction networks. *Syst Biol (Stevenage)*, 152(4):249–255, 2005.
- S. Klamt. Generalized concept of minimal cut sets in biochemical networks. *Biosystems*, 83(2-3):233–247, 2006.
- S. Klamt and J. Stelling. Combinatorial complexity of pathway analysis in metabolic networks. *Mol Biol Rep*, 29(1-2):233–236, 2002.

- K. Klement, C. Melle, U. Murzik, S. Diekmann, J. Norgauer, and P. Hemmerich. Accumulation of annexin A5 at the nuclear envelope is a biomarker of cellular aging. *Mech Ageing Dev*, 2012.
- D. J. Korz, U. Rinas, K. Hellmuth, E. A. Sanders, and W. D. Deckwer. Simple fed-batch technique for high cell density cultivation of *Escherichia coli*. *J. Biotechnol.*, 39(1):59–65, 1995.
- M. Kosar, J. Bartkova, S. Hubackova, Z. Hodny, J. Lukas, and J. Bartek. Senescence-associated heterochromatin foci are dispensable for cellular senescence, occur in a cell type- and insult-dependent manner and follow expression of p16(ink4a). *Cell Cycle*, 10(3):457–468, 2011.
- V. Krizhanovsky, W. Xue, L. Zender, M. Yon, E. Hernando, and S. W. Lowe. Implications of cellular senescence in tissue damage response, tumor suppression, and stem cell biology. *Cold Spring Harb Symp Quant Biol*, 73:513–522, 2008.
- J. O. Krömer, C. Wittmann, H. Schröder, and E. Heinzle. Metabolic pathway analysis for rational design of L-methionine production by *Escherichia coli* and *Corynebacterium glutamicum*. *Metab Eng*, 8:353–369, 2006.
- K.-I. Kucho, K. Okamoto, S. Tabata, H. Fukuzawa, and M. Ishiura. Identification of novel clock-controlled genes by cDNA macroarray analysis in *Chlamydomonas reinhardtii*. *Plant Mol Biol*, 57(6):889–906, 2005.
- T. Kuilman and D. S. Peeper. Senescence-messaging secretome: SMS-ing cellular stress. *Nat Rev Cancer*, 9(2):81–94, 2009.
- T. Kuilman, C. Michaloglou, W. J. Mooi, and D. S. Peeper. The essence of senescence. *Genes Dev*, 24(22):2463–2479, 2010.
- W. Kutta. Beitrag zur näherungsweise integration von differentialgleichungen. *Z. Math. und Phys.*, 46:435–453, 1901.
- H. M. Lam, K. Coschigano, C. Schultz, R. Melo-Oliveira, G. Tjaden, I. Oliveira, N. Ngai, M. H. Hsieh, and G. Coruzzi. Use of arabidopsis mutants and genes to study amide amino acid biosynthesis. *Plant Cell*, 7(7):887–898, 1995.

E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczkzy, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chisoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blöcker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kasprzyk,

- S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen, J. Szustakowski, and I. H. G. S. Consortium. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921, 2001.
- C. Lawless, C. Wang, D. Jurk, A. Merz, T. von Zglinicki, and J. F. Passos. Quantitative assessment of markers for cell senescence. *Exp Gerontol*, 45(10):772–778, 2010.
- S. P. Lee, M. L. Censullo, H. G. Kim, and M. K. Han. Substrate-length-dependent activities of human immunodeficiency virus type 1 integrase in vitro: differential DNA binding affinities associated with different lengths of substrates. *Biochemistry*, 34(32):10215–23, 1995.
- S. Y. Lee. High cell-density culture of *Escherichia coli*. *Trends Biotechnol.*, 14(3):98–105, 1996.
- N. E. Lewis, K. K. Hixson, T. M. Conrad, J. A. Lerman, P. Charusanti, A. D. Polpitiya, J. N. Adkins, G. Schramm, S. O. Purvine, D. Lopez-Ferrer, K. K. Weitz, R. Eils, R. König, R. D. Smith, and B. Ø. Palsson. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol Syst Biol*, 6:390, 2010.
- N. E. Lewis, H. Nagarajan, and B. Ø. Palsson. Constraining the metabolic genotype-phenotype relationship using a phylogeny of *in silico* methods. *Nat Rev Microbiol*, 10(4):291–305, 2012.
- J. C. Liao, S. Y. Hou, and Y. P. Chao. Pathway analysis, engineering, and physiological considerations for redirecting central metabolism. *Biotechnol Bioeng*, 52(1):129–140, 1996.
- D. Ludwig, D. D. Jones, and C. S. Holling. Qualitative analysis of Insect Outbreak

- Systems: Spruce Budworm and Forest. *The Journal of Animal Ecology*, 47(1): 315–332, 1978.
- G. W. Luli and W. R. Strohl. Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations. *Appl. Environ. Microbiol.*, 56(4):1004–11, 1990.
- Q. Luo, Y. L. Shen, D. Z. Wei, and W. Cao. Optimization of culture on the overproduction of TRAIL in high-cell-density culture by recombinant *Escherichia coli*. *Appl Microbiol Biotechnol*, 71(2):184–91, 2006.
- R. Mahadevan and C. H. Schilling. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng*, 5(4):264–276, 2003.
- R. Mahadevan, J. S. Edwards, and F. J. Doyle. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys J*, 83(3):1331–1340, 2002.
- S. C. Makrides. Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol Rev*, 60(3):512–538, 1996.
- F. A. Mallette, M.-F. Gaumont-Leclerc, and G. Ferbeyre. The DNA damage signaling pathway is a critical mediator of oncogene-induced senescence. *Genes Dev*, 21(1):43–48, 2007.
- R. Marcotte, C. Lacelle, and E. Wang. Senescent fibroblasts resist apoptosis by downregulating caspase-3. *Mech Ageing Dev*, 125(10-11):777–783, 2004.
- M. Margulies, M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y.-J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. I. Alenquer, T. P. Jarvie, K. B. Jirage, J.-B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, and J. M. Rothberg. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057):376–380, 2005.

- B. Marquez and F. Van Bambeke. ABC multidrug transporters: target for modulation of drug pharmacokinetics and drug-drug interactions. *Current drug targets*, 12(5):600–20, 2011.
- B. T. Martin, T. Jager, R. M. Nisbet, T. G. Preuss, and V. Grimm. Predicting population dynamics from the properties of individuals: a cross-level test of dynamic energy budget theory. *Am Nat*, 181(4):506–519, 2013.
- M. L. Mavrovouniotis, G. Stephanopoulos, and G. Stephanopoulos. Computer-aided synthesis of biochemical pathways. *Biotechnol Bioeng*, 36(11):1119–1132, 1990.
- M. C. McKenna. Substrate competition studies demonstrate oxidative metabolism of glucose, glutamate, glutamine, lactate and 3-hydroxybutyrate in cortical astrocytes from rat brain. *Neurochemical research*, 37(11):2613–26, 2012.
- L. Michaelis and M. Menten. Die Kinetik der Invertinwirkung. *Biochem. Z.*, 49:333–369, 1913.
- M. Mittag. Conserved circadian elements in phylogenetically diverse algae. *Proc Natl Acad Sci U S A*, 93:14401–14404, 1996.
- A. J. Nakamura, C. E. Redon, W. M. Bonner, and O. A. Sedelnikova. Telomere-dependent and telomere-independent origins of endogenous DNA damage in tumor cells. *Aging (Albany NY)*, 1(2):212–218, 2009.
- M. Narita, S. N. nez, E. Heard, M. Narita, A. W. Lin, S. A. Hearn, D. L. Spector, G. J. Hannon, and S. W. Lowe. Rb-mediated heterochromatin formation and silencing of E2F target genes during cellular senescence. *Cell*, 113(6):703–716, 2003.
- M. A. Oberhardt, B. Ø. Palsson, and J. A. Papin. Applications of genome-scale metabolic reconstructions. *Mol Syst Biol*, 5:320, 2009.
- O. L. Oke. Nitrite Toxicity to Plants. *Nature*, 212:528, 1966.
- C. Pantoja and M. Serrano. Murine fibroblasts lacking p21 undergo senescence and are resistant to transformation by oncogenic Ras. *Oncogene*, 18(35):4974–4982, 1999.

- J. A. Papin, J. Stelling, N. D. Price, S. Klamt, S. Schuster, and B. Ø. Palsson. Comparison of network-based pathway analysis methods. *Trends Biotechnol*, 22(8):400–405, 2004.
- W. M. Pardridge and T. B. Choi. Neutral amino acid transport at the human blood-brain barrier. *Federation proceedings*, 45(7):2073–8, 1986.
- J. H. Park and S. Y. Lee. Fermentative production of branched chain amino acids: a focus on metabolic engineering. *Appl. Microbiol. Biotechnol.*, 85(3):491–506, 2010a.
- J. H. Park and S. Y. Lee. Metabolic pathways and fermentative production of l-aspartate family amino acids. *Biotechnol. J.*, 5(6):560–77, 2010b.
- J. F. Passos, G. Nelson, C. Wang, T. Richter, C. Simillion, C. J. Proctor, S. Miwa, S. Olijslagers, J. Hallinan, A. Wipat, G. Saretzki, K. L. Rudolph, T. B. L. Kirkwood, and T. von Zglinicki. Feedback between p21 and reactive oxygen production is necessary for cell senescence. *Mol Syst Biol*, 6:347, 2010.
- M. Pearson, R. Carbone, C. Sebastiani, M. Cioce, M. Fagioli, S. Saito, Y. Higashimoto, E. Appella, S. Minucci, P. P. Pandolfi, and P. G. Pelicci. PML regulates p53 acetylation and premature senescence induced by oncogenic Ras. *Nature*, 406(6792):207–210, 2000.
- T. Pfeiffer, I. Sánchez-Valdenebro, J. C. Nuño, F. Montero, and S. Schuster. METATOOL: for studying metabolic networks. *Bioinformatics*, 15(3):251–257, 1999.
- M. Piques, W. X. Schulze, M. Höhne, B. Usadel, Y. Gibon, J. Rohwer, and M. Stitt. Ribosome and transcript copy numbers, polysome occupancy and enzyme dynamics in *Arabidopsis*. *Mol Syst Biol*, 5:314, 2009.
- J. R. Pomeroy. Uncovering mechanisms of bistability in biological systems. *Current opinion in biotechnology*, 19(4):381–8, 2008.
- M. G. Poolman, D. A. Fell, and C. A. Raines. Elementary modes analysis of photosynthate metabolism in the chloroplast stroma. *Eur J Biochem*, 270(3):430–439, 2003.

- R. D. Portugal, M. G. P. Land, and B. F. Svaiter. A computational model for telomere-dependent cell-replicative aging. *Biosystems*, 91(1):262–267, 2008.
- N. D. Price, J. L. Reed, and B. Ø. Palsson. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat Rev Microbiol*, 2(11):886–897, 2004.
- A. Raue, V. Becker, U. Klingmüller, and J. Timmer. Identifiability and observability analysis for experimental design in nonlinear dynamical models. *Chaos*, 20(4):045105, 2010.
- C. Reder. Metabolic control theory: a structural approach. *J Theor Biol*, 135(2):175–201, 1988.
- S. Ressler, J. Bartkova, H. Niederegger, J. Bartek, K. Scharffetter-Kochanek, P. Jansen-Dürr, and M. Wlaschek. p16INK4A is a robust in vivo biomarker of cellular aging in human skin. *Aging Cell*, 5(5):379–389, 2006.
- U. Rinas, K. Hellmuth, R. Kang, A. Seeger, and H. Schlieker. Entry of *Escherichia coli* into stationary phase is indicated by endogenous and exogenous accumulation of nucleobases. *Appl. Environ. Microbiol.*, 61(12):4147–51, 1995.
- S. J. Robles and G. R. Adami. Agents that cause DNA double strand breaks lead to p16INK4a enrichment and the premature senescence of normal fibroblasts. *Oncogene*, 16(9):1113–1123, 1998.
- J. D. Rochaix, M. Goldschmidt-Clermont, and M. S., editors. *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.
- F. Rodier and J. Campisi. Four faces of cellular senescence. *J Cell Biol*, 192(4):547–556, 2011.
- I. Rubelj and Z. Vondracek. Stochastic mechanism of cellular aging—abrupt telomere shortening as a model for stochastic nature of cellular aging. *J Theor Biol*, 197(4):425–438, 1999.
- C. Runge. Über die numerische Auflösung von Differentialgleichungen. *Math. Ann*, pages 167–178, 1895.

- E. Ruppin, J. A. Papin, L. F. de Figueiredo, and S. Schuster. Metabolic reconstruction, constraint-based analysis and game theory to probe genome-scale metabolic networks. *Curr Opin Biotechnol*, 21(4):502–510, 2010.
- J. M. Savinell and B. Ø. Palsson. Optimal selection of metabolic fluxes for in vivo measurement. I. Development of mathematical methods. *J Theor Biol*, 155(2):201–214, 1992a.
- J. M. Savinell and B. Ø. Palsson. Network analysis of intermediary metabolism using linear optimization. II. Interpretation of hybridoma cell metabolism. *J Theor Biol*, 154(4):455–473, 1992b.
- S. Schäuble, I. Heiland, O. Voytsekh, M. Mittag, and S. Schuster. Predicting the Physiological Role of Circadian Metabolic Regulation in the Green Alga *Chlamydomonas reinhardtii*. *PLoS One*, 6(8):e23026, 2011a.
- S. Schäuble, S. Schuster, and C. Kaleta. Hands-on metabolism analysis of complex biochemical networks using elementary flux modes. *Methods Enzymol*, 500:437–456, 2011b.
- S. Schäuble, K. Klement, S. Marthandan, S. Münch, I. Heiland, S. Schuster, P. Hemmerich, and S. Diekmann. Quantitative model of cell cycle arrest and cellular senescence in primary human fibroblasts. *PLoS One*, 7(8):e42150, 2012.
- S. Schäuble, A. K. Stavrum, P. Puntervoll, S. Schuster, and I. Heiland. Effect of substrate competition in kinetic models of metabolic networks. *FEBS Lett*, 587(17):2818–2824, 2013.
- M. B. Scholz, C.-C. Lo, and P. S. G. Chain. Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Curr Opin Biotechnol*, 23(1):9–15, 2012.
- R. Schuetz, N. Zamboni, M. Zampieri, M. Heinemann, and U. Sauer. Multidimensional optimality of microbial metabolism. *Science*, 336(6081):601–604, 2012.
- P. Schürmann and J.-P. Jacquot. Plant Thioredoxin Systems Revisited. *Annu Rev Plant Physiol Plant Mol Biol*, 51:371–400, 2000.
- C. Schuster, Stefan und Hilgetag. On Elementary Flux Modes in biochemical reaction systems at steady state. *J Biol Syst*, 2:165–182, 1994.

