DISS. ETH NO. 18728

Reverse engineering of a metabolic, noise-exploiting control design

A dissertation submitted to

ETH ZURICH

for the degree of
Doctor of Sciences

presented by

Timo Oliver Kotte

Dipl.-Ing., University of Stuttgart, Germany
M.Sc., Georgia Institute of Technology, USA
born March, 22nd, 1978
citizen of Germany and Finland

accepted on the recommendation of

Prof. Dr. Sven Panke
Prof. Dr. Uwe Sauer
Prof. Dr. John Lygeros
Prof. Dr. Matthias Heinemann
2009

And how wonderful it is!
In the limitless ocean of Myself,
waves of beings arise, collide,
play for a while, and subside
— as is their nature.

Ashtavakra Gita, verse 2.25
unknown author

In memory of my Dad.

Contents

Contents

Summary

Most of life's diversity arises from genetic differences between species, and between individuals within a species. However, diversity also arises on the phenotypic level, e.g. through the diversification of a genetically uniform cell population into multiple stable, distinct and co–existing phenotypes. Our general understanding of phenotypic diversification is still in its infancy, yet this phenomenon is medically relevant as it has been linked to, for instance, HIV–1 latency and the resistance of pathogenic bacteria to antimicrobial agents.

Therefore, the fundamental research presented in this thesis aims at advancing our understanding of the general principles surrounding and generating phenotypic diversification of genetically uniform cell populations. Following up on a substantiated hypothesis derived from a chance observation, this thesis identifies such general principles using as model system the adaptation of *Escherichia coli*'s central carbon metabolism from glucose to acetate growth.

As phenotypic diversification is an emergent property of multiple molecular interactions, the identification of the generating mechanism amidst the molecular complexity is not straight–forward and requires the help of mechanistic models. However, the development of such models is severely hampered by uncertain values for the many parameters appearing in such models. To be able to nonetheless construct and analyze such a model, a novel approach to obtain a sound system understanding despite non–identifiable parameter values is presented in Chapter 2. This so–termed *divide–and–conquer approach* draws on steady state -omics measurement data and exploits a decomposition of the global parameter estimation problem into independent subproblems. The solutions to these subproblems are joined to the complete space of global optima, which can be easily analyzed. The conditions at which the decomposition occurs are mathematically derived, and practical strategies to fulfill these conditions are outlined. The approach is demonstrated with a toy model.

In Chapter 3, using the divide–and–conquer approach, a large–scale differential equation model of *E. coli*'s central metabolism and its enzymatic and genetic regulation is constructed. With this model, the system–wide adaptations of metabolic operation between glycolytic and gluconeogenic carbon sources (such as glucose and acetate), which are on the molecular level not yet understood, are investigated in mechanistic detail. First, it is shown that these adaptations can emerge from the interplay of already known interactions and thus do not require e.g. a hitherto not identified transmembrane sensor for acetate. Second, it is shown that the successful in silico adaptations follow the here proposed general mechanism of in-

Summary

direct carbon source recognition via *distributed sensing of intracellular metabolic fluxes*. This mechanism uses two distinct general motifs, termed *pathway usage* and *flux direction*, to establish *flux–signalling metabolites*, whose bindings to target transcription factors form *molecular sensors for intracellular metabolic flux*. These sensors are embedded in *global feedback loop architectures* that orchestrate the regulatory adjustments to recognized changes in carbon source availability. By connecting system-level understanding to molecular-level knowledge, these general principles improve our understanding of bacterial metabolism's operation in environments with fluctuating carbon sources.

In Chapter 4, using measurement techniques with single–cell resolution, it is experimentally proven that a genetically uniform *E. coli* population diversifies into two stable phenotypic subpopulations following a carbon source shift from glucose to gluconeogenic substrates such as acetate. This population–level strategy differs from the known diversification strategy of stochastic switching and is introduced as *responsive diversification*. It is further shown that the molecular mechanism responsible for generating the phenotypic subpopulations resides at the core of central metabolism, is a subcircuit of the regulatory circuitry modeled in Chapter 3, is centered on a there proposed flux sensor, and comprises two intertwined feedback and feedforward loops on the levels of metabolic and genetic regulation. This circuit ensures that only the cells most capable of gluconeogenic growth are allowed to grow, and consequently encodes the novel function of *performance–based selection* of phenotypes. These results show that within metabolism, phenotypic multistability, which has so far only been observed in certain substrate uptake pathways that feed into central metabolism, generalizes to central metabolism and can thus be viewed as an inherent feature of its design.

The observation of responsive diversification within central metabolism naturally raises the question *why* such diversification has evolved. Therefore, the optimality of this adaptation strategy is investigated in Chapter 5 with the help of evolutionary game theory. By means of a mathematical model, it is shown that in certain stochastic environments, population growth is *not* maximized by maximization of each cell's growth rate but by responsive diversification of phenotypes. Further, it is shown that this behavior 'resolves' an *apparent* tragedy of the commons dilemma, where selfish behavior of individuals is suboptimal to the whole population, through the adaptation strategy of bet–hedging, where different cells place different bets on the uncertain future. This population–level study is connected to the intracellular level by arguing that the two phenotypes specialize in opposing extremes of two molecular trade–offs; one of these trade–offs rests on the principle of performance–based selection proposed in Chapter 4 and is as such conceptually novel. When taken together, these results show that responsive diversification of phenotypes upon carbon source fluctuations can *simultaneously* optimize two intracellular trade–offs, increase the robustness of the total population to stresses, and maximize the size of the total population in the absence of stresses.

The general principles proposed by the here presented fundamental research,

along with the discovery of phenotypic bistability within central metabolism, may guide systematic disease research in finding ways to force cells trapped in sick phenotypes back to healthy phenotypes without introducing mutations.

Zusammenfassung

Die Vielfalt des Lebens beruht grösstenteils auf genetischen Unterschieden zwischen den Arten sowie zwischen den Individuen innerhalb einer Art. Zusätzlich entsteht diese Vielfalt aber auch auf phänotypischer Ebene, zum Beispiel durch die Diversifizierung einer genetisch identischen Zellpopulation in mehrere stabile, unterschiedliche und koexistierende Phänotypen. Unser Grundlagenverständnis der phänotypischen Diversifizierung steckt noch in den Kinderschuhen, jedoch ist dieses Phänomen medizinisch bedeutsam da es zum Beispiel mit der Latenz des HIV–1 Virus und der Resistenz pathogener Bakterien gegenüber antimikrobiellen Wirkstoffen in Verbindung gebracht wird.

Daher zielt die in dieser Arbeit vorgelegte Grundlagenforschung auf eine Vertiefung unseres Verständnisses der allgemeingültigen Grundsätze ab, welche phänotypische Diversifizierung erzeugen und sie umgeben. Diese Arbeit identifiziert solche allgemeingültigen Grundsätze durch das Folgen einer fundierten Hypothese, welche aus einer zufälligen Beobachtung abgeleitetet wurde. Als Modellsystem dient *Escherichia coli*s Anpassung des Zentralstoffwechsels von Glucose– zu Acetat–Wachstum.

Da phänotypische Diversifizierung aus dem Zusammenspiel vieler molekularer Interaktionen emergiert, ist die Identifikation des sie erzeugenden Mechanismus inmitten der molekularen Komplexität nicht trivial und erfordert die Hilfe mechanistischer Modelle. Allerdings wird die Entwicklung solcher Modelle dadurch stark behindert, dass die meisten Werte für die zahlreichen, in solchen Modellen vorkommenden Parameter unsicher sind. Um dennoch solche Modelle zu entwickeln und zu analysieren, wird in Kapitel 2 ein neuartiger Ansatz präsentiert, der es erlaubt, ein solides Systemverständnis trotz Parameterunsicherheiten zu erhalten. Dieser hier so benannte *Teile–und–Erobere Ansatz* stützt sich auf im stationären Zustand gewonnene -Omik Messdaten und basiert auf einer Zergliederung des globalen Parameterschätz-Problems in unabhängige Teilprobleme. Die Lösungen dieser Teilprobleme werden zum gesamten Lösungsraum der globalen Optima zusammengefügt, welcher leicht analysiert werden kann. Die Bedingungen für das Auftreten einer solchen Zergliederung werden mathematisch hergeleitet, und es werden praktisch anwendbare Strategien erläutert um diese Bedingungen herbeizuführen. Der Ansatz wird anhand eines kleinen Beispiel-Modells demonstriert.