- R. Schuster and S. Schuster. Refined algorithm and computer program for calculating all non-negative fluxes admissible in steady states of biochemical reaction systems with or without some flux rates fixed. *Comput Appl Biosci*, 9(1):79–85, 1993.
- S. Schuster, D. A. Fell, T. Pfeiffer, T. Dandekar, and P. Bork. *Elementary Modes Analysis Illustrated with Human Red Cell Metabolism*. BioThermoKinetics in the Post Genomic Era. Chalmers, Göteborg, 1998.
- S. Schuster, T. Dandekar, and D. A. Fell. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol*, 17(2):53–60, 1999.
- S. Schuster, D. A. Fell, and T. Dandekar. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat Biotechnol*, 18(3):326–332, 2000.
- S. Schuster and D. Kenanov. Adenine and adenosine salvage pathways in erythrocytes and the role of S-adenosylhomocysteine hydrolase. A theoretical study using elementary flux modes. *FEBS J*, 272(20):5278–5290, 2005.
- S. Schuster, T. Pfeiffer, and D. A. Fell. Is maximization of molar yield in metabolic networks favoured by evolution? *J Theor Biol*, 252(3):497–504, 2008.
- J.-M. Schwartz, C. Gaugain, J. C. Nacher, A. de Daruvar, and M. Kanehisa. Observing metabolic functions at the genome scale. *Genome Biol*, 8(6):R123, 2007.
- D. Segrè, D. Vitkup, and G. M. Church. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci U S A*, 99(23):15112–15117, 2002.
- A. Seressiotis and J. A. Bailey. Mps: An algorithm and data base for metabolic pathway synthesis. *Biotechnol Lett*, 8:837–842, 1986.
- M. Serrano, A. W. Lin, M. E. McCurrach, D. Beach, and S. W. Lowe. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. *Cell*, 88(5):593–602, 1997.
- J. W. Shay and W. E. Wright. Senescence and immortalization: role of telomeres and telomerase. *Carcinogenesis*, 26(5):867–874, 2005.

- D. N. Shelton, E. Chang, P. S. Whittier, D. Choi, and W. D. Funk. Microarray analysis of replicative senescence. *Curr Biol*, 9(17):939–945, 1999.
- J. Shiloach and U. Rinas. *Glucose and acetate metabolism in E. coli - System level analysis and biotechnological applications in protein production processes*, pages 377–400. Springer-Verlag, Berlin, Heidelberg, New York, 2009.
- J. Shiloach and R. Fass. Growing *E. coli* to high cell density—a historical perspective on method development. *Biotechnol Adv*, 23(5):345–357, 2005.
- E. Sikora, T. Arendt, M. Bennett, and M. Narita. Impact of cellular senescence signature on ageing research. *Ageing Res Rev*, 10(1):146–152, 2011.
- J. R. Smith and R. G. Whitney. Intraclonal variation in proliferative potential of human diploid fibroblasts: stochastic mechanism for cellular aging. *Science*, 207(4426):82–84, 1980.
- W. W. Soon, M. Hariharan, and M. P. Snyder. High-throughput sequencing for biology and medicine. *Mol Syst Biol*, 9:640, 2013.
- P. D. Sozou and T. B. Kirkwood. A stochastic model of cell replicative senescence based on telomere shortening, oxidative stress, and somatic mutations in nuclear and mitochondrial DNA. *J Theor Biol*, 213(4):573–586, 2001.
- A. H. Stouthamer. A theoretical study on the amount of atp required for synthesis of microbial cell material. *Antonie Van Leeuwenhoek*, 39(3):545–65, 1973.
- S. H. Strogatz. *Nonlinear Dynamics and Chaos*. Westview Press, 2000.
- H. Takai, A. Smogorzewska, and T. de Lange. DNA damage foci at dysfunctional telomeres. *Curr Biol*, 13(17):1549–1556, 2003.
- Z. Tan. Intramitotic and intraclonal variation in proliferative potential of human diploid cells: explained by telomere shortening. *J Theor Biol*, 198(2):259–268, 1999.
- M. Terzer and J. Stelling. Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics*, 24(19):2229–2235, 2008.

- B. Teusink, J. Passarge, C. A. Reijenga, E. Esgalhado, C. C. van der Weijden, M. Schepper, M. C. Walsh, B. M. Bakker, K. van Dam, H. V. Westerhoff, and J. L. Snoep. Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. *Eur J Biochem*, 267(17):5313–5329, 2000.
- I. Thiele and B. Ø. Palsson. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc*, 5(1):93–121, 2010.
- I. Thiele, N. Swainston, R. M. T. Fleming, A. Hoppe, S. Sahoo, M. K. Aurich, H. Haraldsdottir, M. L. Mo, O. Rolfsson, M. D. Stobbe, S. G. Thorleifsson, R. Agren, C. Bölling, S. Bordel, A. K. Chavali, P. Dobson, W. B. Dunn, L. Endler, D. Hala, M. Hucka, D. Hull, D. Jameson, N. Jamshidi, J. J. Jonsson, N. Juty, S. Keating, I. Nookaew, N. L. Novère, N. Malys, A. Mazein, J. A. Papin, N. D. Price, E. Selkov, M. I. Sigurdsson, E. Simeonidis, N. Sonnenschein, K. Smallbone, A. Sorokin, J. H. G. M. van Beek, D. Weichart, I. Goryanin, J. Nielsen, H. V. Westerhoff, D. B. Kell, P. Mendes, and B. Ø. Palsson. A community-driven global reconstruction of human metabolism. *Nat Biotechnol*, 2013.
- O. Toussaint, E. E. Medrano, and T. von Zglinicki. Cellular and molecular mechanisms of stress-induced premature senescence (SIPS) of human diploid fibroblasts and melanocytes. *Exp Gerontol*, 35(8):927–945, 2000.
- O. Toussaint, J. Remacle, J.-F. Dierick, T. Pascal, C. Fripiat, V. Royer, J. P. Magalhães, S. Zdanov, and F. Chainiaux. Stress-induced premature senescence: from biomarkers to likelihood of in vivo occurrence. *Biogerontology*, 3(1-2):13–17, 2002.
- O. Toussaint, G. Weemaels, F. Debacq-Chainiaux, K. Scharffetter-Kochanek, and M. Wlaschek. Artefactual effects of oxygen on cell culture models of cellular senescence and stem cell biology. *J Cell Physiol*, 226(2):315–321, 2011.
- C. T. Trinh and F. Srienc. Metabolic engineering of *Escherichia coli* for efficient conversion of glycerol to ethanol. *Appl Environ Microbiol*, 75(21):6696–6705, 2009.
- C. T. Trinh, R. Carlson, A. Wlaschin, and F. Srienc. Design, construction and performance of the most efficient biomass producing *E. coli* bacterium. *Metab Eng*, 8(6):628–638, 2006.

- C. T. Trinh, P. Unrean, and F. Sreenc. Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. *Appl Environ Microbiol*, 74(12):3634–3643, 2008.
- J. J. Tyson and B. Novak. Temporal organization of the cell cycle. *Curr Biol*, 18(17):R759–R768, 2008.
- P. Unrean and F. Sreenc. Predicting the adaptive evolution of thermoanaerobacterium saccharolyticum. *J Biotechnol*, 158(4):259–266, 2012.
- P. Unrean, C. T. Trinh, and F. Sreenc. Rational design and construction of an efficient *E. coli* for production of diapolycopendioic acid. *Metab Eng*, 12(2):112–122, 2010.
- A. Varma, B. W. Boesch, and B. Ø. Palsson. Biochemical production capabilities of escherichia coli. *Biotechnol Bioeng*, 42(1):59–73, 1993.
- A. Varma and B. Ø. Palsson. Metabolic flux balancing: Basic concepts, scientific and practical use. *Nat Biotechnol*, 12:994–998, 1994.
- L. Venkatraman, H. Li, C. F. Dewey, J. K. White, S. S. Bhowmick, H. Yu, and L. Tucker-Kellogg. Steady states and dynamics of urokinase-mediated plasmin activation in silico and in vitro. *Biophysical journal*, 101(8):1825–34, 2011.
- J. Vijg. The role of DNA damage and repair in aging: new approaches to an old problem. *Mech Ageing Dev*, 129(7-8):498–502, 2008.
- C. Wagner. Nullspace approach to determine the elementary modes of chemical reaction systems. *J. Phys. Chem. B*, 108(7):2425–2431, 2004.
- C. Wagner and R. Urbanczik. The geometry of the flux cone of a metabolic network. *Biophys J*, 89(6):3837–3845, 2005.
- H. Waltenberger, C. Schneid, J. O. Grosch, A. Bareiss, and M. Mittag. Identification of target mRNAs for the clock-controlled RNA-binding protein Chlamy 1 from *Chlamydomonas reinhardtii*. *Mol Genet Genomics*, 265:180–188, 2001.
- Y. Wang, J. A. Eddy, and N. D. Price. Reconstruction of genome-scale metabolic models for 126 human tissues using mcadre. *BMC Syst Biol*, 6:153, 2012.

- J. D. Watson and F. H. Crick. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.
- M. R. Watson. Metabolic maps for the Apple II. *Biochem. Soc. Trans.*, 12:1093–1094, 1984.
- M. R. Watson. A discrete model of bacterial metabolism. *Comput Appl Biosci*, 2(1):23–27, 1986.
- W. Weckwerth. Unpredictability of metabolism—the key role of metabolomics science in combination with next-generation genome sequencing. *Anal Bioanal Chem*, 400(7):1967–1978, 2011.
- V. F. Wendisch, M. Bott, and B. J. Eikmanns. Metabolic engineering of *Escherichia coli* and *Corynebacterium glutamicum* for biotechnological production of organic acids and amino acids. *Curr Opin Microbiol*, 9(3):268–274, 2006.
- H. V. Westerhoff and B. Ø. Palsson. The evolution of molecular biology into systems biology. *Nat Biotechnol*, 22(10):1249–1252, 2004.
- A. P. Wlaschin, C. T. Trinh, R. Carlson, and F. Sreenc. The fractional contributions of elementary modes to the metabolism of *Escherichia coli* and their estimation from reaction entropies. *Metab Eng*, 8:338–352, 2006.
- A. J. Wolfe. The acetate switch. *Microbiol. Mol. Biol. Rev.*, 69(1):12–50, 2005.
- M. Yeung, I. Thiele, and B. Ø. Palsson. Estimation of the number of extreme pathways for metabolic networks. *BMC Bioinformatics*, 8(1):363, 2007.
- W. Zhao, Z. X. Lin, and Z. Q. Zhang. Cisplatin-induced premature senescence with concomitant reduction of gap junctions in human fibroblasts. *Cell Res*, 14(1):60–66, 2004.
- H. Zur, E. Rupp, and T. Shlomi. iMAT: an integrative metabolic analysis tool. *Bioinformatics*, 26(24):3140–3142, 2010.

A. Supplementary material

Effect of substrate competition in kinetic models of metabolic networks.

Schäuble S, Stavrum A K, Puntervoll P, Schuster S, Heiland I, *FEBS Lett*, 587(17): 2818–2824, 2013.

SUPPLEMENTARY MATERIAL

Supplementary Material

Sascha Schäuble^{1*}, Anne Kristin Stavrum^{2*}, Pål Puntervoll³, Stefan Schuster⁴, and Ines Heiland^{5†}

¹Theoretical Systems Biology Group, Friedrich-Schiller-University Jena, Germany

²Department of Informatics, University of Bergen, Bergen, Norway

³Computational Biology Unit, Uni Computing, Bergen, Norway

⁴Department of Bioinformatics, Friedrich-Schiller-University Jena, Germany

⁵Department of arctic and marine biology, University of Tromsø, Norway

*These authors have contributed equally to this work

†corresponding author: *ines.heiland@uit.no*

June 17, 2013

Contents

1	Full derivation of kinetic rate law for monomolecular reactions	2
2	Proof of substitution rule	5
3	Kinetic rate laws for bi-molecular reactions including substrate competition	7
3.1	Ordered Bi-Bi system	7
3.2	Theorell-Chance Bi Bi system	8
3.3	Ping Pong Bi Bi system	8
4	Detailed model descriptions	9
4.1	Case A - Competition between different pathways	9
4.1.1	Irreversible kinetics	9
4.1.2	Reversible kinetics	10
4.2	Case B - Competition in a branched pathway	12
4.2.1	Irreversible kinetics	12
4.2.2	Reversible kinetics	13
4.3	Case C - Multistep reactions with alternating enzymes	14
4.3.1	Irreversible kinetics	14
4.3.2	Reversible kinetics	15

1 Full derivation of kinetic rate law for monomolecular reactions

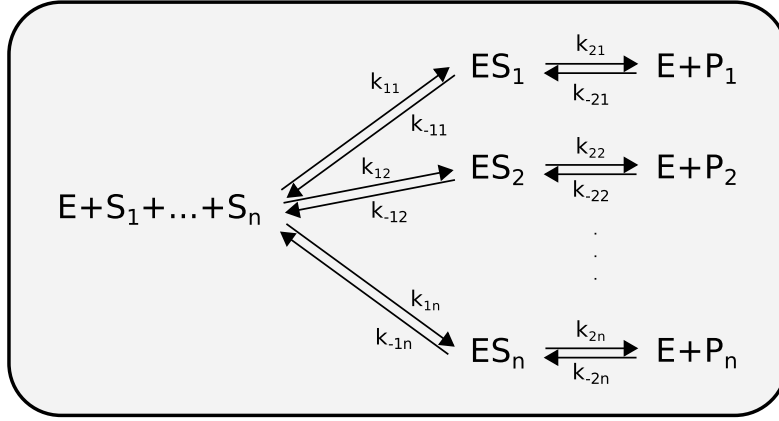


Figure 1: Reaction diagram for two substrates competing for the active site of an enzyme.

List of differential equations for the metabolic compounds in Figure 1:

$$\begin{aligned}
 \dot{S}_1 &= -k_{11}E \cdot S_1 + k_{-11}ES_1 \\
 \dot{S}_2 &= -k_{12}E \cdot S_2 + k_{-12}ES_2 \\
 &\vdots \\
 \dot{S}_n &= -k_{1n}E \cdot S_n + k_{-1n}ES_n \\
 \dot{E} &= -(k_{11}E \cdot S_1 + k_{12}E \cdot S_2 + \dots + k_{1n}E \cdot S_n) \\
 &\quad + (k_{-11}ES_1 + k_{-12}ES_2 + \dots + k_{-1n}ES_n) \\
 &\quad + (k_{21}ES_1 + k_{22}ES_2 + \dots + k_{2n}ES_n) \\
 &\quad - (k_{-21}E \cdot P_1 + k_{-22}E \cdot P_2 + \dots + k_{-2n}E \cdot P_n) \\
 \dot{ES}_1 &= k_{11}E \cdot S_1 - k_{-11}ES_1 - k_{21}ES_1 + k_{-21}E \cdot P_1 \\
 \dot{ES}_2 &= k_{12}E \cdot S_2 - k_{-12}ES_2 - k_{22}ES_2 + k_{-22}E \cdot P_2 \\
 &\vdots \\
 \dot{ES}_n &= k_{1n}E \cdot S_n - k_{-1n}ES_n - k_{2n}ES_n + k_{-2n}E \cdot P_n \\
 \dot{P}_1 &= k_{21}ES_1 - k_{-21}E \cdot P_1 \\
 \dot{P}_2 &= k_{22}ES_2 - k_{-22}E \cdot P_2 \\
 &\vdots \\
 \dot{P}_n &= k_{2n}ES_n - k_{-2n}E \cdot P_n
 \end{aligned}$$

The generalised rate law for the production of P_1 with n competing substrates for the binding site of the enzyme E and the substitutions $V_f = k_{21}E_t$, $V_r = k_{-11}E_t$ reads:

$$\dot{P}_1 = v_{0_n} = \frac{V_f \frac{S_1}{K_m^{S_1}} - V_r \frac{P_1}{K_m^{P_1}}}{\sum_{i=1}^n \left(\frac{S_i}{K_m^{S_i}} + \frac{P_i}{K_m^{P_i}} \right) + 1} \quad (1)$$

The proof is given via mathematical induction.