In Kapitel 3 wird der Teile–und–Erobere Ansatz angewandt um ein umfangreiches Differentialgleichungsmodell von *E. coli*s Zentralstoffwechsel und seiner enzymatischen und genetischen Regulation zu entwickeln. Mit diesem Modell werden die systemweiten Stoffwechsel-Anpassungen zwischen Wachstum auf glyko-

lytischen und gluconeogenesischen Kohlenstoff–Quellen (wie Glukose und Acetat), welche auf molekularer Ebene noch nicht verstanden sind, in mechanistischem Detail untersucht. Zunächst wird gezeigt, dass diese Anpassungen aus dem Zusammenspiel bereits bekannter Wechselwirkungen hervortreten können und somit zum Beispiel ein bisher nicht identifizierter Transmembran-Sensor für Acetat nicht erforderlich ist. Darüber hinaus wird gezeigt, dass den erfolgreichen Anpassungen im Modell der hier vorgestellte, allgemeingültige Mechanismus der indirekten Substraterkennung durch *verteilte Messung intrazellulärer metabolischer Flüsse* zugrunde liegt. Dieser Mechanismus verwendet zwei unterschiedliche allgemeingültige Motive, benannt *Pfadnutzung* und *Flussrichtung*, um *Fluss–signalisierende Metabolite* zu erzeugen, deren Bindungen an Transkriptionsfaktoren *molekulare Sensoren für intrazelluläre metabolische Flüsse* erzeugt. Diese Sensoren sind in *globale Rückkopplungsarchitekturen* eingebettet, welche die regulatorischen Anpassungen an erkannte Substrat–Schwankungen aufeinander abstimmen. Diese allgemeinen Grundsätze verbinden globales Systemverständnis mit detailliertem biochemischen Wissen und vertiefen dadurch unserer Verständnis über die Funktionsweise des bakteriellen Stoffwechsels in Umgebungen mit schwankenden Kohlenstoff-Quellen.

Mittels Messtechniken mit Einzelzell-Auflösung wird in Kapitel 4 experimentell nachgewiesen, dass sich eine genetisch identische *E. coli* Population in zwei stabile phänotypische Subpopulationen diversifiziert sobald sie mit einem Substratwechsel von Glukose zu einem gluconeogenesischen Substrat wie Acetat konfrontiert wird. Diese Adaptationsstrategie auf Populationsebene unterscheidet sich von der bekannten Strategie der Diversifizierung durch stochastisches Umstellen und wird hier als *responsive Diversifizierung* benannt. Ausserdem wird gezeigt, dass sich der für die Erzeugung der phänotypischen Subpopulationen verantwortliche molekulare Mechanismus im Zentrum des zentralen Stoffwechsels befindet. Dieser Regulationsmechanismus ist ein Teil der in Kapitel 3 modellierten Regulation, basiert auf einem dort identifizierten Fluss–Sensor, und besteht aus zwei ineinander geflochtenen Feedback und Feedforward–Regulationen auf den Ebenen der metabolischen und genetischen Regulation. Dieser Mechanismus stellt sicher, dass nur die Zellen wachsen dürfen welche besonders gut auf gluconeogenesisches Wachstum eingestellt sind. Damit ermöglicht dieser Mechanismus die konzeptionell neue Funktion der *leistungsbasierten Selektion von Phänotypen*. Diese Ergebnisse zeigen, dass innerhalb des Stoffwechsels phänotypische Multistabilität nicht nur wie bisher vermutet in vereinzelten Substrat–Aufnahmepfaden auftritt, sondern auch im Zentralstoffwechsel erzeugt wird und daher als ein grundlegendes Designprinzip des Zentralstoffwechsels betrachtet werden kann.

Die Beobachtung responsiver Diversifizierung innerhalb des Zentralstoffwechsels geht einher mit die Frage *weshalb* eine solche Diversifizierung evolviert ist. Daher wird die Optimalität dieser Anpassungsstrategie in Kapitel 5 mit Hilfe der evolutionären Spieltheorie untersucht. Mittels eines mathematischen Modells wird gezeigt, dass in bestimmten stochastischen Umgebungen das Wachstum einer Zellpopulation *nicht* durch die Maximierung des Wachstums jeder einzelnen Zelle

maximiert wird, sondern durch responsive Diversifizierung in phänotypische Subpopulationen. Darüber hinaus wird gezeigt, dass dieses Verhalten mittels der Strategie der *Absicherung von Wetten*, bei der verschiedene Zellen verschiedene Wetten auf die unbekannte Zukunft abschliessen, ein augenscheinliches *Tragik der Allmende Dilemma* 'löst', bei dem egoistisches Verhalten von Individuen suboptimal für die Gesamtpopulation ist. Diese Adaptationsstrategie auf Populationsebene wird mittels der Argumentation, dass sich die beiden Phänotypen in entgegengesetzte Extreme zweier molekularer Kompromisse spezialisieren, mit der intrazellulären Ebene verknüpft. Einer dieser molekularen Kompromisse beruht auf dem Prinzip der in Kapitel 4 vorgeschlagenen leistungsbasierten Selektion und ist daher konzeptionell neu. Zusammengenommen zeigen diese Resultate, dass die responsive Diversifizierung von Phänotypen nach Substratwechseln *gleichzeitig* die Robustheit der Gesamtpopulation gegenüber Stress und die Grösse der Gesamtpopulation in Abwesenheit von Stress erhöhen sowie zwei molekulare Kompromisse optimieren kann.

Die aus der hier vorgelegten Grundlagenforschung hervorgegangenen allgemeingültigen Grundsätze sowie der Nachweis phänotypischer Bistabilität im Zentralstoffwechsel können der systematischen Krankheitsforschung als Leitfaden dienen, um in einem krankhaften Phänotyp gefangene Zellen ohne zusätzliche Mutationen auf einen gesunden Phänotyp zurückzuzwingen.

Chapter 1

Introduction

1.1 Relevance of this thesis

Most of life's diversity arises from genetic differences between species, and between individuals within a species. However, the diversity of life is not limited to genetic differences between individual organisms or cells. A second, epigenetic layer of diversity is added in the process of 'translating' the genetic information into actual cellular functioning. Such diversity, which is generated on the phenotypic level, becomes apparent when organisms or cells with identical genomes differ from each other. For instance, the fingerprints of genetically identical human twins can be easily distinguished [82], and genetically identical cats display different coat patterns and personalities [192].

Unfortunately, natural selection eventually resulting from such multi–layered diversification becomes threatening when this selection concerns diseases. For instance, tumors are long known to consist of a highly diversified cancer cell population [59]. A small fraction of these cells often survives therapeutic treatment and rapidly proliferates into a new tumor; furthermore, having emerged out of natural

selection in treatment conditions, the newly emerged tumor is often more resistant to future treatment [59]. Similarly, a small fraction of a pathogenic bacterial population often survives antibiotic treatment and continues to proliferate thereafter in the infected host. For instance, the opportunistic pathogen *Staphylococcus aureus* is one of the major causes of community–acquired diseases ranging from superficial skin infections to life–threatening systemic infections and very effectively resists treatment with antimicrobial agents [120]. A last prominent example is given by the exceptionally rapid diversification of the virus HIV, making drug discovery a very challenging task [173, 174].

To find cures for such diseases, a general understanding of the diversification–generating molecular processes is invaluable [97, 99]. A significant part of the diversification occurs on the genetic layer, e.g. through the increased mutation rates of cancer cells. However, phenotypic diversification of genetically identical cells is also known to significantly contribute to the diversification of diseases. For instance, in an isogenic population of bacteria, resistance to antimicrobial agents is transiently acquired by a phenotypic subpopulation [12]. Also, HIV–1 expresses two phenotypic subpopulations that correspond to latent and productive HIV–1 infection [27, 224]. Whereas the fundamentals of genotypic diversification have been studied for decades, phenotypic diversification has only recently received community–wide attention (for instance, see the reviews [166, 168, 193]), which is why our understanding of phenotypic diversification is still in its infancy [161]. Hence, there is a pressing need for fundamental research on phenotypic diversification.

The aim of the fundamental research presented in this thesis is to advance our understanding of the general principles surrounding and generating phenotypic diversification of isogenic cell populations. Because such general principles can be easier identified when the cellular background is comparatively simple, this thesis investigates the model bacterium *Escherichia coli*. More specifically, because metabolism is known to be closely connected to several major diseases (for instance, see [126, 135, 222]), and the central metabolism of *E. coli* is an exceptionally well–studied and well–characterized system, this thesis focuses on the emergence of phenotypic multistability in *E. coli*'s central metabolism.

This introduction is structured as follows. First, as this thesis identifies general principles using a microbe, the complexity of the microbial world and the general approaches followed by this thesis to address this complexity are briefly classified. This classification is followed by brief reviews of several fields that are of relevance to this thesis. The first of these reviews concerns the emergence of phenotypic multistability from nonlinear interactions and stochastic events, the second the general modeling approaches in computational systems biology, and the third evolutionary game theory. These reviews are followed by an outline of this thesis.