Base case: $n = 2$

Conservation relation:

$$E_t = E + ES_1 + ES_2 \quad (2)$$

Deduction

$$\underline{ES_1 = 0}$$

$$\begin{aligned} k_{11}E \cdot S_1 + k_{-21}E \cdot P_1 - (k_{-11} + k_{21})ES_1 &= 0 \\ E &= \frac{(k_{-11} + k_{21})ES_1}{k_{11}S_1 + k_{-21}P_1} \end{aligned} \quad (3)$$

$$\underline{ES_2 = 0}$$

$$E = \frac{(k_{-12} + k_{22})ES_2}{k_{12}S_2 + k_{-22}P_2} \quad (4)$$

$$\underline{(3) = (4)}$$

$$ES_2 = ES_1 \frac{(k_{-11} + k_{21})(k_{12}S_2 + k_{-22}P_2)}{(k_{11}S_1 + k_{-21}P_1)(k_{-12} + k_{22})} \quad (5)$$

$$\underline{(5) \rightarrow (2)}$$

$$\begin{aligned} ES_2 &= E_t - ES_1 - E \\ \underbrace{ES_1 \frac{(k_{-11} + k_{21})(k_{12}S_2 + k_{-22}P_2)}{(k_{11}S_1 + k_{-21}P_1)(k_{-12} + k_{22})}}_{ES_2} &= E_t - ES_1 - \underbrace{\frac{(k_{-11} + k_{21})ES_1}{k_{11}S_1 + k_{-21}P_1}}_E \\ E_t &= ES_1 \left(1 + \frac{(k_{-11} + k_{21})(k_{12}S_2 + k_{-22}P_2)}{(k_{11}S_1 + k_{-21}P_1)(k_{-12} + k_{22})} + \frac{(k_{-11} + k_{21})}{k_{11}S_1 + k_{-21}P_1} \right) \\ ES_1 &= \frac{E_t \frac{k_{11}S_1 + k_{-21}P_1}{k_{-11} + k_{21}}}{1 + \frac{k_{11}S_1 + k_{-21}P_1}{k_{-11} + k_{21}} + \frac{k_{12}S_2 + k_{-22}P_2}{k_{-12} + k_{22}}} \end{aligned} \quad (6)$$

$$\underline{(6) \rightarrow (3)}$$

$$E = \frac{E_t}{1 + \frac{k_{11}S_1 + k_{-21}P_1}{k_{-11} + k_{21}} + \frac{k_{12}S_2 + k_{-22}P_2}{k_{-12} + k_{22}}} \quad (7)$$

$$\underline{(6), (7) \rightarrow \dot{P}_1}$$

$$\begin{aligned} \dot{P}_1 &= k_2^1 ES_1 - k_{-2}^1 E \cdot P_1 \\ &= \frac{k_{21}E_t \frac{k_{11}S_1}{k_{-11} + k_{21}} + k_{21}E_t \frac{k_{-21}P_1}{k_{-11} + k_{21}} - k_{-21}E_t P_1}{\frac{S_2}{K_m^{S_2}} + \frac{P_2}{K_m^{P_2}} + \frac{S_1}{K_m^{S_1}} + \frac{P_1}{K_m^{P_1}} + 1} \end{aligned}$$

$$\begin{aligned}
\dot{P}_1 = v_0 &= \frac{k_{21}E_t \frac{S_1}{K_m^{S_1}} - k_{-11}E_t \frac{P_1}{K_m^{P_1}}}{\frac{S_1}{K_m^{S_1}} + \frac{P_1}{K_m^{P_1}} + \frac{S_2}{K_m^{S_2}} + \frac{P_2}{K_m^{P_2}} + 1} \\
&= \frac{V_f \frac{S_1}{K_m^{S_1}} - V_r \frac{P_1}{K_m^{P_1}}}{\sum_{i=1}^2 \left(\frac{S_i}{K_m^{S_i}} + \frac{P_i}{K_m^{P_i}} \right) + 1} \tag{8}
\end{aligned}$$

→ Eq. (1) holds for some n according to induction hypothesis → Show that it holds for $n + 1$

Inductive step

(1) holds for some n according to induction hypothesis → Show that it holds for $n + 1$

Steady state assumption yields:

$$\dot{E}S_1 = \dot{E}S_2 = \dots = \dot{E}S_n = \dot{E}S_{n+1} = 0$$

Conservation relation yields:

$$ES_2 = E_t - E - ES_1 - ES_3 - ES_4 - \dots - ES_n - ES_{n+1} \tag{9}$$

$$\begin{aligned}
&\underbrace{ES_1 \frac{(k_{-11} + k_{21})(k_{12}S_2 + k_{-22}P_2)}{(k_{11}S_1 + k_{-21}P_1)(k_{-12} + k_{22})}}_{ES_2} = \\
&\quad E_t - \underbrace{\frac{(k_{-11} + k_{21})ES_1}{k_{11}S_1 + k_{-21}P_1}}_E - ES_1 - \\
&\quad \underbrace{ES_1 \frac{(k_{-11} + k_{21})(k_{13}S_3 + k_{-23}P_3)}{(k_{11}S_1 + k_{-21}P_1)(k_{-13} + k_{23})}}_{ES_3} - \\
&\quad \underbrace{ES_1 \frac{(k_{-11} + k_{21})(k_{14}S_4 + k_{-14}P_4)}{(k_{11}S_1 + k_{-21}P_1)(k_{-14} + k_{24})}}_{ES_4} - \\
&\quad \dots - \\
&\quad \underbrace{ES_1 \frac{(k_{-11} + k_{21})(k_{1n}S_n + k_{-1n}P_n)}{(k_{11}S_1 + k_{-21}P_1)(k_{-1n} + k_{2n})}}_{ES_n} - \\
&\quad \underbrace{ES_1 \frac{(k_{-11} + k_{21})(k_{1n+1}S_{n+1} + k_{-1n+1}P_{n+1})}{(k_{11}S_1 + k_{-21}P_1)(k_{-1n+1} + k_{2n+1})}}_{ES_{n+1}}
\end{aligned}$$

$$E_t = ES_1 \left(1 + \frac{(k_{-11} + k_{21})}{k_{11}S_1 + k_{-21}P_1} + \frac{(k_{-11} + k_{21})(k_{12}S_2 + k_{-22}P_2)}{(k_{11}S_1 + k_{-21}P_1)(k_{-12} + k_{22})} + \frac{(k_{-11} + k_{21})(k_{13}S_3 + k_{-23}P_3)}{(k_{11}S_1 + k_{-21}P_1)(k_{-13} + k_{23})} + \dots + \frac{(k_{-11} + k_{21})(k_{1n}S_n + k_{-2n}P_n)}{(k_{11}S_1 + k_{-21}P_1)(k_{-1n} + k_{2n})} + \frac{(k_{-11} + k_{21})(k_{1n+1}S_{n+1} + k_{-2n+1}P_{n+1})}{(k_{11}S_1 + k_{-21}P_1)(k_{-1n+1} + k_{2n+1})} \right)$$

$$ES_1 = \frac{E_t \frac{k_{11}S_1 + k_{-21}P_1}{k_{-11} + k_{21}}}{1 + \frac{k_{11}S_1 + k_{-21}P_1}{k_{-11} + k_{21}} + \frac{k_{12}S_2 + k_{-22}P_2}{k_{-12} + k_{22}} + \dots + \frac{k_{1n}S_n + k_{-2n}P_n}{k_{-1n} + k_{2n}} + \frac{k_{1n+1}S_{n+1} + k_{-2n+1}P_{n+1}}{k_{-1n+1} + k_{2n+1}}} \quad (10)$$

(10) \rightarrow (9)

$$E = \frac{E_t}{1 + \frac{k_{11}S_1 + k_{-21}P_1}{k_{-11} + k_{21}} + \frac{k_{12}S_2 + k_{-22}P_2}{k_{-12} + k_{22}} + \dots + \frac{k_{1n}S_n + k_{-2n}P_n}{k_{-1n} + k_{2n}} + \frac{k_{1n+1}S_{n+1} + k_{-2n+1}P_{n+1}}{k_{-1n+1} + k_{2n+1}}} \quad (11)$$

(10), (11) $\rightarrow \dot{P}_1$

$$\dot{P}_1 = k_2^1 ES_1 - k_{-2}^1 E \cdot P_1$$

$$\begin{aligned} \dot{P}_1 &= \frac{k_{21}E_t \frac{S_1}{K_m^{S_1}} - k_{-11}E_t \frac{P_1}{K_m^{P_1}}}{\frac{S_1}{K_m^{S_1}} + \frac{S_2}{K_m^{S_2}} + \frac{P_2}{K_m^{P_2}} + \frac{P_1}{K_m^{P_1}} + \dots + \frac{S_n}{K_m^{S_n}} + \frac{P_n}{K_m^{P_n}} + \frac{S_{n+1}}{K_m^{S_{n+1}}} + \frac{P_{n+1}}{K_m^{P_{n+1}}} + 1} \\ &= \frac{k_{21}E_t \frac{S_1}{K_m^{S_1}} - k_{-11}E_t \frac{P_1}{K_m^{P_1}}}{\sum_{i=1}^n \left(\frac{S_i}{K_m^{S_i}} + \frac{P_i}{K_m^{P_i}} \right) + \frac{S_{n+1}}{K_m^{S_{n+1}}} + \frac{P_{n+1}}{K_m^{P_{n+1}}} + 1} \\ \dot{P}_1 &= \frac{k_{21}E_t \frac{S_1}{K_m^{S_1}} - k_{-11}E_t \frac{P_1}{K_m^{P_1}}}{\sum_{i=1}^{n+1} \left(\frac{S_i}{K_m^{S_i}} + \frac{P_i}{K_m^{P_i}} \right) + 1} \end{aligned}$$

□

2 Proof of substitution rule

Classic reversible Michaelis-Menten kinetics:

$$v = \frac{V_f \frac{S_1}{K_m^{S_1}} - V_r \frac{P_1}{K_m^{P_1}}}{\frac{S_1}{K_m^{S_1}} + \frac{P_1}{K_m^{P_1}} + 1} \quad (12)$$

Now, by adding the competition of one competing substrate and one product as competitive inhibition by changing the K_m according to the rules described in [1], Eq. (12) becomes:

$$v = \frac{V_f \frac{S_1}{K_m^{S_1} (1 + \frac{S_2}{K_m^{S_2}} + \frac{P_2}{K_m^{P_2}})} - V_r \frac{P_1}{K_m^{P_1} (1 + \frac{S_2}{K_m^{S_2}} + \frac{P_2}{K_m^{P_2}})}}{\frac{S_1}{K_m^{S_1} (1 + \frac{S_2}{K_m^{S_2}} + \frac{P_2}{K_m^{P_2}})} + \frac{P_1}{K_m^{P_1} (1 + \frac{S_2}{K_m^{S_2}} + \frac{P_2}{K_m^{P_2}})} + 1} \quad (13)$$

By multiplying both, numerator and denominator with the Term $(1 + \frac{S_2}{K_m^{S_2}} + \frac{P_2}{K_m^{P_2}})$ Eq. (13) becomes:

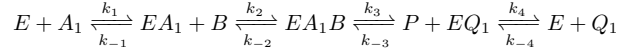
$$\begin{aligned} v &= \frac{V_f \frac{S_1}{K_m^{S_1}} - V_r \frac{P_1}{K_m^{P_1}}}{\frac{S_1}{K_m^{S_1}} + \frac{P_1}{K_m^{P_1}} + 1 + \frac{S_2}{K_m^{S_2}} + \frac{P_2}{K_m^{P_2}}} \\ &= \frac{V_f \frac{S_1}{K_m^{S_1}} - V_r \frac{P_1}{K_m^{P_1}}}{\frac{S_1}{K_m^{S_1}} + \frac{S_2}{K_m^{S_2}} + \frac{P_1}{K_m^{P_1}} + \frac{P_2}{K_m^{P_2}} + 1} \\ &= \frac{V_f \frac{S_1}{K_m^{S_1}} - V_r \frac{P_1}{K_m^{P_1}}}{\sum_{i=1}^2 \left(\frac{S_i}{K_m^{S_i}} + \frac{P_i}{K_m^{P_i}} \right) + 1} \end{aligned} \quad (14)$$

□

3 Kinetic rate laws for bi-molecular reactions including substrate competition

3.1 Ordered Bi-Bi system

Scheme



Following the King-Altman method of deriving a steady-state velocity equation and making use of the Cleland nomenclature as described in Segel (1975), the resulting rate equation without substrate inhibition in terms of kinetic constants reads:

$$v = \frac{V_f V_r \left([A][B] - \frac{[P][Q]}{K_{eq}} \right)}{V_r K_{ia} K_{m_B} + V_r K_{m_B} [A] + V_r K_{m_A} [B] + \frac{V_f K_{m_Q} [P]}{K_{eq}} + \frac{V_f K_{m_P} [Q]}{K_{eq}} + V_r [A][B] + \frac{V_f K_{m_Q} [A][P]}{K_{eq} K_{ia}} + \frac{V_f [P][Q]}{K_{eq}} + \frac{V_r K_{m_A} [B][Q]}{K_{iq}} + \frac{V_r [A][B][P]}{K_{ip}} + \frac{V_f [B][P][Q]}{K_{ib} K_{eq}}} \quad (15)$$

where

$$\begin{aligned} V_f &= \frac{k_3 k_4}{(k_3 + k_4)} & V_r &= \frac{k_{-1} k_{-2}}{(k_{-1} + k_{-2})} & K_{eq} &= \frac{k_1 k_2 k_3 k_4}{k_{-1} k_{-2} k_{-3} k_{-4}} \\ K_{m_A} &= \frac{k_3 k_4}{k_1 (k_3 + k_4)} & K_{m_B} &= \frac{k_4 (k_{-2} + k_3)}{k_2 (k_3 + k_4)} & K_{ia} &= \frac{k_{-1}}{k_1} \\ K_{ib} &= \frac{k_{-1} + k_{-2}}{k_2} & K_{m_P} &= \frac{k_{-1} (k_{-2} + k_3)}{k_{-3} (k_{-1} + k_{-2})} & K_{m_Q} &= \frac{k_{-1} k_{-2}}{k_{-4} (k_{-1} + k_{-2})} \\ K_{ip} &= \frac{k_3 + k_4}{k_{-3}} & K_{iq} &= \frac{k_4}{k_{-4}} \end{aligned}$$

Now, we introduce species A_2 and Q_2 , which compete with A and Q , respectively, for the single binding site of the free enzyme. It is reasonable to assume that only one species at a time can bind to the catalytic site.

The competition follows the mechanism of a competitive inhibition and thus, only affects the respective kinetic's k_m value, but not the maximal velocities in either direction. Thus, the terms K_{m_A} and K_{m_Q} in Eq. (15) need to be extended by the term $(1 + \frac{[A_2]}{K_{m_{A_2}}} + \frac{[Q_2]}{K_{m_{Q_2}}})$:

$$v = \frac{V_f V_r \left([A][B] - \frac{[P][Q]}{K_{eq}} \right)}{V_r K_{ia} K_{m_B} + V_r K_{m_B} [A] + V_r K_{m_A} \left(1 + \frac{[A_2]}{K_{m_{A_2}}} + \frac{[Q_2]}{K_{m_{Q_2}}} \right) [B] + \frac{V_f K_{m_Q} \left(1 + \frac{[A_2]}{K_{m_{A_2}}} + \frac{[Q_2]}{K_{m_{Q_2}}} \right) [P]}{K_{eq}} + \frac{V_f K_{m_P} [Q]}{K_{eq}} + V_r [A][B] + \frac{V_f K_{m_Q} \left(1 + \frac{[A_2]}{K_{m_{A_2}}} + \frac{[Q_2]}{K_{m_{Q_2}}} \right) [A][P]}{K_{eq} K_{ia}} + \frac{V_f [P][Q]}{K_{eq}} + \frac{V_r K_{m_A} \left(1 + \frac{[A_2]}{K_{m_{A_2}}} + \frac{[Q_2]}{K_{m_{Q_2}}} \right) [B][Q]}{K_{iq}} + \frac{V_r [A][B][P]}{K_{ip}} + \frac{V_f [B][P][Q]}{K_{ib} K_{eq}}} \quad (16)$$

The k_m values for B and P species remain unaffected, as the competition occurs only at the initial step of the reaction in either direction. Consequently, only the species A and Q face competitors for the binding site of the free enzyme.