1.2 Complexity in the microbial world

In the microbial world, complexity has multiple sources, is of different types, and occurs on different layers. To obtain a comprehensive understanding of microbial complexity, these different aspects of complexity need to be appreciated in conjunction [96]. This section classifies the types of microbial complexity investigated in this thesis.

A popular notion of complex systems is that very large numbers of simple and identical elements interact with each other to produce complex behaviors [95]. On the intracellular level, however, complexity does *not* arise from interactions between a large number of identical elements. Rather, intracellular complexity arises from selective interactions between a large number of functionally specialized elements, most prominently proteins. The behavior of cells arising out of such complex interaction patterns is probably best described through the term 'coherent' [95]. We thus need to understand how *coherent* behavior arises out of *selective* interactions between *functionally specialized* elements. It has long been known that the behavior of a complex system depends at least as much on the interactions between the system's components as on the properties of the components themselves [227]. Hence, we need to understand the properties of individual genes and proteins *and also* understand the structure, function and dynamics of the system that is formed through interactions between these elements [96]. This aspect of microbial complexity, which is further introduced in Sections 1.3.1, 1.3.2 and 1.3.3, is researched by the scientific discipline of systems biology. The systems biology approach of this thesis is classified in Section 1.3.4 and applied in Chapters 2, 3 and 4 of this thesis.

On the population level, the emergence of complexity *does* follow the popular notion of arising from interactions between many identical elements — complexity arises from the interplay of genotypic and phenotypic (sub)populations, each comprising large numbers of almost identical individuals, with each other and the environment. It has long been noted that this type of complexity can lead to counter–intuitive and complex phenomena [218], giving rise to the scientific discipline of game theory with applications mostly in the social sciences. Traditional game theory has been merged with population biology to the field of evolutionary game theory, which aims at understanding the phenomena that emerge when a large number of almost identical cells interact with each other and the environment [129]. This aspect of microbial complexity, along with a brief review of the scientific discipline of evolutionary game theory, is further introduced in Section 1.4. Population–level complexity is the topic of Chapter 5 of this thesis.

1.3 Complexity on the intracellular level

The specific aspect of intracellular complexity investigated by this thesis is the emergence of multiple stable, distinct, and co–existing phenotypes within an isogenic cell population. Such phenotypic diversification, leading to phenotypic multi-

stability, arises from the interplay of nonlinear molecular interactions and stochastic events [166, 168, 193]. In the following sections, intracellular nonlinearity and stochasticity are first separately reviewed in Sections 1.3.1 and 1.3.2, respectively. Then, the emergence of phenotypic multistability from the interplay of intracellular nonlinearity and stochasticity is reviewed in Section 1.3.3.

1.3.1 Nonlinearity

If the dependency of an entity y on another entity x is linear, it can be mathematically described through the linear model $y = a\ x + b$, with parameters a and b quantifying the dependency. Such perfect linearity exist only in the world of mathematics, not in real nature. However, in certain situations, linear models are sufficient to capture the *dominant* behavior of a natural process. For instance, when the concentration of an intracellular compound x is sufficiently large and the cellular cytosol is sufficiently well–mixed, the diffusion rate r of an intracellular compound x due to the cell's growth rate μ follows approximately $r = \mu\ x$; if the compound is also subject to degradation, the combined diffusion and degradation rate r_d is approximately given by $r_d = (\mu + k_d)\ x$, with k_d the first–order rate constant of the degradation. Another example is provided by an enzyme whose catalyzed reaction rate r can be approximated (i) over the whole range of its substrate concentration S with a nonlinear kinetics, e.g. of Michaelis–Menten type $r = v_{max}\ S/(S + K_S)$ with v_{max} the maximal reaction rate and K_S the affinity constant for the substrate, and (ii) for the range of small substrate concentrations with the linear model $r = v_{max}\ S/K_S$. Because linear systems behave in a straight–forward and easily predictable manner, linearizations are in the field of engineering extensively applied to reduce the complexity of mathematical models and to thus facilitate the construction of man–made systems.

Whereas the possible behavior of linear systems is simple and limited, even small nonlinear systems can exhibit astoundingly complex behavior. For instance, depending on the choice of parameters, a simple predator–prey model exhibits oscillations and chaotic behavior [216]. The dynamic behavior of nonlinear intracellular systems is with success (for instance, see [144]) investigated by the field of biochemical system theory, using among other tools phase–plane plots and bifurcation analyses [183]. The method of bifurcation analysis is also applied in Chapter 4 of this thesis.

The analysis of nonlinear processes within living cells has led to a quantitative understanding of many, partly counter–intuitive phenomena. One example of such a counter–intuitive phenomenon is provided by the branch point effect [109]. This effect combines two simple Michaelis–Menten–type enzymes catalyzing the first diverging reactions after a branch point of metabolic pathways into an ultrasensitive switch, which can abruptly change the flux distribution through the two diverging pathways. The effect occurs for instance at the branch point of the TCA and the glyoxylate shunt in *E. coli*, where the Hill coefficient of the equivalent sigmoidal switch has been quantified as 8 or higher (a Hill–coefficients of 2 is already quite

high). It was later shown that such ultrasensitivity can also occur in converging as well as in linear pathways [49]. The branch point effect appears in Chapter 3 of this thesis.

Another characteristic of nonlinear systems that is not exhibited by linear systems is the capability for multiple steady states. In the context of cellular operation, multiple *stable* steady states correspond to multiple stable protein concentrations, which translate into multiple stable and distinct phenotypes among which a particular cell may 'choose'. The mechanism that generates a bistable protein concentration is illustrated in Figure 1.1 and rests on the existence of three intersections between the protein's concentration–dependent curves of production and combined protein dilution & degradation rates. Whereas protein dilution and degradation rates naturally depend on the protein concentration (see above), a protein production rate depends on the protein concentration only if these two entities mutually affect each other, requiring them to be involved in some form of feedback loop architecture. To enable three intersections between the two rate curves, the feedback architectures must be of a certain type, for instance positive or double–negative [193]. Further, the production rate curve must be of a certain shape and is most prominently S–shaped, i.e. sigmoidal [193]. Such a sigmoidal shape arises from cooperative binding events, which usually occur during the process of gene expression [193]. In Chapter 4 of this thesis, it will be shown that a sigmoidal shape of the protein production curve can also result from propagation of cooperative effects introduced by allosteric enzyme regulation on the metabolic layer.

1.3.2 Stochastic events

All events inside a biological cell ultimately depend on probabilistic collisions between molecules [147]. From such basic physical principles, it has long been predicted [186] that the concentrations of intracellular molecules are to a potentially large degree statistically uncertain, or 'noisy'. Such intracellular randomness has also long been experimentally observed [20, 125, 137, 197] and theoretically analyzed [21].

The advent of new measurement techniques with single–cell resolution, most notably fluorescence microscopy and flow cytometry [32] in conjunction with fluorescent protein reporters [195, 207], together with a renewed interest in the theoretical implications of molecular noise [5, 130, 131], encouraged the first quantitative studies of intracellular randomness [18, 28, 50, 116, 143]. These studies focused on random events in the central dogma of molecular biology, meaning in replication, gene activation, transcription, and translation; they showed that autorepression of replication and transcription suppresses noise, and that there are differences between the noise statistics of prokaryotes and eukaryotes. It has further been shown that noise levels increase with decreasing molecular abundances [14]. Taken together, these studies point to low–copy mRNAs as the primary source of cellular randomness, a view that has been adopted by the community [166, 168, 193].
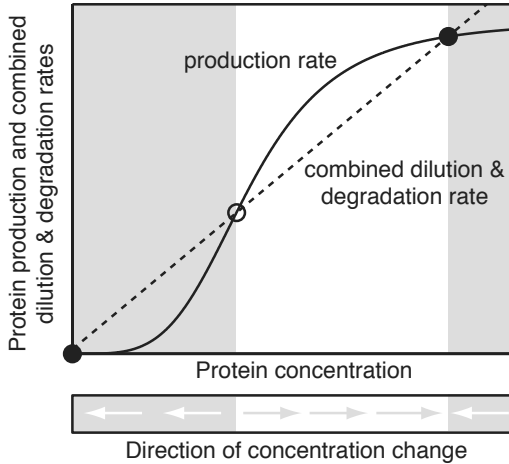
Figure 1.1: Generation of phenotypic multistability. Two simultaneously available stable steady states (filled circles), with convergence regions separated by an unstable steady state (empty circle), the watershed, arise where the rates of protein production and combined degradation & dilution intersect and are thus equal. The arrows in the bottom bar point into the direction of system dynamics; when protein production exceeds dilution & degradation, the concentration increases, and vice versa. A diversification of phenotypes with bistable protein concentrations occurs when intracellular randomness causes some cells to reside in the low and others in the high stable steady state.