3.2 Theorell-Chance Bi Bi system

Theorell and Chance proposed 1951 an Ordered Bi Bi mechanism without a central complex which contains both substrates. Using the King-Altman/Cleland scheme leads to a velocity equation equal to (15) without terms of combined ABP and BPQ species. Thus, from Eq. (16) we can directly derive the velocity equation, where two additional species A_2 and Q_2 compete for the binding site of the free enzyme species:

$$v = \frac{V_f V_r \left([A][B] - \frac{[P][Q]}{K_{eq}} \right)}{V_r K_{ia} K_{m_B} + V_r K_{m_B} [A] + V_r K_{m_A} \left(1 + \frac{[A_2]}{K_{m_{A_2}}} + \frac{[Q_2]}{K_{m_{Q_2}}} \right) [B] + \frac{V_f K_{m_Q} \left(1 + \frac{[A_2]}{K_{m_{A_2}}} + \frac{[Q_2]}{K_{m_{Q_2}}} \right) [P]}{K_{eq}} + \frac{V_f K_{m_P} [Q]}{K_{eq}} + V_r [A][B] + \frac{V_f K_{m_Q} \left(1 + \frac{[A_2]}{K_{m_{A_2}}} + \frac{[Q_2]}{K_{m_{Q_2}}} \right) [A][P]}{K_{eq} K_{ia}} + \frac{V_f [P][Q]}{K_{eq}} + \frac{V_r K_{m_A} \left(1 + \frac{[A_2]}{K_{m_{A_2}}} + \frac{[Q_2]}{K_{m_{Q_2}}} \right) [B][Q]}{K_{iq}}} \quad (17)$$

3.3 Ping Pong Bi Bi system

The Ping Pong Bi Bi mechanism distinguishes from an Ordered Bi Bi as here, a product is released between the addition of two substrates. Again, by applying the King-Altman method a velocity equation can be derived:

$$v = \frac{V_f V_r \left([A][B] - \frac{[P][Q]}{K_{eq}} \right)}{V_r K_{m_B} [A] + V_r K_{m_A} [B] + \frac{V_f K_{m_Q} [P]}{K_{eq}} + \frac{V_f K_{m_P} [Q]}{K_{eq}} + V_r [A][B] + \frac{V_f K_{m_Q} [A][P]}{K_{eq} K_{ia}} + \frac{V_f [P][Q]}{K_{eq}} + \frac{V_r K_{m_A} [B][Q]}{K_{iq}}} \quad (18)$$

The definition of the kinetic constants is the same as for the Ordered Bi Bi system, except for K_{ib} and K_{ip} :

$$K_{ib} = \frac{k_{-3}}{k_3} \qquad K_{ip} = \frac{k_2}{k_{-2}}$$

Following the argument given above, only the K_m values are affected, when additional A_2 and Q_2 species compete with A and Q , respectively for the binding site of the free enzyme. Consequently, Eq. (18) can be modified in the same manner:

$$v = \frac{V_f V_r \left([A][B] - \frac{[P][Q]}{K_{eq}} \right)}{V_r K_{m_B} [A] + V_r K_{m_A} \left(1 + \frac{[A_2]}{K_{m_{A_2}}} + \frac{[Q_2]}{K_{m_{Q_2}}} \right) [B] + \frac{V_f K_{m_Q} \left(1 + \frac{[A_2]}{K_{m_{A_2}}} + \frac{[Q_2]}{K_{m_{Q_2}}} \right) [P]}{K_{eq}} + \frac{V_f K_{m_P} [Q]}{K_{eq}} + V_r [A][B] + \frac{V_f K_{m_Q} \left(1 + \frac{[A_2]}{K_{m_{A_2}}} + \frac{[Q_2]}{K_{m_{Q_2}}} \right) [A][P]}{K_{eq} K_{ia}} + \frac{V_f [P][Q]}{K_{eq}} + \frac{V_r K_{m_A} \left(1 + \frac{[A_2]}{K_{m_{A_2}}} + \frac{[Q_2]}{K_{m_{Q_2}}} \right) [B][Q]}{K_{iq}}} \quad (19)$$

Note that Eq. (17) and (19) differ only by the term $V_r K_{ia} K_{m_B}$, which is missing in the denominator of the Ping Pong Bi Bi mechanism. Hence, if the backward maximum velocity V_r , the Michaelis-Menten constant K_{m_B} or the dissociation constant K_{ia} are sufficiently small (close to zero), the Ping-Pong Bi Bi system is virtually not distinguishable from the Theorell-Chance mechanism.

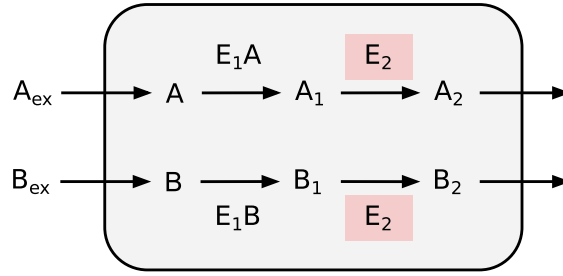
4 Detailed model descriptions

For the sake of simplicity all Michaelis-Menten constants K_{m_x} are the same for A and B species, respectively.

4.1 Case A - Competition between different pathways

$$A(0) = A_1(0) = A_2(0) = B(0) = B_1(0) = B_2(0) = 0$$

4.1.1 Irreversible kinetics



Parameter	V_f	k_A	k_B	K_{m_A}	K_{m_B}	A_{ex}	B_{ex}
Value	1	1	1	0.02	0.02	0.05	0.05

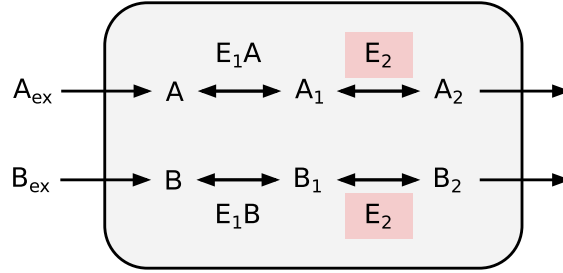
Ordinary differential equation system neglecting substrate competition

$$\begin{aligned}
 \frac{dA(t)}{d(t)} &= k_A A_{ex} - \frac{V_f A(t)}{K_{m_A} + A(t)} \\
 \frac{dA_1(t)}{d(t)} &= \frac{V_f A(t)}{K_{m_A} + A(t)} - \frac{V_f A_1(t)}{K_{m_A} + A_1(t)} \\
 \frac{dA_2(t)}{d(t)} &= \frac{V_f A_1(t)}{K_{m_A} + A_1(t)} - \frac{V_f A_2(t)}{K_{m_A} + A_2(t)} \\
 \frac{dB(t)}{d(t)} &= k_B B_{ex} - \frac{V_f B(t)}{K_{m_B} + B(t)} \\
 \frac{dB_1(t)}{d(t)} &= \frac{V_f B(t)}{K_{m_B} + B(t)} - \frac{V_f B_1(t)}{K_{m_B} + B_1(t)} \\
 \frac{dB_2(t)}{d(t)} &= \frac{V_f B_1(t)}{K_{m_B} + B_1(t)} - \frac{V_f B_2(t)}{K_{m_B} + B_2(t)}
 \end{aligned}$$

Ordinary differential equation system with substrate competition

$$\begin{aligned}
\frac{dA(t)}{dt} &= k_A A_{ex} - \frac{V_f A(t)}{K_{m_A} + A(t)} \\
\frac{dA_1(t)}{dt} &= \frac{V_f A(t)}{K_{m_A} + A(t)} - \frac{V_f A_1(t)}{K_{m_A} \left(1 + \frac{B_1(t)}{K_{m_B}}\right) + A_1(t)} \\
\frac{dA_2(t)}{dt} &= \frac{V_f A_1(t)}{K_{m_A} \left(1 + \frac{B_1(t)}{K_{m_B}}\right) + A_1(t)} - \frac{V_f A_2(t)}{K_{m_A} + A_2(t)} \\
\frac{dB(t)}{dt} &= k_B B_{ex} - \frac{V_f B(t)}{K_{m_B} + B(t)} \\
\frac{dB_1(t)}{dt} &= \frac{V_f B(t)}{K_{m_B} + B(t)} - \frac{V_f B_1(t)}{\left(K_{m_B} \left(1 + \frac{A_1(t)}{K_{m_A}}\right) + B_1(t)\right)} \\
\frac{dB_2(t)}{dt} &= \frac{V_f B_1(t)}{\left(K_{m_B} \left(1 + \frac{A_1(t)}{K_{m_A}}\right) + B_1(t)\right)} - \frac{V_f B_2(t)}{K_{m_B} + B_2(t)}
\end{aligned}$$

4.1.2 Reversible kinetics



Parameter	V_f	V_r	k_A	k_B	K_{m_A}	K_{m_B}	A_{ex}	B_{ex}
Value	1	1	1	1	0.02	0.02	0.05	0.05

Ordinary differential equation system neglecting substrate competition

$$\begin{aligned}
\frac{dA(t)}{d(t)} &= k_A A_{ex} - \frac{\frac{V_f A(t)}{K_{m_A}} - \frac{V_r A_1(t)}{K_{m_A}}}{1 + \frac{A(t)}{K_{m_A}} + \frac{A_1(t)}{K_{m_A}}} \\
\frac{dA_1(t)}{d(t)} &= \frac{\frac{V_f A(t)}{K_{m_A}} - \frac{V_r A_1(t)}{K_{m_A}}}{1 + \frac{A(t)}{K_{m_A}} + \frac{A_1(t)}{K_{m_A}}} - \frac{\frac{V_f A_1(t)}{K_{m_A}} - \frac{V_r A_2(t)}{K_{m_A}}}{1 + \frac{A_1(t)}{K_{m_A}} + \frac{A_2(t)}{K_{m_A}}} \\
\frac{dA_2(t)}{d(t)} &= \frac{\frac{V_f A_1(t)}{K_{m_A}} - \frac{V_r A_2(t)}{K_{m_A}}}{1 + \frac{A_1(t)}{K_{m_A}} + \frac{A_2(t)}{K_{m_A}}} - \frac{V_f A_2(t)}{K_{m_A} + A_2(t)} \\
\frac{dB(t)}{d(t)} &= k_B B_{ex} - \frac{\frac{V_f B(t)}{K_{m_B}} - \frac{V_r B_1(t)}{K_{m_B}}}{1 + \frac{B(t)}{K_{m_B}} + \frac{B_1(t)}{K_{m_B}}} \\
\frac{dB_1(t)}{d(t)} &= \frac{\frac{V_f B(t)}{K_{m_B}} - \frac{V_r B_1(t)}{K_{m_B}}}{1 + \frac{B(t)}{K_{m_B}} + \frac{B_1(t)}{K_{m_B}}} - \frac{\frac{V_f B_1(t)}{K_{m_B}} - \frac{V_r B_2(t)}{K_{m_B}}}{1 + \frac{B_1(t)}{K_{m_B}} + \frac{B_2(t)}{K_{m_B}}} \\
\frac{dB_2(t)}{d(t)} &= \frac{\frac{V_f B_1(t)}{K_{m_B}} - \frac{V_r B_2(t)}{K_{m_B}}}{1 + \frac{B_1(t)}{K_{m_B}} + \frac{B_2(t)}{K_{m_B}}} - \frac{V_f B_2(t)}{K_{m_B} + B_2(t)}
\end{aligned}$$

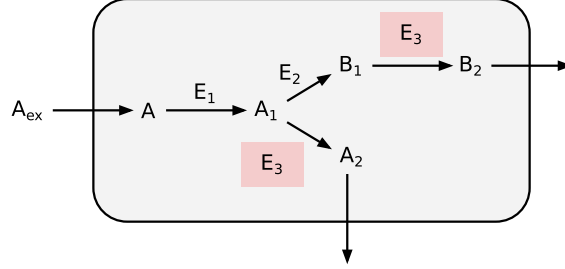
Ordinary differential equation system with substrate competition

$$\begin{aligned}
\frac{dA(t)}{d(t)} &= k_A A_{ex} - \frac{\frac{V_f A(t)}{K_{m_A}} - \frac{V_r A_1(t)}{K_{m_A}}}{1 + \frac{A(t)}{K_{m_A}} + \frac{A_1(t)}{K_{m_A}}} \\
\frac{dA_1(t)}{d(t)} &= \frac{\frac{V_f A(t)}{K_{m_A}} - \frac{V_r A_1(t)}{K_{m_A}}}{1 + \frac{A(t)}{K_{m_A}} + \frac{A_1(t)}{K_{m_A}}} - \frac{\frac{V_f A_1(t)}{K_{m_A}} - \frac{V_r A_2(t)}{K_{m_A}}}{1 + \frac{A_1(t)}{K_{m_A}} + \frac{A_2(t)}{K_{m_A}} + \frac{B_1(t)}{K_{m_B}} + \frac{B_2(t)}{K_{m_B}}} \\
\frac{dA_2(t)}{d(t)} &= \frac{\frac{V_f A_1(t)}{K_{m_A}} - \frac{V_r A_2(t)}{K_{m_A}}}{1 + \frac{A_1(t)}{K_{m_A}} + \frac{A_2(t)}{K_{m_A}} + \frac{B_1(t)}{K_{m_B}} + \frac{B_2(t)}{K_{m_B}}} - \frac{V_f A_2(t)}{K_{m_A} + A_2(t)} \\
\frac{dB(t)}{d(t)} &= k_B B_{ex} - \frac{\frac{V_f B(t)}{K_{m_B}} - \frac{V_r B_1(t)}{K_{m_B}}}{1 + \frac{B(t)}{K_{m_B}} + \frac{B_1(t)}{K_{m_B}}} \\
\frac{dB_1(t)}{d(t)} &= \frac{\frac{V_f B(t)}{K_{m_B}} - \frac{V_r B_1(t)}{K_{m_B}}}{1 + \frac{B(t)}{K_{m_B}} + \frac{B_1(t)}{K_{m_B}}} - \frac{\frac{V_f B_1(t)}{K_{m_B}} - \frac{V_r B_2(t)}{K_{m_B}}}{1 + \frac{A_1(t)}{K_{m_A}} + \frac{A_2(t)}{K_{m_A}} + \frac{B_1(t)}{K_{m_B}} + \frac{B_2(t)}{K_{m_B}}} \\
\frac{dB_2(t)}{d(t)} &= \frac{\frac{V_f B_1(t)}{K_{m_B}} - \frac{V_r B_2(t)}{K_{m_B}}}{1 + \frac{A_1(t)}{K_{m_A}} + \frac{A_2(t)}{K_{m_A}} + \frac{B_1(t)}{K_{m_B}} + \frac{B_2(t)}{K_{m_B}}} - \frac{V_f B_2(t)}{K_{m_B} + B_2(t)}
\end{aligned}$$