However, it has been noted [147] that despite of these findings, low copy mRNA may in fact *not* be the primary source of intracellular randomness, and that the focus of the community on genes, RNAs and proteins may very well turn out to be a purely sociological choice, reflecting how attention is rather given to simplified cartoons of the promoter activation, transcription and translation processes than to the actual physical properties. This objection is supported by several findings. For instance, the instantaneous events assumed by simplified cartoons contrast reality, where each birth or death of a macromolecule involves several small steps and creates a memory between individual events. Accordingly, it has been shown [151] that gestation and senescence periods, which are in the case of mRNA of considerable length due to the elongation phases of transcription and translation, significantly reduce the generated noise without requiring higher molecule abundances. Further, it is known that the anabolic nature of protein synthesis, where smaller subunits are joined together, generates enormous fluctuations in the levels of charged tRNAs [234]. Noisy protein concentrations may thus be a propagated effect of noisy levels of charged tRNA, not mRNA [147]. Lastly, it has been shown [16] that the chromosomal positioning of a gene can have a *larger* influence on noise statistics than the copy number of transcribed mRNA.

Whereas the primary *origin* of intracellular randomness remains unclear [148], a unified theory exists for the *propagation* of noise through the cellular regulatory network [146]. This theory is complemented with experimental studies on noise propagation [6, 18, 76, 132, 150, 152, 182], which show that intracellular noise can be extensively manipulated, such as low–pass filtered or suppressed, by various regulatory architectures.

Propagation of noise through the cellular network can have drastic consequences on network operation. On a detailed level, for instance, molecular noise can significantly *increase* the sensitivity of intracellular regulation [149]; however, it significantly decreases the ultrasensitivity of the branch point effect [22]. Therefore, stochastic fluctuations must be carefully considered when sensitive cellular processes are investigated. On a global level, for instance, there is increasing evidence that aging may be surprisingly dependent on the effects of intracellular randomness [11, 136, 160, 170]. Another important effect of noise propagation is the phenomenon of phenotypic multistability, which is reviewed in the next section.

1.3.3 Phenotypic multistability

The phenomenon of phenotypic multistability, meaning the co–existence of multiple, distinct phenotypes within an isogenic cell population, has long been observed [137, 197]. Although the mechanism for the generation of phenotypic multistability has already been proposed as early as 1961 [134], phenotypic multistability has only recently received widespread attention in the wake of the re–emerging stochasticity field (see Section 1.3.2). Phenotypic multistability emerges when intracellular noise (see Section 1.3.2) causes different cells of an isogenic population to assume different, simultaneously available steady states, or phenotypes (see Figure 1.1 and Section 1.3.1).

Phenotypic multistability has first been discovered in the *lac* operon of *E. coli* [137]. Research on the *lac* operon has continued ever since, and the system is now well characterized through a combination of experiments and system theory analysis [144]. Another classic example of phenotypic bistability is the lysis or lysogeny of bacteriophage λ [65, 156], a virus that infects *E. coli* and, soon after infection, either reproduces quickly, thereby killing the host and releasing phage particles (lytic pathway), or integrates in the genome of the host and remains dormant (lysogenic pathway). The developmental programme that controls the type of infection has become a paradigm for developmental switches [193].

Further examples of phenotypic multistability in the microbial world abound and are by far not limited to the swarming motility [89], the sporulation [212] and the competence for DNA uptake from the environment [202] in *Bacillus subtilis*, the acquisition of antibiotic resistance [12] in *Escherichia coli*, fruiting body formation in *Myxococcus xanthus* [176], flagellin phase variation in *Salmonella enterica* [232], and galactose utilization in *Saccharomyces cerevisiae* [1].

Within metabolism, phenotypic bistability has so far been observed in the substrate utilization pathways of lactose [137, 144] and galactose [1]. As these bistable

substrate utilization pathways are of a peripheral nature, the core of central metabolism into which these pathways feed is still believed to operate homogeneously throughout an isogenic cell population. This traditional view is challenged in Chapter 4 of this thesis.

1.3.4 Computational systems biology

To investigate the aspects of intracellular complexity reviewed in Sections 1.3.1, 1.3.2 and 1.3.3, a computational systems biology approach is followed in Chapters 2, 3 and 4 of this thesis. Therefore, the here chosen computational systems biology approach and the here used types of computational models are briefly reviewed and classified next.

In general, systems biology aims at obtaining a system–level understanding grounded in molecular–level knowledge [96]. To obtain this understanding, systems biology approaches follow an iterative cycle between computational prediction and experimental testing [43]. Computational predictions are generated by the field of computational systems biology, which has two distinct and complementary branches [96]. The first branch generates hypotheses through data–mining, using statistical analyses to extract hidden patterns from huge quantities of experimental data. The second branch generates hypotheses through simulation–based analyses of in silico models. The contributions of Chapters 2, 3 and 4 of this thesis belong to this second branch of computational systems biology.

Simulation–based approaches can be used to understand how very specific molecular interactions work together to accomplish a very specific function. For instance, a comprehensive in silico model has been developed to deconvolve and quantify the contributions of multiple cellular processes on the phosphorylation level of a specific regulatory molecule [24]. However, simulation–based approaches can also be used to uncover *general design principles*, which increase our *general* understanding of cellular operation. Of particular interest for this thesis are general *regulatory motifs*, which are defined as small, repeated, perhaps evolutionarily conserved regulatory subnetworks, classifiable on the basis of function, architecture, dynamics, or biochemical process [45, 69, 164, 228]. Examples of general regulatory motifs include the regulatory motifs of *simple switches* [17, 62], *dual–time switches* [31], *oscillators* [63], *memory* [5], *bandpass frequency filters* [194], *noise filters* [180], and *noise amplifiers* [5].

Novel general regulatory motifs are proposed in Chapters 3 and 4 of this thesis. These motifs are derived with the help of computational models, which are essential to understand the complexity of biological systems [95]. Based on the chosen level of detail, computational models of cellular systems have been classified as interaction–based, constraint–based and mechanism–based [198]. This classification is used also here to place the mechanistic modeling approach of this thesis into a wider framework.

1.3.4.1 Interaction–based models

The least complex type of models are interaction–based models. These models only cover the interactions between components, not the stoichiometry or the rates of these interactions, are static, and do not contain any parameters. Examples of interaction–based models include networks of metabolic reactions [86], protein–protein interaction networks [217], and the yeast genetic interaction network [206].

Interaction–based models are primarily used to analyze network topology, meaning the web of interactions between components, using methods from graph theory [15]. The principle aim of such analyses is to uncover the design principles that underlie network topology. One such design principle is the general organization of metabolic networks into many small, highly connected topological modules that combine in a hierarchical manner into larger, less cohesive units. Within *Escherichia coli*, this hierarchical modularity closely overlaps with known metabolic functions [169]. Modularity is a welcome property for the researcher, as it increases the autonomy of cellular sub–systems and thus allows for their detailed study in isolation of the remainder of the cell [69, 140]. This important finding of interaction–based model analyses, cellular modularity, is exploited in Chapter 3 of this thesis.

Further research has revealed that cellular networks are not only composed of topological modules, but also of few *motifs*, which are small, elementary functional units of few interactions that are statistically overrepresented and conserved throughout the whole network. For instance, it has been determined [190] that the transcriptional regulation network of *Escherichia coli* harbors three distinct, general motifs. Another study [111] showed that the transcriptional regulation network of *Saccharomyces cerevisiae* contains the same three motifs, and three more. Using a more refined detection method, subsequent studies detected even more such motifs in the transcriptional regulation network of *Escherichia coli* [121, 122].

As the architecture of a network constricts its function and reflects its evolution, topological analyses have also attempted to discover general functional or evolutionary principles. Most notably, it was proposed that cellular networks are 'scale–free' [86, 220], meaning that few highly connected hubs hold many less connected components together. It was further proposed that this property might have evolved because it conferred topological, functional and dynamical robustness to the cell [15]. However, later studies [4, 23] have cast severe doubts on these findings and their interpretations. It was concluded that analyses of network topology are restricted to studying the general principles of cellular organization but fail to predict network function and evolution [198]. To overcome these limitations of interaction–based models and, for instance, further examine the modules and motifs uncovered by topological analyses, more sophisticated modeling approaches are required.

1.3.4.2 Constraint–based models

Like interaction–based models, constraint–based models are static, do not contain any parameters, and describe the topology of a network. However, in addition to topological constraints on network behavior, constraint–based models also constrain the network's behavior through considering the stoichiometry of interactions, which includes reaction directions [198]. Because the stoichiometry of metabolic reaction networks is generally well–known, constraint–based models lend themselves especially to the study of metabolic networks. Examples of contraint–based models include those of the metabolic networks of *Escherichia coli* [172], *Saccharomyces cerevisiae* [55], and the opportunistic pathogen *Staphylococcus aureus* N315 [71].