4.2 Case B - Competition in a branched pathway

$$A(0) = A_1(0) = A_2(0) = B_1(0) = B_2(0) = 0$$

4.2.1 Irreversible kinetics



Parameter	V_f	k_A	k_B	K_{m_A}	K_{m_B}	A_{ex}
Value	1	1	1	0.02	0.02	0.05

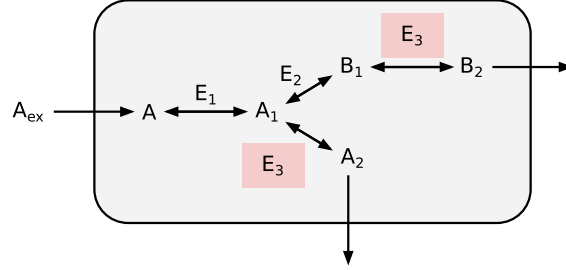
Ordinary differential equation system neglecting substrate competition

$$\begin{aligned}
 \frac{dA(t)}{dt} &= k_A A_{ex} - \frac{V_f A(t)}{K_{m_A} + A(t)} \\
 \frac{dA_1(t)}{dt} &= \frac{V_f A(t)}{K_{m_A} + A(t)} - 2 \frac{V_f A_1(t)}{K_{m_A} + A_1(t)} \\
 \frac{dA_2(t)}{dt} &= \frac{V_f A_1(t)}{K_{m_A} + A_1(t)} - \frac{V_f A_2(t)}{K_{m_A} + A_2(t)} \\
 \frac{dB_1(t)}{dt} &= \frac{V_f A_1(t)}{K_{m_A} + A_1(t)} - \frac{V_f B_1(t)}{K_{m_B} + B_1(t)} \\
 \frac{dB_2(t)}{dt} &= \frac{V_f B_1(t)}{K_{m_B} + B_1(t)} - \frac{V_f B_2(t)}{K_{m_B} + B_2(t)}
 \end{aligned}$$

Ordinary differential equation system with substrate competition

$$\begin{aligned}
 \frac{dA(t)}{dt} &= k_A A_{ex} - \frac{V_f A(t)}{K_{m_A} + A(t)} \\
 \frac{dA_1(t)}{dt} &= \frac{V_f A(t)}{K_{m_A} + A(t)} - \frac{V_f A_1(t)}{K_{m_A} + A_1(t)} - \frac{V_f A_1(t)}{K_{m_A} \left(1 + \frac{B_1(t)}{K_{m_B}}\right) + A_1(t)} \\
 \frac{dA_2(t)}{dt} &= \frac{V_f A_1(t)}{K_{m_A} \left(1 + \frac{B_1(t)}{K_{m_B}}\right) + A_1(t)} - \frac{V_f A_2(t)}{K_{m_A} + A_2(t)} \\
 \frac{dB_1(t)}{dt} &= \frac{V_f A_1(t)}{K_{m_A} + A_1(t)} - \frac{V_f B_1(t)}{K_{m_B} \left(1 + \frac{A_1(t)}{K_{m_A}}\right) + B_1(t)} \\
 \frac{dB_2(t)}{dt} &= \frac{V_f B_1(t)}{K_{m_B} \left(1 + \frac{A_1(t)}{K_{m_A}}\right) + B_1(t)} - \frac{V_f B_2(t)}{K_{m_B} + B_2(t)}
 \end{aligned}$$

4.2.2 Reversible kinetics



Parameter	V_f	V_r	k_A	$K_{m_A}^f$	$K_{m_A}^r$	$K_{m_B}^f$	$K_{m_B}^r$	A_{ex}
Value	1	1	1	0.02	0.02	0.02	0.02	0.05

Ordinary differential equation system neglecting substrate competition

$$\begin{aligned}
\frac{dA(t)}{dt} &= k_A A_{ex} - \frac{\frac{V_f A(t)}{K_{m_A}^f} - \frac{V_r A_1(t)}{K_{m_A}^r}}{1 + \frac{A(t)}{K_{m_A}^f} + \frac{A_1(t)}{K_{m_A}^r}} \\
\frac{dA_1(t)}{dt} &= \frac{\frac{V_f A(t)}{K_{m_A}^f} - \frac{V_r A_1(t)}{K_{m_A}^r}}{1 + \frac{A(t)}{K_{m_A}^f} + \frac{A_1(t)}{K_{m_A}^r}} - \frac{\frac{V_f A_1(t)}{K_{m_A}^f} - \frac{V_r A_2(t)}{K_{m_A}^r}}{1 + \frac{A_1(t)}{K_{m_A}^f} + \frac{A_2(t)}{K_{m_A}^r}} - \frac{\frac{V_f A_1(t)}{K_{m_A}^f} - \frac{V_r B_1(t)}{K_{m_B}^r}}{1 + \frac{A_1(t)}{K_{m_A}^f} + \frac{B_1(t)}{K_{m_B}^r}} \\
\frac{dA_2(t)}{dt} &= \frac{\frac{V_f A_1(t)}{K_{m_A}^f} - \frac{V_r A_2(t)}{K_{m_A}^r}}{1 + \frac{A_1(t)}{K_{m_A}^f} + \frac{A_2(t)}{K_{m_A}^r}} - \frac{V_f A_2(t)}{K_{m_A} + A_2(t)} \\
\frac{dB_1(t)}{dt} &= \frac{\frac{V_f A_1(t)}{K_{m_A}^f} - \frac{V_r B_1(t)}{K_{m_B}^r}}{1 + \frac{A_1(t)}{K_{m_A}^f} + \frac{B_1(t)}{K_{m_B}^r}} - \frac{\frac{V_f B_1(t)}{K_{m_B}^f} - \frac{V_r B_2(t)}{K_{m_B}^r}}{1 + \frac{B_1(t)}{K_{m_B}^f} + \frac{B_2(t)}{K_{m_B}^r}} \\
\frac{dB_2(t)}{dt} &= \frac{\frac{V_f B_1(t)}{K_{m_B}^f} - \frac{V_r B_2(t)}{K_{m_B}^r}}{1 + \frac{B_1(t)}{K_{m_B}^f} + \frac{B_2(t)}{K_{m_B}^r}} - \frac{V_f B_2(t)}{K_{m_B} + B_2(t)}
\end{aligned}$$

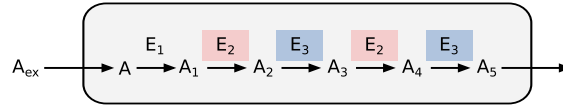
Ordinary differential equation system with substrate competition

$$\begin{aligned}
\frac{dA(t)}{dt} &= k_A A_{ex} - \frac{\frac{V_f A(t)}{K_{m_A}^f} - \frac{V_r A_1(t)}{K_{m_A}^r}}{1 + \frac{A(t)}{K_{m_A}^f} + \frac{A_1(t)}{K_{m_A}^r}} \\
\frac{dA_1(t)}{dt} &= \frac{\frac{V_f A(t)}{K_{m_A}^f} - \frac{V_r A_1(t)}{K_{m_A}^r}}{1 + \frac{A(t)}{K_{m_A}^f} + \frac{A_1(t)}{K_{m_A}^r}} - \frac{\frac{V_f A_1(t)}{K_{m_A}^f} - \frac{V_r A_2(t)}{K_{m_A}^r}}{1 + \frac{A_1(t)}{K_{m_A}^f} + \frac{A_2(t)}{K_{m_A}^r} + \frac{B_1(t)}{K_{m_B}^f} + \frac{B_2(t)}{K_{m_B}^r}} - \frac{\frac{V_f A_1(t)}{K_{m_A}^f} - \frac{V_r B_1(t)}{K_{m_B}^r}}{1 + \frac{A_1(t)}{K_{m_A}^f} + \frac{B_1(t)}{K_{m_B}^r}} \\
\frac{dA_2(t)}{dt} &= \frac{\frac{V_f A_1(t)}{K_{m_A}^f} - \frac{V_r A_2(t)}{K_{m_A}^r}}{1 + \frac{A_1(t)}{K_{m_A}^f} + \frac{A_2(t)}{K_{m_A}^r} + \frac{B_1(t)}{K_{m_B}^f} + \frac{B_2(t)}{K_{m_B}^r}} - \frac{V_f A_2(t)}{K_{m_A} + A_2(t)} \\
\frac{dB_1(t)}{dt} &= \frac{\frac{V_f A_1(t)}{K_{m_A}^f} - \frac{V_r B_1(t)}{K_{m_B}^r}}{1 + \frac{A_1(t)}{K_{m_A}^f} + \frac{B_1(t)}{K_{m_B}^r}} - \frac{\frac{V_f B_1(t)}{K_{m_B}^f} - \frac{V_r B_2(t)}{K_{m_B}^r}}{1 + \frac{A_1(t)}{K_{m_A}^f} + \frac{A_2(t)}{K_{m_A}^r} + \frac{B_1(t)}{K_{m_B}^f} + \frac{B_2(t)}{K_{m_B}^r}} \\
\frac{dB_2(t)}{dt} &= \frac{\frac{V_f B_1(t)}{K_{m_B}^f} - \frac{V_r B_2(t)}{K_{m_B}^r}}{1 + \frac{A_1(t)}{K_{m_A}^f} + \frac{A_2(t)}{K_{m_A}^r} + \frac{B_1(t)}{K_{m_B}^f} + \frac{B_2(t)}{K_{m_B}^r}} - \frac{V_f B_2(t)}{K_{m_B} + B_2(t)}
\end{aligned}$$

4.3 Case C - Multistep reactions with alternating enzymes

$$A(0) = A_1(0) = A_2(0) = A_3(0) = A_4(0) = A_5(0) = 0$$

4.3.1 Irreversible kinetics



Parameter	V_f	k_A	K_{m_A}	A_{ex}
Value	1	1	0.02	0.05

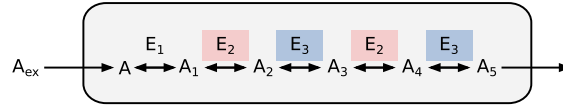
Ordinary differential equation system neglecting substrate competition

$$\begin{aligned}
\frac{dA(t)}{dt} &= k_A A_{ex} - \frac{V_f A(t)}{K_{m_A} + A(t)} \\
\frac{dA_1(t)}{dt} &= \frac{V_f A(t)}{K_{m_A} + A(t)} - \frac{V_f A_1(t)}{K_{m_A} + A_1(t)} \\
\frac{dA_2(t)}{dt} &= \frac{V_f A_1(t)}{K_{m_A} + A_1(t)} - \frac{V_f A_2(t)}{K_{m_A} + A_2(t)} \\
\frac{dA_3(t)}{dt} &= \frac{V_f A_2(t)}{K_{m_A} + A_2(t)} - \frac{V_f A_3(t)}{K_{m_A} + A_3(t)} \\
\frac{dA_4(t)}{dt} &= \frac{V_f A_3(t)}{K_{m_A} + A_3(t)} - \frac{V_f A_4(t)}{K_{m_A} + A_4(t)} \\
\frac{dA_5(t)}{dt} &= \frac{V_f A_4(t)}{K_{m_A} + A_4(t)} - \frac{V_f A_5(t)}{K_{m_A} + A_5(t)}
\end{aligned}$$

Ordinary differential equation system with substrate competition

$$\begin{aligned}
\frac{dA(t)}{dt} &= k_A A_{ex} - \frac{V_f A(t)}{K_{m_A} + A(t)} \\
\frac{dA_1(t)}{dt} &= \frac{V_f A(t)}{K_{m_A} + A(t)} - \frac{V_f A_1(t)}{K_{m_A} \left(1 + \frac{A_3(t)}{K_{m_A}}\right) + A_1(t)} \\
\frac{dA_2(t)}{dt} &= \frac{V_f A_1(t)}{K_{m_A} \left(1 + \frac{A_3(t)}{K_{m_A}}\right) + A_1(t)} - \frac{V_f A_2(t)}{K_{m_A} \left(1 + \frac{A_4(t)}{K_{m_A}}\right) + A_2(t)} \\
\frac{dA_3(t)}{dt} &= \frac{V_f A_2(t)}{K_{m_A} \left(1 + \frac{A_4(t)}{K_{m_A}}\right) + A_2(t)} - \frac{V_f A_3(t)}{K_{m_A} \left(1 + \frac{A_1(t)}{K_{m_A}}\right) + A_3(t)} \\
\frac{dA_4(t)}{dt} &= \frac{V_f A_3(t)}{K_{m_A} \left(1 + \frac{A_1(t)}{K_{m_A}}\right) + A_3(t)} - \frac{V_f A_4(t)}{K_{m_A} \left(1 + \frac{A_2(t)}{K_{m_A}}\right) + A_4(t)} \\
\frac{dA_5(t)}{dt} &= \frac{V_f A_4(t)}{K_{m_A} \left(1 + \frac{A_2(t)}{K_{m_A}}\right) + A_4(t)} - \frac{V_f A_5(t)}{K_{m_A} + A_5(t)}
\end{aligned}$$

4.3.2 Reversible kinetics



Parameter	V_f	V_r	k_A	K_{m_A}	A_{ex}
Value	1	1	1	0.02	0.05

Ordinary differential equation system neglecting substrate competition

$$\begin{aligned}
\frac{dA(t)}{dt} &= k_A A_{ex} - \frac{V_f A(t)}{K_{m_A}} - \frac{V_r A_1(t)}{K_{m_A}} \\
\frac{dA_1(t)}{dt} &= \frac{V_f A(t)}{K_{m_A}} - \frac{V_r A_1(t)}{K_{m_A}} - \frac{V_f A_1(t)}{K_{m_A}} - \frac{V_r A_2(t)}{K_{m_A}} \\
\frac{dA_2(t)}{dt} &= \frac{V_f A_1(t)}{K_{m_A}} - \frac{V_r A_2(t)}{K_{m_A}} - \frac{V_f A_2(t)}{K_{m_A}} - \frac{V_r A_3(t)}{K_{m_A}} \\
\frac{dA_3(t)}{dt} &= \frac{V_f A_2(t)}{K_{m_A}} - \frac{V_r A_3(t)}{K_{m_A}} - \frac{V_f A_3(t)}{K_{m_A}} - \frac{V_r A_4(t)}{K_{m_A}} \\
\frac{dA_4(t)}{dt} &= \frac{V_f A_3(t)}{K_{m_A}} - \frac{V_r A_4(t)}{K_{m_A}} - \frac{V_f A_4(t)}{K_{m_A}} - \frac{V_r A_5(t)}{K_{m_A}} \\
\frac{dA_5(t)}{dt} &= \frac{V_f A_4(t)}{K_{m_A}} - \frac{V_r A_5(t)}{K_{m_A}} - \frac{V_f A_5(t)}{K_{m_A}}
\end{aligned}$$

Ordinary differential equation system with substrate competition

$$\begin{aligned}
\frac{dA(t)}{dt} &= k_A A_{ex} - \frac{V_f A(t)}{K_{m_A}} - \frac{V_r A_1(t)}{K_{m_A}} \\
\frac{dA_1(t)}{dt} &= \frac{V_f A(t)}{K_{m_A}} - \frac{V_r A_1(t)}{K_{m_A}} - \frac{V_f A_1(t)}{K_{m_A}} - \frac{V_r A_2(t)}{K_{m_A}} \\
\frac{dA_2(t)}{dt} &= \frac{V_f A_1(t)}{K_{m_A}} - \frac{V_r A_2(t)}{K_{m_A}} - \frac{V_f A_2(t)}{K_{m_A}} - \frac{V_r A_3(t)}{K_{m_A}} \\
\frac{dA_3(t)}{dt} &= \frac{V_f A_2(t)}{K_{m_A}} - \frac{V_r A_3(t)}{K_{m_A}} - \frac{V_f A_3(t)}{K_{m_A}} - \frac{V_r A_4(t)}{K_{m_A}} \\
\frac{dA_4(t)}{dt} &= \frac{V_f A_3(t)}{K_{m_A}} - \frac{V_r A_4(t)}{K_{m_A}} - \frac{V_f A_4(t)}{K_{m_A}} - \frac{V_r A_5(t)}{K_{m_A}} \\
\frac{dA_5(t)}{dt} &= \frac{V_f A_4(t)}{K_{m_A}} - \frac{V_r A_5(t)}{K_{m_A}} - \frac{V_f A_5(t)}{K_{m_A}}
\end{aligned}$$

References

- [1] T. C. Chou and P. Talaly, "A simple generalized equation for the analysis of multiple inhibitions of michaelis-menten kinetic systems.," *J Biol Chem*, vol. 252, pp. 6438–6442, Sep 1977.