Unlike interaction–based models, constraint–based models are capable to generate predictions on network function and evolutionary optimality. For instance, constraint–based models can be used to predict the viability of knockout mutants with a success rate of approximately 80% in yeast [53]. The prediction of metabolic flux distributions is possible through flux balance analyses [211]. Further, a combined experimental and in silico study revealed that *E. coli* adaptively evolves to attain the growth capacity predicted by the constraint–based model [78]. When combined with metabolome data, constraint–based models can be used to predict active regulatory sites [105], as is done in Chapter 3 of this thesis. When combined with transcriptome data obtained from comparative microarray analyses, constraint–based models can further be used to reveal those metabolites around which the most significant transcriptional changes occur [145]. This approach was used by the author of this thesis to compare the transcriptome of a clinical *Staphylococcus aureus* N315 strain with a normal phenotype to that of its isogenic mutant with a stable small–colony–variant phenotype. This side project, which is not included in this thesis, yielded new insights into the pathogenesis of *S. aureus* [77, 188].

However, the usability of constraint–based models is limited. Because these models are static, they cannot capture dynamic processes and therefore cannot generate detailed, quantitative predictions on cellular dynamics.

1.3.4.3 Mechanism–based models

Unlike interaction–based and constraint–based models, which are static, mechanism–based models are dynamic and can be used to generate detailed, quantitative predictions on the dynamics of cellular processes [198]. By analyzing specific cellular subsystems in mechanistic detail, mechanism–based models exploit the modularity of cellular interaction networks (see Section 1.3.4.1). Examples of mechanism–based models of cellular subsystems include models of glycolysis [24, 36] and the lac operon [144] in *E. coli*, bacterial chemotaxis in *E. coli* and in *B. subtilits* [165], the target–of–rapamycin (TOR) signalling pathway in *S. cerevisiae* [104], the circadan rhythm in *Drosophila melanogaster* [209], and the

epidermal growth factor (EGF) signal–transduction cascade in humans [185].

In conjunction with experiments, mechanism–based modeling approaches can elucidate the function of known cellular interactions (for instance, see [102, 144]), reveal novel interactions and operating principles of specific systems (for instance, see [104]), and extract general design principles out of small model systems (for instance, see [144]). However, unlike interaction–based and constraint–based models, mechanism–based models usually contain many parameters whose values need to be estimated from experimental data [85]. Because the availability of measurement data is often limited [210], the parameter estimation problem is usually underdetermined and remains a major bottleneck in the development of useful mechanistic models. This problem is addressed in Chapter 2 of this thesis, where a method is developed to analyze mechanistic models with unidentifiable parameter values.

Hampered by unidentifiable parameter values and uncertain model structures, most mechanism–based approaches have so far been limited to the study of relatively small cellular subsystems [95]. Two notable exceptions are a model of the human EGF signal transduction cascade [185], which contains over 100 equations and kinetic parameters, and a model of *E. coli*'s glycolysis [24], which comprises 38 enzymatic reactions, more than 50 metabolites, and the expression of 17 enzymes. In Chapter 3 of this thesis, a novel, large–scale model of *E. coli*'s central carbon metabolism is presented. To the author's best knowledge, the presented model is the third large–scale mechanistic model of intracellular processes so far.

What sets the large–scale mechanistic model presented in this thesis apart from the other two such models is the primary focus on *general* principles rather than on investigating specific molecular phenomena: The main result of one of the two above mentioned studies is the identification of the parameter that most sensitively determines the signal efficacy of the EGF–induced response [185]; the main result of the other study is the quantification of the contributions of multiple cellular processes on the phosphorylation level of a specific regulatory molecule [24]. In contrast, the main result of Chapter 3 of this thesis is the identification of the general principle of *distributed sensing of intracellular metabolic fluxes*, which, importantly, *requires* the here investigated high level of complexity to emerge.

1.4 Complexity on the population level

On the population–level, complexity arises from the interplay of genotypic and phenotypic (sub)populations, each comprising large numbers of almost identical individuals, with each other and the environment 1.2. This type of complexity is investigated by the scientific discipline of evolutionary game theory, which is briefly reviewed in this section and applied in Chapter 5 of this thesis to investigate the population–level consequences of *responsive phenotypic diversification*, a general principle introduced in Chapter 4 of this thesis.

1.4.1 Evolutionary game theory

As a cell constantly interacts with other cells and its environment, complex interaction patterns can emerge on the population–level. Therefore, even if the molecular complexity inside a biological cell is understood through its reduction to basic design principles, it may remain unclear which network components and architectural features exist to ensure survival in a particular environment — the identified features may provide a fundamental function, lead to competition, mutualism, commensalism, or parasitism, or may simply be byproducts of evolution without significantly contributing to survival [228].

The field of evolutionary game theory [129], which is a combination of traditional game theory and population biology, aims to understand the emerging complexity on the population–level; it is becoming an increasingly popular tool [9, 47, 64, 123, 124, 219, 229] to explain observed patterns of phenotype expression and the compositional dynamics of viral and bacterial populations. In evolutionary games, different microbial strains compete with each other, using strategies whose payoffs (Darwinian fitness, or average reproductive success) depend on the strategies of other microbial strains. Evolutionary game theory predicts phenotype–expression patterns as evolutionary stable strategies in a game pitting microbe against microbe, and microbe against nature [228].

For instance, through an evolutionary game theory study combined with targeted experiments, it has been explained why certain competitive communities of bacteria coexist. Remarkably, such coexistence occurs because these communities play a *game of rock–paper–scissors*, where rock crushes scissors, scissors cuts paper, and paper covers rock [91]. This game, which is also a *poison–antidote game* [139], arises from a pairwise competition of (i) the expensive and aggressive strategy 'Produce both a toxin and its antidote' (rock), (ii) the cheaper and defensive strategy 'Produce the antidote but not the toxin' (paper), and (iii) the cheapest but risky strategy 'Produce neither toxin nor antidote' (scissors). The payoffs of these three strategies are, in terms of average reproductive success, in the microbial game equivalent to the winning rules of the children's game. This study also highlights the importance of considering localized interactions between cells. Such localized interactions are the focus of *spatial games*, which have also received much attention in recent years [124, 138].

Another remarkable, experimentally verified application of evolutionary game theory explains the steady-state coexistence of two *Saccharomyces cerevisiae*–strains in a well-mixed culture [64]. The coexistence has been shown to follow the rules of a *snowdrift game*, which arises because the performance of either of the two competing strains increases when it becomes rare.

Among the many possible playing strategies known to game theory, two are of particular importance for this thesis. These two strategies, bet–hedging and cooperation, are introduced in the following two sections.

1.4.2 Cooperation

Mutual cooperation is the globally optimal strategy of the *prisoner's dilemma game*, in which selfish behavior of the two competing parties leads to a suboptimal solution. The dilemma arises because defection of one party has potentially disastrous consequences for the 'nice but stupid' party that chose to cooperate. The prisoner's dilemma game occurs frequently in nature; in the microbial world, examples include the lowered fitness of an RNA virus at high rates of co–infection [208], the suboptimal proliferation dynamics of an *E. coli* mutant [219], and the cooperation between tumor cells [8].

When the prisoner's dilemma game is played by more than two parties, it becomes a *social goods* game [68], and the arising dilemma, which has already been conceptualized in 1833 [115], is since 1968 known as the *tragedy of the commons* [67]. The tragedy lies in the access of self–interested individuals to a common resource that can be exploited *either* rapidly *or* efficiently [112, 123]. It has been convincingly argued [67] that unless the access to the common resource is regulated through the imposition of higher–level incentives, selfish behavior of individuals depletes the resource rapidly and inefficiently, with potentially disastrous consequences for the resource, and thus for the population as a whole.

For instance, when human society is confronted with a tragedy of the commons dilemma, which occurs e.g. when the common resource 'clean air' is selfishly polluted e.g. as a consequence of cost reduction, the arising tragedy, 'depletion' of clean air, can only be averted through the imposition of laws that penalize excessive emissions. In biology, the tragedy of the commons dilemma has been applied to numerous, recently reviewed scenarios and has been classified into three types [163]:

- Type 1: A *pre–existing, extrinsic resource* over which individuals in a group or population compete. For instance, competition of phages within the host leads to a suboptimal virulence [90].

- Type 2a: A *social good* formed by cooperating individuals (*cooperative environment*). For instance, competition between genetic lineages within an individual leads to lower individual fitness [56].

- Type 2b: A *social good* formed by cooperating individuals restraining from conflict (*non-competitive environment*). For instance, restraint from plant competition for light enables all plants to allocate more resources to reproduction unless individuals forgo the noncompetitive environment created by abstaining from growing taller [225].