Quantitative Model of Cell Cycle Arrest and Cellular Senescence in
Primary Human Fibroblasts.

Schäuble S, Klement K, Marthandan S, Munch S, Heiland I,
Schuster S, Hemmerich P, Diekmann S, Heiland I, *PLoS ONE*, 7(8):
e42150, 2012.

SUPPLEMENTARY MATERIAL

Supplement S1 Bifurcation analysis of the stress response $F(t)$

$$\frac{dF}{dt} = 0$$

$$F_1: F_1 = 0$$

$$F_2: T \left(1 - \frac{F(t)}{K} \right) = \frac{F(t)}{1+F(t)^2} \quad (\text{S1.1})$$

When is $T \left(1 - \frac{F(t)}{K} \right)$ at a tangent to $\frac{F(t)}{1+F(t)^2}$?

$$\begin{aligned} \frac{d}{dF} \left(T \left(1 - \frac{F(t)}{K} \right) \right) &= \frac{d}{dF} \frac{F(t)}{1+F(t)^2} \\ -\frac{T}{K} &= \frac{1-F(t)^2}{(1+F(t)^2)^2} \end{aligned} \quad (\text{S1.2})$$

Substituting (S1.2) into (S1.1) yields

$$\begin{aligned} T - \frac{F(t)^2 - 1}{(1+F(t)^2)^2} F(t) &= \frac{F(t)}{1+F(t)^2} \\ T &= \frac{F(t)^3 - F(t) + F(t) + F(t)^3}{(1+F(t)^2)^2} \\ T &= \frac{2F(t)^3}{(1+F(t)^2)^2} \end{aligned} \quad (\text{S1.3})$$

Thus, a saddle-node bifurcation occurs if T increases and becomes greater than (S1.3). The current stable fixpoint is annihilated by the instable one. Increasing F further leads to a rapid change to the only stable fixpoint left, resembling bistable behavior.

Supplement S2 Sensitivity analysis

The sensitivity of each parameter k was approximated by fixating all parameters except for k , which was uniformly varied in the vicinity of 10% around its optimal value. Repeated 100 times for each parameter, this resulted in a set of 100 simulation runs per parameter k . The variance of the final population doubling (PD) values of these runs was further analyzed and interpreted as sensitivity measure to variation of parameter k .

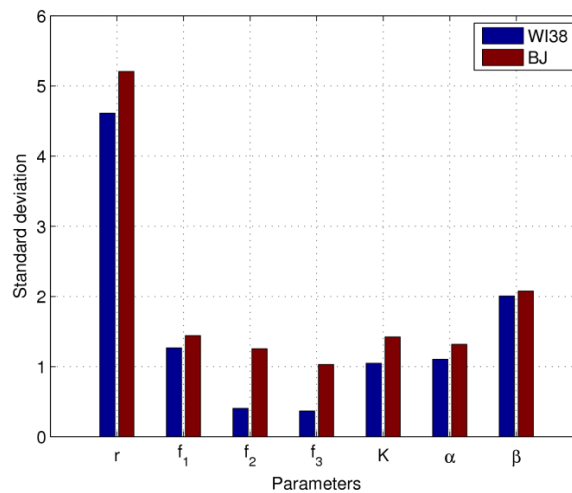
The sensitivity s_k for variation of parameter k was calculated via

$$s_k = \frac{1}{n-1} \left(\sum_{i=1}^n (x_i - \bar{x})^2 \right)^{\frac{1}{2}},$$

with n being the number of simulation runs and x_i the final PD value of simulation i .

This procedure was conducted using the BJ and WI-38 replicative senescence data and has been repeated for every parameter given by model Equations 3a-d and 4. The sensitivities resulting from the variation of single parameter values are shown in Figure S2.1.

Figure S2.1



Sensitivity of model parameters as described by Eq. 3a-d and Eq. 4. The model is most sensitive to the variation of the growth rate r , resembling its particular importance for the maximal replicative capacity of cell cultures. Regarding the parameters f_{1-3} , BJ as well as WI-38 cells show qualitatively the highest sensitivity for $P \rightarrow C$ transition, followed by its back transition (with rate f_2) and the $C \rightarrow S$ transition rate f_3 . Interestingly, WI-38 cells are notably less sensitive to f_2 and f_3 variation than BJ cells. β shows a higher sensitivity than α , as it is a concentration dependent parameter, whereas concentration independent α is responsible for linearly accumulating long term stress.

Supplement S3 Analytical Analysis of the complete cellular senescence model

Assumption: Let $F(t)$ be constant: $f_1 F(t) \equiv \hat{f}_1$

The ODE system reduces to:

$$\begin{aligned}\dot{P} &= rP - \hat{f}_1 P + f_2 C = (r - \hat{f}_1)P + f_2 C \\ \dot{C} &= \hat{f}_1 P - f_2 C - f_3 C \\ \dot{S} &= f_3 C\end{aligned}\tag{S3.1}$$

Observation: As the concentration of P increases with parameter r , no conservation relation can be retrieved from this system.

However, calculating the fractions $R_1 = \frac{P}{P+C+S}$, $R_2 = \frac{C}{P+C+S}$ and $R_3 = \frac{S}{P+C+S}$, yields the conservation relation:

$$R_1 + R_2 + R_3 = 1\tag{S3.2}$$

The population doublings are calculated as follows:

$$PD = \log_2([P + C + S])\tag{S3.3}$$

By substituting (S3.1) into (S3.3) we derive:

$$\dot{PD} = \log_2(P + C + S)\tag{S3.4}$$

$$= \frac{1}{\ln 2} \frac{1}{P + C + S} (\dot{P} + \dot{C} + \dot{S})\tag{S3.5}$$

$$\stackrel{(1)}{=} \frac{r}{\ln 2} \frac{P}{P + C + S} = \frac{r}{\ln 2} R_1\tag{S3.6}$$

Thus, the time dependent concentration change of \dot{PD} is zero, if r or R_1 is zero and greater than zero otherwise, since $r, R_1 > 0$. Consequently, R_1 is zero, if P is zero, implying that no proliferating cells are left to the population.

Moreover, applying the conservation relation (S3.2), it is sufficient to derive a solution for R_1 and R_2 in order to know the concentration fraction of R_3 .

Isoclines

$$\begin{aligned}
\dot{R}_1 &= \frac{\dot{P}(P + C + S) - P(\dot{P} + \dot{C} + \dot{S})}{(P + C + S)^2} \\
&= \frac{(r - \hat{f}_1)P}{P + C + S} + \frac{f_2 C}{P + C + S} - \frac{rP^2}{(P + C + S)^2} \\
&= -rR_1^2 + (r - \hat{f}_1)R_1 - f_2 R_2
\end{aligned} \tag{S3.7}$$

$$\begin{aligned}
\dot{R}_2 &= \frac{\dot{C}}{P + C + S} - \frac{rPC}{(P + C + S)^2} \\
&= \hat{f}_1 R_1 - (f_2 + f_3)R_2 - rR_1 R_2
\end{aligned} \tag{S3.8}$$

By setting (S3.7) and (S3.8) equal to zero we derive the following isoclines:

$$\begin{aligned}
\dot{R}_1 &= 0 : \\
R_2 &= \frac{r}{f_2} R_1^2 + \frac{\hat{f}_1 - r}{f_2} R_1 = y_1
\end{aligned} \tag{S3.9}$$

$$\begin{aligned}
\dot{R}_2 &= 0 : \\
R_2(rR_1 + f_2 + f_3) &= \hat{f}_1 R_1 \\
R_2 &= \frac{\hat{f}_1 R_1}{rR_1 + f_2 + f_3} = y_2
\end{aligned} \tag{S3.10}$$

Further analysis of (S3.9) and (S3.10):

$$y_1' = \frac{2r}{f_2} R_1 + \frac{\hat{f}_1 - r}{f_2} \tag{S3.11}$$

$$y_1'' = \frac{2r}{f_2} \tag{S3.12}$$

$$\begin{aligned}
y_1' &= 0 : \\
2rR_1 + \hat{f}_1 - r &= 0 \\
R_1^0 &= \frac{r - \hat{f}_1}{2r}
\end{aligned} \tag{S3.13}$$

$$y_1 = 0, y_2 = 0 :$$

$$R_1 \left(\frac{r}{f_2} R_1 + \frac{\hat{f}_1 - r}{f_2} \right) = 0$$

$$y_1^0 = 0, \tag{S3.14}$$

$$\frac{r - \hat{f}_1}{f_2} \frac{f_2}{r} = y_1^0 = \frac{r - \hat{f}_1}{r} \tag{S3.15}$$

$$y_2^0 = 0 \tag{S3.16}$$

$$\lim_{t \rightarrow \infty} y_2 = \frac{\hat{f}_1}{r} \tag{S3.17}$$

It is now convenient to look into a graphical representation of y_1 and y_2 . According to (S3.9), (S3.10) and (S3.13- S3.17) the graphs look as follows:

Figure S3.1: $r > \hat{f}_1$

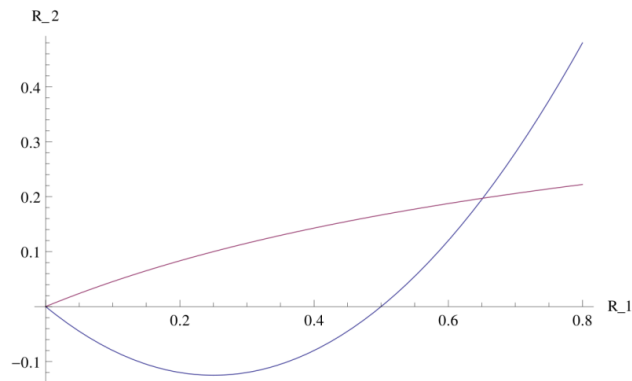


Figure S3.2: $r = \hat{f}_1$

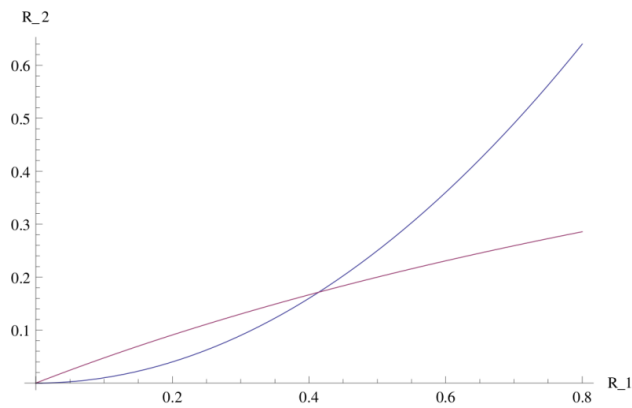
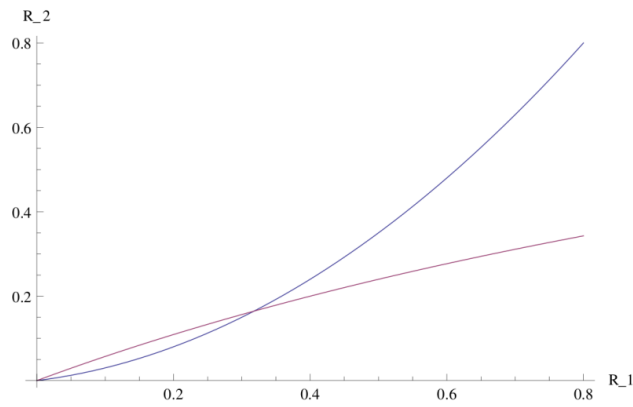
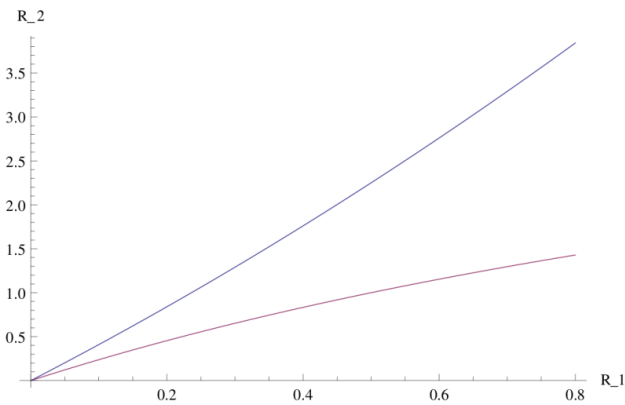


Figure S3.3: $r < \hat{f}_1$ 

The parameter settings that relate to Figure S3.1 and S3.2 always yield a stable nonzero solution for R_1 and R_2 and thus $\dot{P}D \neq 0$. Now, case 3 (Figure S3.3) is of particular interest, as further decrease of r relative to \hat{f}_1 yields:

Figure S3.4: $r \ll \hat{f}_1$ 

Parameter settings according to Figure S3.3, again yield a nonzero stable solution for R_1 and R_2 , whereas in the case of Figure S3.4, the only stable solution for R_1 and R_2 is zero.

We derive this solution by computing:

$$y_1'(0) \geq y_2'(0) \quad (\text{S3.18})$$

$$y_1'(0) = \frac{\hat{f}_1 - r}{f_2} \quad (\text{S3.19})$$

$$\frac{\hat{f}_1(f_2 + f_3)}{(f_2 + f_3)^2} = y'_2(0) = \frac{\hat{f}_1}{f_2 + f_3} \quad (\text{S3.20})$$

The fixed point at $(0,0)$ in the (R_1, R_2) -plane is stable, if (S3.18) holds.

$$\begin{aligned} \frac{\hat{f}_1 - r}{f_2} &\geq \frac{\hat{f}_1}{f_2 + f_3} \\ \hat{f}_1 f_3 &\geq r(f_2 + f_3) \\ \hat{f}_1 &\geq r \left(\frac{f_2}{f_3} + 1 \right) \end{aligned} \quad (\text{S3.21})$$

$$r \leq \frac{\hat{f}_1 f_3}{f_2 + f_3} \quad (\text{S3.22})$$

Thus, R_1 and consequently (S3.4) equals zero, if (S3.21) or (S3.22) holds, respectively, or is greater than zero otherwise.

Predicting the Physiological Role of Circadian Metabolic Regulation
in the Green Alga *Chlamydomonas reinhardtii*.

Schäuble S, Heiland I, Voytsekh O, Mittag M, Schuster S, *PLoS ONE*, 6(8): e23026, 2011a.

SUPPLEMENTARY MATERIAL

Supplementary Material

Table S1

Abbreviation	Name
13BPG	1,3-Bisphosphoglycerate
23DHD	2,3-dihydrodipicolinate
2PG	2-Phospho-glycerate
3PG	3-Phospho-glycerate
6PGN	6-Phospho-gluconate
Ac	Acetate
AcCoA	Acetyl-coenzyme-A
AcD	Acetaldehyde
AcGlu	Acetyl-glutamate
AcGlu5s	Acetyl-glutamate 5-semialdehyde
AcGluP	Acetylglutamyl-phosphate
Aconitate	Aconitate
AcOrn	Acetyl-ornithine
AcP	Acetylphosphate
ADP	Adenosine diphosphate
aKetG	α -Ketoglutarate
Ala	Alanine
AMP	Adenosine monophosphate
Arg	Arginine
ArgSucc	Arginino-succinate
Asn	Asparagine
Asp	Aspartate
AspSemi	Aspartate-semialdehyde
Asyl4Po	Aspartyl-4-phosphate
ATP	Adenosine triphosphate
biPO ₄ ³⁻	Pyrophosphate
CarPO	Carbamoyl-phosphate
Citr	Citrulline
Citrate	Citrate
CO ₂	Carbon dioxide
CoA	Coenzyme A
DHAP	Dihydroxyacetone-phosphate

continued on next page

Abbreviation	Name
DiAPi	L,L-Diaminopimelate
E4P	Erythrose-4-phosphate
F16BP	Fructose-1,6-biphosphate
F6P	Fructose-6-phosphate
Fum	Fumerate
G6P	Glucose-6-phosphate
GAP	Glyceraldehyde-3-phosphate
GL6P	Glucono-1,5-lactone-6-phosphate
Gln	Glutamine
Glu	Glutamate
Glu5P	Glutamate-5-phosphate
Glu5s	Acetyl-glutamate-5-semialdehyde
Gly	Glycine
GlyOx	Glyoxylate
H	Hydrogen
H ₂ O	Water
HCO ₃ ⁻	Hydrogencarbonate
Isocitrate	Isocitrate
Lys	Lysine
Malate	Malate
MDiAPi	Meso-diaminopimelate
NAD(H)	Nicotinamide adenine dinucleotide
NADP(H)	Nicotinamide adenine dinucleotide phosphate
NH ₄ ⁺	Ammonium
NO ₂ ⁻	Nitrite
NO ₃ ⁻	Nitrate
Orn	Ornithine
Oxaloacetate	Oxaloacetate
oxFdx	oxidized Ferredoxin
PEP	Phosphoenolpyruvate
PO ₄ ³⁻	Phosphate
Pyr	Pyruvate
R5P	Ribose-5-phosphate
redFdx	reduced Ferredoxin
Ru5P	Ribulose-5-phosphate
S17BP	Sedoheptulose-1,7-bisphosphate

continued on next page

Abbreviation	Name
S7P	Sedoheptulose-7-phosphate
SuccCoA	Succinylcoenzyme-A
Succinate	Succinate
tHD	Tetrahydrodipicolinate
UQ	Ubiquinone
UQH ₂	Ubiquinol
X5P	Xylulose-5-phosphate

Supplementary Material

Table S2

Abbrevia- tion	Name	EC- number	JGI-ID	UG _{≥7} -repeat
Anase	Asparaginase	3.5.1.1	187983	3*i
ACD	Acetaldehyde Dehydrogenase	1.2.1.10	133318	1*i
AdK	Adenylate Kinase	2.7.4.3	24512	
			103322	
			114363	
			118113	1*i
			129362	
			133184	
			194134	
			194947	
			196741	
ACH	Aconitate Hydratase	4.2.1.3	129025	
			195293	
ACS	Acetyl CoA Synthetase	6.2.1.1	196311	1*i
			119897	1*i
			391501	
			194063	
AGK	Acetylglutamate Kinase	2.7.2.8	78991	
			143603	1*i
AGS	Argininosuccinate Synthase	6.3.4.5	58140	
AGT	Alanine-glyoxylate transaminase	2.6.1.44	133057	
			194541	
			205967	1*i
AHD	Aspartate Kinase	2.7.2.4	161455	
			196316	
			196925	
AK	Acetate Kinase	2.7.2.1	128476	1*i
			129982	1*i
AKDG	α-Ketoglutarate Dehydrogenase Complex	1.8.1.4	57890	1*i
		1.8.1.5	205763	
		2.3.1.61	145395	
		2.3.1.62	145987	

continued on next page

Abbrevia- tion	Name	EC- number	JGI-ID	UG _{≥7} -repeat
AKDG	α-Ketoglutarate Dehydrogenase Complex	1.2.4.2	79471	
ALD	Aldehyde Dehydrogenase	1.2.1.3	135609	1*i
ALDO	Aldolase	4.1.2.13	24459	
			29185	
			152892	
			196304	
AOD	Acetylornithinase	3.5.1.16	155546	
ARG1	Acetyl-gamma-glutamyl-phosphate Reductase	1.2.1.38	191987	1*i, 1*3' UTR
ARG9	Acetylornithine Aminotransferase	2.6.1.11	119395	
			139007	
AS	Asparagine Synthase	6.3.5.4	140252	
			167865	
ASL	Argininosuccinate Lyase	4.3.2.1	101662	1*3' UTR
ASSD	Aspartate Semialdehyde Dehydrogenase	1.2.1.11	148810	
AST	Aspartate Aminotransferase	2.6.1.1	118364	2*i
			126943	2*i
			129557	
			146923	
			170056	1*i
			174097	2*i
			186959	2*i
			206390	2*i
BP	Inorganic Pyrophosphatase	3.6.1.1	133620	
			137778	
			174103	
CIS	Citrate Synthase	2.3.3.1	24263	2*i
			194915	
CMPS	Carbamoyl-phosphate Synthase	6.3.5.5	128227	
			195255	
DPA	Diaminopimelic Acid Aminotransferase	2.6.1.-	129557	
DPD	Diaminopimelate Decarboxylase	4.1.1.20	146886	3*i
DPE	Diaminopimelate Epimerase	5.1.1.7	150957	
DPR	Dihydrodipicolinate Reductase	1.3.1.26	205760	
DPR	Dihydrodipicolinate Reductase	1.3.1.26	395433	1*i
DPS	Dihydrodipicolinate Synthase	4.2.1.52	126518	

continued on next page

Abbrevia- tion	Name	EC- number	JGI-ID	UG _{≥7} -repeat
En	Enolase	4.2.1.11	83064	
FBA	Sedoheptulose-Bisphosphate Aldolase	4.1.2.-	24459	
FBP	Fructose-1,6-Bisphosphatase	3.1.3.11	24084	
Fdx GOGAT	Glutamate Synthase, Ferredoxin-dependent	1.4.7.1	140487	
FR	Ferredoxin Reductase		195553	
FUM	Fumarate Hydratase	4.2.1.2	195953 282848	
G6PI	Phosphoglucose Isomerase	5.3.1.9	135220	1*3' UTR
GAPD	Glyceraldehyde-3-phosphate Dehydrogenase	1.2.1.12	16652 102889 129019 140618 153894 195910 196442 196443	2*i
GAPN	Triosephosphate Dehydrogenase	1.2.1.9	102889	
GDH	Glutamate Dehydrogenase	1.4.1.3	82916 137469	2*i 1*i
GLD	Glucose-6-phosphate-1-Dehydrogenase	1.1.1.49	119861 309903 173841	1*i, 1*e
GNA	Glutamate Acetyltransferase	2.3.1.35	288806	3*i
GS	Glutamine Synthetase	6.3.1.2	129468 133971 136895 147468	1*3' UTR 1*3' UTR
GSD	Glutamate-5-semialdehyde Dehydrogenase	1.2.1.41	130812	
ICL	Isocitrate Lyase	4.1.3.1	104431 191668	1*i
IDH	Isocitrate Dehydrogenase (NADP ⁺ -dependent) Isocitrate Dehydrogenase (NAD ⁺ -dependent)	1.1.1.42	196567 196042	1*3' UTR

continued on next page

Abbrevia- tion	Name	EC- number	JGI-ID	UG _{≥7} -repeat
IDH	Isocitrate Dehydrogenase (NAD ⁺ -dependent)		196044	1*i
MAS	Malate Synthase	2.3.3.9	196328	1*i
MDH	Malate Dehydrogenase	1.1.1.37	60444 126023 137163 158129	1*i
MME	Malate Dehydrogenase (NADP ⁺ dep.)	1.1.1.40	126820 147722 196351 196831 196832	3*i 3*i 2*i
MME	Malate Dehydrogenase (NAD ⁺ dep.)	1.1.1.38	196833	
NAD(P)H GOGAT	Glutamate Synthase, NAD(P)H-dependent	1.4.1.14	205746	
NAGS	Acetylglutamate Synthase	2.3.1.1	130199 288806	2*i
NiR	Nitrite Reductase	1.7.7.1	192085	6*i, 3' UTR
NK	NAD ⁺ Kinase	2.7.1.23	123446 165793 196779	1*i
NP	NADH-ubiquinone oxidoreductase	1.6.5.3	24195 54440 57090 58686 59411 77311 79362 127317 127639 131464 132151 132909 135635 139850 143441	

continued on next page

Abbrevia- tion	Name	EC- number	JGI-ID	UG _{≥7} -repeat
NP	NADH-ubiquinone oxidoreductase	1.6.5.3	145512	
			149240	
			164272	
			164424	
			174569	2*i
			182226	
			182302	
			182980	
			184222	
			184606	1*i
			185013	
			186185	
			186342	
			187994	
			188142	
			190543	
			190916	
			191146	
			193762	
			194458	
NR	Nitrate Reductase	1.7.1.1	184661	2*i
			139758	1*3' UTR
OTA	Ornithine Transaminase	2.6.1.13	195386	
OTC	Ornithine Transcarbamylase	2.1.3.3	188762	
PAT	Phosphate Acetyltransferase	2.3.1.8	11226	3*i
			191051	
PCK	Phosphoenolpyruvate Carboxykinase	4.1.1.49	196612	1*i
PDC	Pyruvate Dehydrogenase Complex	1.8.1.4	57890	1*i
			205763	
			145395	
			149206	
			149709	
			187285	
			196500	

continued on next page

Abbrevia- tion	Name	EC- number	JGI-ID	UG _{≥7} -repeat
PDC	Pyruvate Dehydrogenase Complex	1.2.4.1	139515	
			193810	1*i
			196469	
			206010	1*e/3' UTR
			190446	1*i
			155587	1*i
PEPC	Phosphoenolpyruvate Carboxylase	4.1.1.31	80312	2*i
			182821	
PFK	Phosphofructokinase	2.7.1.11	411593	2*i
			196310	
			196430	1*i
			196624	
PGD	6-Phosphogluconate Dehydrogenase	1.1.1.44	115511	
			120516	
			128576	
			158911	
			192597	
PGK	Phosphoglycerate Kinase	2.7.2.3	36313	
			132210	
			196383	
PGL	6-Phosphogluconolactonase	3.1.1.31	146607	
			390565	1*i
PGM	Phosphoglycerate Mutase	5.4.2.1	8761	
			21373	2*i
			30383	
			325517	
			119977	
			161085	
			196305	
PPDK	Pyruvate-phosphate Dikinase	2.7.9.1	206694	
			196616	
PRF	Pyruvate-Ferredoxin Oxidoreductase	1.2.7.1	206677	
PROB	Glutamate-5-kinase	2.7.2.11	170370	
			189050	
PYC	Pyruvate carboxylase	6.4.1.1	402089	1*i
PyrD	Pyruvate Decarboxylase	4.1.1.1	127786	1*i
PyrK	Pyruvate Kinase	2.7.1.40	104490	

continued on next page

Abbrevia- tion	Name	EC- number	JGI-ID	UG _{≥7} -repeat
PyrK	Pyruvate Kinase	2.7.1.40	107530	
			118203	
			119280	
			122254	
			136854	
			149896	
			196263	
			196261	
			196263	
RPE	Ribulose-5-phosphate 3-Epimerase	5.1.3.1	6964	
			135614	
RPI	Ribose-5-phosphate Isomerase	5.3.1.6	55838	
			205912	
SCS	Succinate CoA Synthetase	6.2.1.5	24101	1 ^{*i}
			56839	
			196569	1 ^{*i}
			196570	
SDH	Succinate Dehydrogenase	1.3.5.1	142231	1 ^{*i}
			183570	
			195641	
			394775	2 ^{*i}
TAL	Transaldolase	2.2.1.2	146574	
			176076	4 ^{*i}
			287436	
TIM	Triosephosphate Isomerase	5.3.1.1	26265	
TRK ₁ / TRK ₂	Transketolase	2.2.1.1	141319	

Table S1: Overview of modeled metabolites and corresponding abbreviations.

Table S2: Overview of modeled enzymes and corresponding EC-numbers, abbreviations as well as JGI database IDs (cre v4.0; <http://genome.jgi-psf.org/Chlre4/>). The code of the UG-repeats is as follows: i – intron, e – exon, 5'/3' UTR – the 5' or 3' untranslated region of an enzyme. For bold marked UG-repeat entries CHLAMY1 binding has been shown experimentally.

Metabolic costs of amino acid and protein production in *Escherichia coli*.

Kaleta C, Schäuble S, Rinas U, Schuster S. *Biotechnol J*, 8(9):
1105–1114, 2013.

SUPPLEMENTARY MATERIAL

- 2 -

Supplementary Text

S1 Metabolic network

We assumed that ammonium is the nitrogen and sulfate the sulfur source. Additionally, we provided an inflow of oxygen, water, carbon dioxide as well as protons, and allowed for the outflow of every compound that can be exported into the extra-cellular space (see (Feist et al., 2007) for details). We set the reaction glycine hydroxymethyltransferase (protein GlyA) to irreversible status in the direction of glycine production. Otherwise, the model can produce every amino acid from carbon dioxide, ammonium and sulfate alone in the ‘LP unlimited energy’ computations.

S2 Computation of total synthesis cost of amino acids present in the cell

For the computation of total amino acid synthesis cost, first the number of each amino acid present in one *E. coli* cell had to be determined. In ref. (Feist et al., 2007), the concentrations of all amino acids per gram dry-cell-mass in *E. coli* are documented. Multiplying this number with the dry-cell mass of an *E. coli* cell ($3 \cdot 10^{-13}$ g) and the number of molecules in one millimole ($6.022 \cdot 10^{20}$), for each amino acid, we can determine the number of molecules of each amino acid in one *E. coli* cell. These numbers are (in molecules per cell): Ala, $9.3E+7$; Arg, $5.3E+7$; Asn, $4.4E+7$; Asp, $4.4E+7$; Cys, $1.7E+7$; Glu, $4.8E+7$; Gln, $4.8E+7$; Gly, $1.1E+8$; His, $1.7E+7$; Ile, $5.2E+7$; Leu, $8.1E+7$; Lys, $6.2E+7$; Met, $2.8E+7$; Phe, $3.2E+7$; Pro, $4.0E+7$; Ser, $3.9E+7$; Thr, $4.6E+7$; Trp, $1.0E+7$; Tyr, $2.5E+7$ and Val, $7.6E+7$. Subsequently, this number has been multiplied with the inverse of the yield of the amino acid produced from glucose (from the ‘manual’ computations) to obtain the number of glucose molecules required to produce the corresponding amount of amino acid molecules.

S3 Linear programming

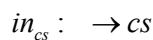
Given the stoichiometric matrix S of the metabolic network, in which rows correspond to metabolites and columns to reactions, we assumed in all cases that all intermediates are balanced (steady-state condition) and that reversible reactions are only used in thermodynamically feasible directions. For computational reasons we split reversible reactions into irreversible forward and backward steps. Hence, the steady-state and irreversibility conditions can be formulated as

$$Sv = 0$$

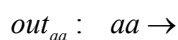
$$v \geq 0.$$

‘LP standard’

In order to determine the maximal yield of a particular amino acid aa from a given carbon source cs , the extra-cellular form of the carbon source was added as inflow of the form



to the network and an outflow of the amino acid in the form



was added, yielding the modified stoichiometric matrix S^1 . Subsequently, flux through in_{cs} was constrained to a maximum of 1 and the flux through out_{aa} was maximized while leaving all other fluxes unconstrained:

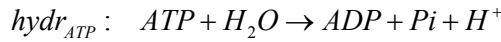
- 3 -

$$\begin{aligned}
 &\max \quad v_{out_{aa}} \\
 &s.t. \\
 &\mathbf{S}^1 \mathbf{v} = \mathbf{0} \\
 &\mathbf{v} \geq \mathbf{0} \\
 &v_{in_{cs}} \leq 1.
 \end{aligned}$$

The maximal yield of amino acid *aa* was then calculated as

$$Y_{aa}^{\max} = \frac{v_{out_{aa}}}{v_{in_{cs}}}.$$

Maximal ATP co-production of yield optimal amino acid biosynthetic flux distributions, $Y_{aa}^{\max, ATP}$ (used in Supplementary Material 1), was calculated by maximizing the flux through an artificial ATP hydrolysis reaction of the form

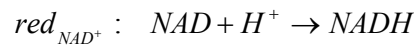


that was added to \mathbf{S}^1 to yield \mathbf{S}^2 , while fixing the flux through the amino acid outflow reaction to its maximum:

$$\begin{aligned}
 &\max \quad v_{hydr_{ATP}} \\
 &s.t. \\
 &\mathbf{S}^2 \mathbf{v} = \mathbf{0} \\
 &\mathbf{v} \geq \mathbf{0} \\
 &v_{in_{cs}} \leq 1 \\
 &v_{out_{aa}} = Y_{aa}^{\max}.
 \end{aligned}$$

‘LP unlimited energy’

In order to simulate the unlimited availability of energy, we introduced a reaction that provides energy by reducing NAD. This reaction has the form



and was added to the network to yield \mathbf{S}^3 . After adding the reduction reaction, computations were performed as described for ‘LP standard’ by replacing \mathbf{S}^1 with \mathbf{S}^3 . Note that besides the conversion of ADP and Pi to ATP through ATP synthase this reaction also allows for the unlimited reduction of $NADP^+$ through a transhydrogenase.