Additionally, biological tragedy of the commons have been classified into two classes [163]:

- Collapsing: A situation in which selfish individual behavior results in the entire resource vanishing. For instance, when Cape honey bee workers cease

to help the colony and instead invest in their own selfish reproduction, very few individuals become workers, and in turn, the colony collapses [127].

- Component: A situation in which selfish individual behavior results in a lower average fitness for the group, but the group is still able to persist on the resource in question (Type 1) or benefit to some degree from the social good (Types 2a and 2b); the resource has not disappeared completely. The above examples provided to illustrate the three types of tragedies are all component tragedies.

The tragedy of the commons dilemma is especially widespread in microbial populations [159]. However, it arises only if there are no direct benefits to restraint [67]. Therefore, apparently resolved tragedies might, upon close examination, turn out not to be tragedies in the first place [163]. For instance, sentinel behavior in meerkats has the direct personal advantage from being watchful [40]; thus, sentinel behavior is not a cooperative but a selfish bet–hedging strategy (see next section), with the sentinel betting on the presence of nearby predators. Care must therefore be taken when distinguishing cooperative from selfish behavior, as is demonstrated in Chapter 5 of this thesis through close examination of an apparent tragedy in the microbial world.

1.4.3 Bet–hedging

Facing an uncertain future, microbial populations are known to diversify into multiple phenotypes such that after unexpected environmental changes, at least one of the phenotypes will be well–adapted [41]. Such diversification has been shown, both theoretically [106, 107, 205] and experimentally [2], to be the optimal response in certain stochastic environments. Importantly, it was found that phenotypic variation is an evolvable trait, with interphenotype switching rates, like those between the two stable states of gene expression in a bistable system, being tuned to the frequency of environmental changes [2].

However, although phenotypic diversification does in these environments maximize the representation of the genotype in future generations, this strategy is selfish and as such fundamentally different from cooperation: Individuals independently place their own (selfish) bets on the uncertain future; as different individuals place different bets, the whole population effectively spreads its risks, or 'hedges its bets' [187]. In recent years, bet–hedging has received much interest as a strategy of microbes against novel and hostile environments [7, 201, 229].

One example of bet–hedging in microbes is the sporulation of *B. subtilis*. Under conditions of nutrient starvation, an isogenic *B. subtilis* population splits into two phenotypic subpopulations. One subpopulation sporulates, and the other utilizes alternative metabolites to continue growth [214]. As benefit, sporulating cells harvest resistance to various environmental conditions; these cells ensure the preservation of the clonal lineage. The two benefits of cells that delay or avoid sporulation are (i) slow proliferation on nutrients released from lysis of other cells, a

behavior termed cannibalism or fratricide [39], and (ii) the capability for rapid resumption of growth in the event of returning nutrient conditions. In contrast, cells that have sporulated are committed to a long-term process of spore formation and subsequent germination. Therefore, each of the two phenotypes is a form of specialization that increases efficiency in one area at the expense of the other [213].

The classic example of microbial bet–hedging, however, is given by persister cells [113]. Persister cells are in a stochastically acquired [12], transient, growth–impaired state that confers resistance to antibiotics. Somewhat analogous to the bet–hedging strategy of sporulation, persistence is a form of bet–hedging that ensures survival during catastrophes [107]. Importantly, although bet–hedging is primarily a strategy to anticipate unexpected changes in environmental conditions, it is also a strategy of a *social goods* game — persistence of a phenotypic subpopulation can benefit another, actively metabolizing subpopulation as persister cells do not compete for limited resources [61]. Such a conclusion also holds true for *B. subtilis* spores. In Chapter 5 of this thesis, a situation is investigated where bet–hedging has a similar social component.

1.5 Outline of this thesis

This thesis aims at advancing our understanding of the general principles surrounding and generating phenotypic diversification of isogenic cell populations. The model system used to identify such principles is the central carbon metabolism of the bacterium *Escherichia coli*. The research presented in this thesis originated from the chance observation that following a carbon source shift from glucose to acetate, certain *E. coli* mutant strains do not visibly grow for a reproducible 5 days before resuming exponential growth fairly suddenly. This and other early observations led to the substantiated hypothesis that such a prolonged 'lag time' is due to diversification of the total population into a growing and a non–growing phenotypic subpopulation. A prolonged 'lag time' would arise if the growing cells were initially so small in number that their growth, obscured by the majority of non–growing cells, needs 4–5 days before it visibly affects the size of the total cell population (see Figure 4.1). As phenotypic diversification has not yet been identified in central metabolism, its discovery in this cellular core process, along with the identification of the generating molecular mechanism and the underlying adaptation strategy on the population level, would be a significant finding.

As phenotypic multistability is an emergent property of nonlinear interactions and stochastic processes, requiring feedback loop architectures and cooperative binding events (see Section 1.3), its identification amidst the molecular complexity of central metabolism is not straight–forward. To examine and understand the dynamics arising from such cellular complexity, mechanistic models are essential [95]. However, a translation of our current knowledge of *E. coli*'s central metabolism into mechanistic models is severely hampered by uncertain values for

the many parameters appearing in such models. To be able to nonetheless construct and analyze such a model, a novel approach to deal with underdetermined biochemical models is presented in Chapter 2.

In Chapter 3, the novel approach developed in Chapter 2 is used to construct a large–scale mechanistic model of *E. coli*'s central metabolism and its regulation. This model demonstrates that from the interplay of known, local interactions in *E. coli*'s central metabolism, the system-wide adaptations of metabolic operation between glycolytic and gluconeogenic carbon sources (such as glucose and acetate) can emerge. The model further demonstrates that these adaptations are centered on a new regulatory motif, which establishes *molecular sensors for intracellular metabolic flux.*

In Chapter 4, experimental proof is provided for the diversification of an isogenic *E. coli* population into two stable phenotypic subpopulations following a carbon source shift from glucose to acetate. It is further shown that the mechanism responsible for generating the phenotypic subpopulations is a subcircuit of the regulatory circuitry modeled in Chapter 3 and is centered on a there proposed flux sensor. These results show that phenotypic diversification generalizes to central metabolism and can thus be viewed as an inherent feature of its design. Among the new, general principles derived from these results is the novel population–level adaptation strategy of *responsive diversification.*

The occurrence of multiple stable phenotypes in the here studied model system naturally raises the question *why* such phenotypic diversification has evolved. Therefore, the optimality of the population–level adaptation strategy introduced in Chapter 4, responsive diversification, is investigated in Chapter 5 with the help of evolutionary game theory.

Chapter 2

A divide–and–conquer approach to analyze underdetermined biochemical models

Oliver Kotte and Matthias Heinemann

2.1 Summary

To obtain meaningful predictions from dynamic computational models, their uncertain parameter values need to be estimated from experimental data. Due to the usually large number of parameters compared to the available measurement data, these estimation problems are often underdetermined meaning that the solution is a multidimensional space. In this case, the challenge is yet to obtain a sound system understanding despite non–identifiable parameter values, e.g. through identifying those parameters that most sensitively determine the model's behavior. Here, we

present the so–called divide–and–conquer approach — a strategy to analyze underdetermined biochemical models. The approach draws on steady state -omics measurement data and exploits a decomposition of the global estimation problem into independent subproblems. The solutions to these subproblems are joined to the complete space of global optima, which can be easily analyzed. We derive the conditions at which the decomposition occurs, outline strategies to fulfill these conditions, and — using an example model — illustrate how the approach uncovers the most important parameters and suggests targeted experiments without knowing the exact parameter values.

2.2 Introduction

Mathematical models are capable to reproduce and predict complex cellular responses, and are as such invaluable in advancing our understanding of living cells [95]. A common type of mathematical models are differential equation models, which are especially suited to investigate the dynamic behavior arising from molecular interactions. Such models often contain many parameters whose values are uncertain but affect the simulated responses [79]. Therefore, these parameters are usually either directly measured or collectively estimated from experimental data, a process which most commonly involves the maximization of the maximum likelihood, often in the form of the minimization of a least squares distance between the simulation and the data, and the proper pre- and post-estimation diagnostics [85].

Because differential equation models of biochemical systems typically contain many uncertain parameters whereas the availability of measurement data is often limited [210], the parameter estimation problem is often underdetermined and remains a major bottleneck in the development of useful models. However, recent research suggests that the knowledge of all parameter values may not be necessary to obtain good predictions. First, the model structure can tightly constrain the possible responses such that astonishingly accurate predictions are possible even without estimating the parameters [34]. Second, 'sloppiness' seems to be a universal property of systems biology models, meaning that most parameter values are unimportant because the system response is sensitively determined by the combination of only few parameter values [66]. These observations lead to a question: 'If I do not have enough measurement data to identify my parameter values, can I still obtain a sound system understanding and derive good predictions despite of my problem being underdetermined?'

In this chapter, we present an approach that is capable to achieve exactly this, given that certain conditions on the underdetermined parameter estimation problem can be fulfilled. This so–called divide–and–conquer approach exploits a division of the estimation problem (not the model itself) into many independent subproblems of smaller dimension. The decomposition yields the *complete* solution space of the underdetermined estimation problem in a structured form, which facilitates

a subsequent, systematic analysis of that solution space. The analysis can reveal the possible responses within the solution space and identify which parameters most sensitively determine these, and what effect a variation of these parameter values has on the response.

This chapter is structured as follows. First, we derive the necessary and sufficient conditions to trigger a decomposition into subproblems. Next, we show how this decomposition can be exploited through the divide–and–conquer approach, and discuss its application to real-world problems in systems biology. Then, to demonstrate the approach, we establish and analyze the complete solution space of an underdetermined model that overarches the metabolic and transcriptional regulation levels.

2.3 Method

In this section, we first present the mathematical foundation of the divide–and–conquer approach and then the method itself.

2.3.1 Mathematical foundation

The divide–and–conquer approach exploits a decomposition of the global parameter estimation problem into smaller subproblems. This decomposition occurs only when certain conditions are fulfilled. To derive these conditions, we successively specialize the general formulation of the global estimation problem to a formulation composed of independent subproblems. The conditions imposed during this specialization are the necessary and sufficient conditions to trigger the decomposition.

The general parameter estimation problem is stated as finding the set of parameters $\mathbf{p}$ within upper and lower bounds, $\mathbf{p}^U$ and $\mathbf{p}^L$, that minimizes a scalar cost function J. The cost function measures the goodness of the model prediction $\mathbf{y}(\mathbf{p}, t)$ with respect to an experimentally measured data set $\mathbf{y}_{msd}(t)$, and may include a diagonal scaling matrix $\mathbf{W}(t)$ with non–negative elements. The model prediction $\mathbf{y}$, which is calculated from the differential state variables $\mathbf{x}$ with the predictor function $\mathbf{k}$, is constrained by the system dynamics $\mathbf{f}$, which governs the time progression of $\mathbf{x}$. The problem can also include a set of parameters $\mathbf{q}$ that are not estimated. The mathematical formulation of this problem is:

Find $\mathbf{p}$ to minimize the sum of squared errors

$$J = \int_{t_0}^{t_f} (\mathbf{y}_{msd}(t) - \mathbf{y}(\mathbf{p}, t))^T \mathbf{W}(t)(\mathbf{y}_{msd}(t) - \mathbf{y}(\mathbf{p}, t))dt \qquad (2.1)$$

subject to the constraints

$$\frac{d\mathbf{x}}{dt} - \mathbf{f}(\mathbf{x}, \mathbf{p}, \mathbf{q}, t) = 0 \qquad (2.2)$$

$$\mathbf{y} - \mathbf{k}(\mathbf{x}, \mathbf{p}, \mathbf{q}, t) = 0 \qquad (2.3)$$

$$\mathbf{x}(t_0) = \mathbf{x}_0 \qquad (2.4)$$

$$\mathbf{p}^L \leq \mathbf{p} \leq \mathbf{p}^U. \qquad (2.5)$$

Mathematically, this is a nonlinear optimization problem with differential–algebraic constraints, which is commonly solved using a suitable optimizer (see Fig. 2.1a).

Before we specialize Equations 2.1–2.5 to a formulation composed of independent subproblems, we take all the summands appearing in $\mathbf{f}$ and list them in a rate vector $\mathbf{v}$. We define a stoichiometric matrix $\mathbf{S}$ such that $\mathbf{f} = \mathbf{S} \cdot \mathbf{v}$, and rewrite Eq. 2.2 to

$$\frac{d\mathbf{x}}{dt} - \mathbf{S} \cdot \mathbf{v}(\mathbf{x}, \mathbf{p}, \mathbf{q}, t) = 0. \qquad (2.6)$$

As most biological measurements are taken at discrete time points, we limit our investigation to cost functions of the form

$$J = \sum_{i=1}^{m} (\mathbf{y}_{msd}(t_i) - \mathbf{y}(\mathbf{p}, t_i))^T \mathbf{W}(t_i)(\mathbf{y}_{msd}(t_i) - \mathbf{y}(\mathbf{p}, t_i)) \qquad (2.7)$$

where t_i is the i–th of m measurement time points.

Next, we specialize the general formulation given by Equations 2.3–2.7 through imposing a condition on the measurement data set.

Condition 1: At all measurement time points t_i, the measurement data set must consist of all differential state variables x and all rates v, such that

$$\mathbf{y}_{msd}^T = \left(\mathbf{x}_{msd}^T \ \mathbf{v}_{msd}^T\right), \qquad (2.8)$$

which implies

$$\mathbf{k}^T = \left(\mathbf{x}^T \ \mathbf{v}^T\right). \qquad (2.9)$$

In a later section, we comment on how these conditions can be fulfilled in real–world problems. We continue with including a condition on the model structure.

Condition 2: The model structure must allow for an *exact* fit to *all* measurement data points, such that

$$J = \sum_{i=1}^{m} \begin{pmatrix} \mathbf{x}_{msd}(t_i) - \mathbf{x}(\mathbf{p}, t_i) \\ \mathbf{v}_{msd}(t_i) - \mathbf{v}(\mathbf{p}, t_i) \end{pmatrix}^T \mathbf{W}(t_i) \begin{pmatrix} \mathbf{x}_{msd}(t_i) - \mathbf{x}(\mathbf{p}, t_i) \\ \mathbf{v}_{msd}(t_i) - \mathbf{v}(\mathbf{p}, t_i) \end{pmatrix} = 0. \qquad (2.10)$$

As the sum–of–squares J is strictly non–negative, a parameter set leading to $J = 0$ must be a global optimum. In practice, this condition requires an underdetermined estimation problem.
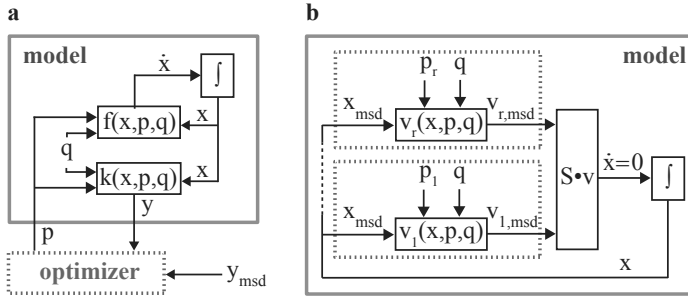
Figure 2.1: Comparison of the divide–and–conquer approach with the conventional parameter estimation strategy. **a**, the conventional strategy. The optimizer (dotted box) uses the output y of a model simulation to optimize the parameters p with respect to the experimental data y_{msd}. **b**, the divide–and–conquer approach. Given certain conditions, the estimation problem decomposes into multiple independent subproblems (dotted boxes), for which the complete analytical solution spaces can be derived by solving a system of algebraic equations.

With these conditions fulfilled, the global estimation problem reduces to finding a solution $\mathbf{p}$, $\mathbf{p}^L \leq \mathbf{p} \leq \mathbf{p}^U$, that satisfies

$$v_j\left(\mathbf{x}_{msd}(t_i), \mathbf{p}, \mathbf{q}, t_i\right) - v_{msd,j}(t_i) = 0 \tag{2.11}$$

for $i = 1 \ldots m$ and $j = 1 \ldots r$, where r is the number of components in $\mathbf{v}$. Thus, Eq. 2.11 comprises $m \cdot r$ equations. As $\mathbf{p}$ is the only unknown, this in practice underdetermined equation system can be solved to derive the complete solution space of $\mathbf{p}$. Because these equations are coupled solely through $\mathbf{p}$, this potentially very large–dimensional solution space can be decomposed into many smaller–dimensional subspaces by removing the coupling through a further condition on the model structure.

Condition 3: The parameters to be estimated, $\mathbf{p}$, consist of $1 \leq s \leq r$ disjunct sets

$$\mathbf{p}^T = (\mathbf{p}_1^T \ \mathbf{p}_2^T \ \ldots \ \mathbf{p}_s^T) \tag{2.12}$$

such that each set $(\mathbf{p}_k, \mathbf{q})$ fully parameterizes a subset of the rate equations. If disjunct subsets do not exist ($s = 1$), then the complete solution space can be derived through Eq. 2.11 but not decomposed into smaller–dimensional subspaces. If each p_j appears in only one rate equation ($s = r$), then the solution space can be maximally decomposed into pairwise independent subspaces.

The parameter estimation problem is thus reduced to finding $\mathbf{p}_k$ such that

$$v_j\left(\mathbf{x}_{msd}(t_i), \mathbf{p}_k, \mathbf{q}, t_i\right) - v_{msd,j}(t_i) = 0 \tag{2.13}$$

for $i = 1 \ldots m$, $j = 1 \ldots r$ and $k = 1 \ldots s$.