- 4 -

Supplementary Tables

Table S1 Metabolic investment for the production of amino acids from precursors. The number of moles of precursors, energy containing metabolites and other compounds that are consumed to produce one mole of each amino acid is indicated. Negative numbers indicate production. The last column ("ATP cons.") indicates the overall ATP consumption when assuming $2 \text{ ATP} = 2 \text{ FADH} = 1 \text{ NADH} = 1 \text{ NADPH}$ and starting biosynthesis from the precursors. For a list of abbreviations of precursors see Table S6. Abbreviations: AA, amino acid.

AA	Precursors								Energy			Others			ATP cons.
	Ru5P	E4P	PEP	PG	Pyr	AcCoA	AKG	OAA	ATP	NADH	NADPH	NH ₃	Mltbf	CO ₂	
Ala	0	0	0	0	1	0	0	0	0	0	1	1	0	0	2
Arg	0	0	0	0	0	0	1	0	5	-1	3	4	0	1	9
Asn	0	0	0	0	0	0	0	1	2	0	1	2	0	0	4
Asp	0	0	0	0	0	0	0	1	0	0	1	1	0	0	2
Cys	0	0	0	1	0	0	0	0	6	-1	3	1	0	0	10
Glu	0	0	0	0	0	0	1	0	0	0	1	1	0	0	2
Gln	0	0	0	0	0	0	1	0	1	0	1	2	0	0	3
Gly	0	0	0	1	0	0	0	0	0	-1	1	1	-1	0	0
His	1	0	0	0	0	0	0	0	6	-2	2	3	1	0	6
Ile	0	0	0	0	1	0	0	1	2	0	5	1	0	-1	10
Leu	0	0	0	0	2	1	0	0	0	-1	2	1	0	-2	2
Lys	0	0	0	0	1	0	0	1	2	0	4	2	0	-1	10
Met	0	0	0	0	0	0	0	1	10	-1	6	1	1	0	20
Phe	0	1	2	0	0	0	0	0	1	0	2	1	0	-1	5
Pro	0	0	0	0	0	0	1	0	1	0	3	1	0	0	7
Ser	0	0	0	1	0	0	0	0	0	-1	1	1	0	0	0
Thr	0	0	0	0	0	0	0	1	2	0	3	1	0	0	8
Trp	1	1	1	0	0	0	0	0	5	-1	1	2	0	-1	5
Tyr	0	1	2	0	0	0	0	0	1	-1	2	1	0	-1	3
Val	0	0	0	0	2	0	0	0	0	0	2	1	0	-1	4

- 5 -

Table S2 Number of moles of carbon source required (CS req.) and moles of ATP produced (ATP prod.) (assuming 1 NADH = 1 NADPH = 2 FADH = 2 ATP) for the production of one mole of each amino acid precursor.

Precursor	Glucose		Glycerol		Acetate	
	CS req.	ATP prod.	CS req.	ATP prod.	CS req.	ATP prod.
Ribulose-5-P (Ru5P)	1	3	1.66	1.66	3.33	-1.67
Erythrose-4-P (E4P)	0.75	1	1.33	1.33	2.67	-1.33
Phosphoenolpyruvat (PEP)	0.5	2	1	4	2	2
2-phosphoglycerate (PG)	0.5	2	1	4	2	2
Pyruvate (Pyr)	0.5	3	1	5	2	3
Acetyl-CoA (AcCoA)	0.5	5	1	7	1	1
α -Ketoglutarate (AKG)	1	9	2	13	3	4
Oxaloacetate (OAA)	0.5	2	1	4	2	3

- 6 -

Table S3 Characteristics of the production of amino acids from glucose for different calculation schemes as described in the main document. “Yield” indicates the maximal number of moles of the respective amino acid that can be produced from one mole of glucose. “ATP Cons.” corresponds to the number of moles of ATP that are required to produce one mole of the amino acid (assuming that 2 mol ATP can be produced from 1 mole NADH). “Carb. Yield” indicates the fraction of carbon atoms from the carbon source that are retained in the amino acid using the respective pathway. A number >1 indicates that there is a net consumption of carbon dioxide. Otherwise, carbon dioxide is released.

AS	Glucose								
	Manual			LP standard			LP unlimited energy		
	Yield (mol/mol)	ATP Cons.	Carb. Yield	Yield (mol/mol)	ATP Cons.	Carb. Yield	Yield (mol/mol)	ATP Cons.	Carb. Yield
Ala	2.000	-1	1.000	2.000	-0.25	1.000	2.000	-0.25	1.000
Arg	0.929	0	0.929	0.885	0	0.885	1.333	38.5	1.333
Asn	1.733	2	1.156	1.742	0	1.161	2.000	2	1.333
Asp	2.000	0	1.333	1.862	0	1.241	2.000	1	1.333
Cys	1.130	8	0.565	1.032	0	0.516	2.000	12.5	1.000
Glu	1.000	-7	0.833	1.152	0	0.960	1.333	30.25	1.111
Gln	1.000	-6	0.833	1.190	0	0.992	1.333	29.25	1.111
Gly	2.000	-2	0.667	2.731	0	0.910	4.000	4.125	1.333
His	0.820	3	0.820	0.885	0	0.885	1.200	14	1.200
Ile	0.839	7	0.839	0.750	0	0.750	1.000	9	1.000
Leu	0.667	-9	0.667	0.752	0	0.752	0.800	26	0.800
Lys	0.839	5	0.839	0.800	0	0.800	1.000	6.75	1.000
Met	0.689	18	0.574	0.621	0	0.517	2.000	36.25	1.667
Phe	0.571	0	0.857	0.560	0	0.839	0.600	3.25	0.900
Pro	1.000	-2	0.833	1.014	0	0.845	1.333	34.5	1.111
Ser	2.000	-2	1.000	2.000	-1	1.000	2.000	-1	1.000
Thr	1.368	6	0.912	1.297	0	0.865	2.000	7.25	1.333
Trp	0.444	-1	0.815	0.471	0	0.863	0.500	3.25	0.917
Tyr	0.571	-2	0.857	0.581	0	0.871	0.600	1.5	0.900
Val	1.000	-2	0.833	1.000	-0.75	0.833	1.000	-0.75	0.833

- 7 -

Table S4 Characteristics of the production of amino acids from glycerol for different calculation schemes. For a description see Table S.

AS	Glycerol								
	Manual			LP standard			LP unlimited energy		
	Yield (mol/mol)	ATP Cons.	Carb. Yield	Yield (mol/mol)	ATP Cons.	Carb. Yield	Yield (mol/mol)	ATP Cons.	Carb. Yield
Ala	1.000	-3	1.000	1.000	-2.5	1.000	1.000	-2.5	1.000
Arg	0.500	-4	1.000	0.512	0	1.024	0.667	35.125	1.333
Asn	1.000	0	1.333	1.000	-0.25	1.333	1.000	-0.25	1.333
Asp	1.000	-2	1.333	1.000	-1.25	1.333	1.000	-1.25	1.333
Cys	0.652	6	0.652	0.598	0	0.598	1.000	10.25	1.000
Glu	0.500	-11	0.833	0.605	0	1.008	0.667	26.875	1.111
Gln	0.500	-10	0.833	0.607	0	1.011	0.667	25.875	1.111
Gly	1.000	-4	0.667	1.539	0	1.026	2.000	3	1.333
His	0.483	4.33	0.966	0.489	0	0.978	0.600	10.5833	1.200
Ile	0.484	3	0.968	0.436	0	0.871	0.500	4.5	1.000
Leu	0.333	-15	0.667	0.381	0	0.762	0.400	20.375	0.800
Lys	0.484	1	0.968	0.466	0	0.931	0.500	2.25	1.000
Met	0.394	16	0.657	0.358	0	0.596	1.000	34	1.667
Phe	0.300	-4.33	0.900	0.300	-3.5833	0.900	0.300	-3.5833	0.900
Pro	0.500	-6	0.833	0.567	0	0.945	0.667	31.125	1.111
Ser	1.000	-4	1.000	1.000	-3.25	1.000	1.000	-3.25	1.000
Thr	0.778	4	1.037	0.753	0	1.004	1.000	5	1.333
Trp	0.250	-2	0.917	0.250	-4.25	0.917	0.250	-4.25	0.917
Tyr	0.300	-6.33	0.900	0.300	-5.3333	0.900	0.300	-5.3333	0.900
Val	0.500	-6	0.833	0.500	-5.25	0.833	0.500	-5.25	0.833

- 8 -

Table S5 Characteristics of the production of amino acids from acetate for different calculation schemes. For a description see Table S.

AS	Acetate								
	Manual			LP standard			LP unlimited energy		
	Yield (mol/mol)	ATP Cons.	Carb. Yield	Yield (mol/mol)	ATP Cons.	Carb. Yield	Yield (mol/mol)	ATP Cons.	Carb. Yield
Ala	0.50	-1	0.75	0.50	1.00	0.75	0.50	1.00	0.75
Arg	0.27	5	0.81	0.25	0.00	0.74	0.33	-7.50	1.00
Asn	0.47	1	0.93	0.47	0.00	0.95	0.50	-0.75	1.00
Asp	0.50	-1	1.00	0.50	0.25	1.00	0.50	0.25	1.00
Cys	0.32	8	0.48	0.27	0.00	0.40	0.50	-12.25	0.75
Glu	0.33	-2	0.83	0.33	0.75	0.83	0.33	0.75	0.83
Gln	0.33	-1	0.83	0.33	1.75	0.83	0.33	1.75	0.83
Gly	0.50	-2	0.50	0.71	0.00	0.71	1.00	-3.50	1.00
His	0.19	7.67	0.57	0.22	0.00	0.66	0.30	-13.25	0.90
Ile	0.21	6	0.62	0.20	0.00	0.60	0.25	-7.00	0.75
Leu	0.20	-5	0.60	0.20	3.25	0.60	0.20	3.25	0.60
Lys	0.22	4	0.66	0.21	0.00	0.64	0.25	-4.75	0.75
Met	0.19	17	0.47	0.16	0.00	0.41	0.50	-35.00	1.25
Phe	0.14	2.33	0.64	0.14	0.00	0.64	0.15	-2.75	0.68
Pro	0.29	3	0.73	0.29	0.00	0.71	0.33	-3.50	0.83
Ser	0.50	-2	0.75	0.50	1.25	0.75	0.50	1.25	0.75
Thr	0.37	5	0.74	0.35	0.00	0.70	0.50	-6.00	1.00
Trp	0.11	6	0.62	0.12	0.00	0.65	0.13	-3.25	0.69
Tyr	0.15	0.33	0.67	0.15	0.00	0.66	0.15	-1.00	0.68
Val	0.25	-2	0.63	0.25	2.25	0.63	0.25	2.25	0.63

Table S6 Abbreviations.

Abbr.	Metabolite	Abbr.	Metabolite
2PG	2-phosphoglycerate	Glyc(ex)	extra-cellular glycerol
Ac(ex)	extra-cellular acetate	Icit	Isocitrate
AcCoA	acetyl-Coenzym A	Mal	L-malate
AKG	2-oxoglutarate	OAA	Oxaloacetate
DHAP	dihydroxyacetonephosphate	PEP	Phosphoenolpyruvate
E4P	D-Erythrose-4-phosphate	Pyr	Pyruvate
F6P	D-fructose-6-phosphate	Ru5P	ribulose-5-phosphate
FDP	D-Fructose-1,6-bisphosphate	S7P	sedoheptulose-7-phosphate
G3P	glyceraldehyde-3-phosphate	Succ	Succinate
G6P	glucose-6-phosphate	Xu5P	xylulose-5-phosphate
Glc(ex)	extra-cellular D-glucose		

References

- Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., Broadbelt, L. J., Hatzimanikatis, V., Palsson, B. Ø., 2007. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* 3, 121.

One further supplementary table of metabolic costs with or without including polymerization costs is available at the website of the article:

<http://onlinelibrary.wiley.com/doi/10.1002/biot.201200267/supinfo>

B. Beitragende Autoren

Titel	Literaturangabe	Autoren	Anteil
Hands-on metabolism: analysis of complex biochemical networks using elementary flux modes.	<i>Methods Enzymol</i> 500, 2011, 437–456	Schäuble S , Schuster S, Kaleta C	70% 5% 25%
Predicting the Physiological Role of Circadian Metabolic Regulation in the Green Alga <i>Chlamydomonas reinhardtii</i>	<i>PLoS ONE</i> , 6, 2011, e23026	Schäuble S , Heiland I, Voytsekh O, Mittag M, Schuster S	40% 40% 5% 5% 10%
Metabolic costs of amino acid and protein production in <i>Escherichia coli</i>	<i>Biotechnol J</i> , 2013, 8, 1105–1114	Kaleta C, Schäuble S , Rinas U, Schuster S	50% 25% 20% 5%
Effect of substrate competition in kinetic models of metabolic networks.	<i>FEBS Lett</i> , 2013, 587, 2818–2824	Schäuble S , Stavrum A K, Puntervoll P, Schuster S, Heiland I	40% 40% 5% 5% 10%
Quantitative Model of Cell Cycle Arrest and Cellular Senescence in Primary Human Fibroblasts.	<i>PLoS ONE</i> , 7, 2012, e42150	Schäuble S , Klement K, Marthandan S, Munch S, Heiland I, Schuster S, Hemmerich P, Diekmann S	45% 10% 10% 3% 2% 5% 10% 15%

.....
bestätigt, Jun.-Prof. Dr. Christoph Kaleta

.....
bestätigt, Prof. Dr. Stefan Schuster

C. Selbständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Hilfsmittel angefertigt habe. Mir ist die geltende Promotionsordnung der Biologisch-Pharmazeutischen Fakultät bekannt. Bei der Auswahl und Auswertung des Materials, sowie auch bei der Herstellung des Manuskripts haben mich meine Kollegen am Lehrstuhl für Bioinformatik unter der Leitung von Prof. Dr. Stefan Schuster unterstützt. Ich habe weder die Hilfe eines Promotionsberaters in Anspruch genommen, noch haben Dritte unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die vorgelegte Dissertation wurde zuvor nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht. Ferner habe ich weder die gleiche, noch eine abgewandelte Form der vorliegenden Arbeit, oder eine andere Abhandlung bei einer anderen Hochschule als Dissertation eingereicht.

Jena, den 27. September 2013

.....
Sascha Schäuble