As Eq. 2.13 comprises $m \cdot r$ algebraic equations in s decoupled sets, the global problem has been successfully decomposed into independent subproblems of smaller dimension. Eq. 2.13 states that a parameter set $\mathbf{p}$ is a global optimum if all its subsets $\mathbf{p}_k$ parameterize the rate equations v_j in which they appear such that the measured rates are reproduced exactly for the measured states, as illustrated in Fig. 2.1b. Note that a verbose description of this section can be found in Section 2.6.1.

2.3.2 The divide–and–conquer approach

The divide–and–conquer approach is the consistent exploitation of the decomposition of the global estimation problem into independent subproblems. This approach encompasses both the derivation of the complete solution space of an underdetermined problem, and the efficient analysis of that space.

The derivation of the complete solution space can be structured into three steps. In the first step, complete sets of state and rate data are obtained to fulfill Condition 1. In the second step, the degree of decomposition is chosen according to Condition 3. In the third step, the complete solution space of the underdetermined problem is derived by fulfilling Condition 2. Here, it is important to understand that the divide–and–conquer approach does not give the solution to the estimation problem as upper and lower bounds on single parameters values, but as an analytic description of the parameter subspaces that together form the complete solution space.

Once the complete solution space is analytically known, it can be quickly and systematically analyzed to obtain a sound understanding of the possible system responses within this parameter space. It is also important to note that the divide–and–conquer approach first neglects the noisy nature of biochemical measurements, then derives insights based on the assumed 'perfect' data set, and finally assesses the robustness of the derived insights with respect to data noise.

The following three sections discuss the derivation of the complete solution space in detail. After that, we illustrate the systematic analysis of that solution space with an example model, and show how the robustness of the obtained insights with respect to data noise can be assessed.

2.3.2.1 Step 1: Obtaining complete steady state data sets

Recent developments in high–throughput experimental methods provide us with ever more comprehensive measurement data sets in steady state conditions, e.g. of the cell's proteome and metabolome [81]. However, such data sets, whether assembled from literature and/or own measurements, are to date often incomplete. To fulfill Condition 1, which requires complete data sets, we therefore propose to extend incomplete measurement data sets to larger sets of *observed* data. Depending on the problem, these data sets of observables can be complete and therefore applicable for the subsequent parameter estimation, as previously suggested [60].

To obtain a complete set of observables, the unmeasured data can be inferred from the measured data with the help of models, simple or sophisticated, that are based on biological knowledge. For instance, a computational model can be used to observe metabolic reaction rates from measured ^{13}C–labeling patterns of amino acids. Similarly, an incomplete set of measured metabolite concentrations can be extended to a complete set of observed metabolite concentrations by using network–embedded thermodynamic (NET) analysis [105], which is capable to infer unmeasured metabolite concentrations within certain limits.

To observe missing rates, the consistency condition

$$dx/dt = v_+ - v_- = 0, \tag{2.14}$$

can be exploited, which states that in steady state, the sum of all compound production rates v_+ must equal the sum of all compound consumption, dilution, and degradation rates, v_-. If one of the rates in v_- or v_+ is unknown, then Eq. 2.14 can be applied to determine the missing rate from the known rates. This equation underlies flux balance analysis, which is capable to observe metabolic reaction rates from physiological data using a stoichiometric metabolic network model. On a smaller scale, Eq. 2.14 can be used to, for instance, observe a steady state protein production rate from known protein degradation and dilution rates.

If more than one of the rates in v_- or v_+ is unknown, then some of the missing rates can be observed with simple linear models. For instance, a compound dilution rate can be observed through $v_{dil,x} = \mu \cdot x$ with x and the growth rate μ known, or a compound degradation rate through $v_{degr,x} = k_{degr} \cdot x$ with x and the degradation rate constant k_{degr} known. Note that if these simple linear models contain parameters of the dynamic computational model (e.g. μ and k_{degr}), then these parameter values are already fixed and may not appear in the estimation problem of the divide–and–conquer approach (i.e. these parameters are included in **q**, not **p**).

In the unlikely case that all of the rates in Eq. 2.14 are measured, then these measurements will most likely not add up to zero due to measurement errors, implying that the model cannot reproduce the steady state *exactly* (Condition 2 is violated). Therefore, these measurement errors must be 'corrected', e.g. by minimally adjusting the data to fulfill Eq. 2.14. Note that at a later stage, it can be assessed if such data 'correction' and the neglected measurement noise sensitively affect the insights obtained with the assumed 'perfect' data set.

Lastly, data sets such as transcriptome data are often acquired only as relative measures. Fortunately, relative data of a compound concentration x is sufficient if in the model x appears always paired with a multiplicative parameter p. Then, rate equations of x are of the form $r = f(p \cdot x)$, and the parameter estimation of p can correct for an arbitrarily chosen absolute concentration of x. Such situations occur, for instance, with x as an enzyme or mRNA concentration, p as the respective rate constants, and r as metabolic reaction rate or translation rate, respectively.

2.3.2.2 Step 2: Choosing the degree of decomposition

To increase the degree of decomposition, and thereby the transparency of the later solution space to the modeler, the parameters are divided into three disjunct sets, $\mathbf{p_A}$, $\mathbf{p_B}$, and $\mathbf{q}$, such that $\mathbf{p_A}$ contains those parameters that appear in only one equation v_j, $\mathbf{p_B}$ contains those that appear in more than one equation v_j, and $\mathbf{q}$ contains the parameters that are not subject to the estimation. In biochemical models, parameters typically have a specific mechanistic meaning and as such tend to appear in only one equation v_j. Therefore, $\mathbf{p_B}$ usually contains only few parameters but is not necessarily empty.

The maximal degree of decomposition ($s = r$) is reached when $\mathbf{p_B} = \emptyset$. If $\mathbf{p_B} \neq \emptyset$, then the degree of decomposition can be increased by excluding a parameter $p \in \mathbf{p_B}$ from the estimation, which moves this parameter from $\mathbf{p_B}$ into $\mathbf{q}$. For instance, if p is known to be poorly identifiable, which is probable as p appears in multiple underdetermined equations, then assuming a literature value for p may be the better alternative anyway. If this is not justified, then all equations containing p form one subproblem.

2.3.2.3 Step 3: Determining the complete solution space

To obtain the complete parameter space that reproduces the data exactly, the complete data sets from step 1 are plugged into the kinetic model equations. These equations thus become algebraic functions of the parameters. Due to the decomposition of the parameter space performed in step 2, this set of functions consists of independent subsets. Each of these independent subsets comprises a system of equations with α_j unknowns (the number of parameters to be estimated in the subset) and β_j constraints (the number of equations derived from plugging the data into the subset's kinetic model equations). The estimation problem is thus decomposed into independent subproblems, which correspond to solving each subset's system of algebraic equations. These subproblems can be either overdetermined, exactly determined, or underdetermined:

- If $\beta_j > \alpha_j$, then the j-th subproblem is overdetermined, which prohibits an exact fit to the data and therefore the application of the divide–and–conquer approach. However, this case is unlikely to occur, as complete steady state data sets are usually obtained for only few conditions, whereas rate equations typically contain multiple uncertain parameters. An exception to this rule are first-order kinetics, which usually either approximate higher–dimensional kinetics and can be substituted by those, or do not contain any parameters to be estimated, such as a linear dilution rate equation with known growth rate μ.

- If $\beta_j = \alpha_j$, then the j-th subproblem is exactly determined. If a solution exists, it is unique — the data constrains the α_j-dimensional parameter space to a single point.

- If $\beta_j < \alpha_j$, then the j-th subproblem is underdetermined. If a solution exists, the data reduces the α_j–dimensional parameter space to a solution space of dimension $\alpha_j - \beta_j$.

If for any subproblem j, an exact fit to the data cannot be achieved and the subproblem is not overdetermined, then a discrepancy between the model structure and the available data has been identified and localized. The discrepancy can be removed either by changing the model structure, e.g. to a rate law that reproduces the data, or, if there is reason to doubt the quality of the data, by resorting to another set of measurements.

When all discrepancies between the model structure and the data are removed, then Condition 2 is fulfilled. The solutions to the subproblems are then joined to the global solution Ω of the parent parameter estimation problem. If all subproblems are exactly determined, the solution is a single point. In most cases, however, at least one of the subproblems is underdetermined, and the solution is therefore a multidimensional space. A significant advantage of the divide–and–conquer approach is that it yields the *complete* solution space of the parameters in the form of *analytically* known manifolds on which all global solutions are located. Using an example, we next illustrate how this analytically known solution space can be efficiently and thoroughly analyzed to derive a sound system understanding.

2.4 Example

We illustrate the application of the divide–and–conquer approach by deriving and analyzing the complete solution space of the small model system depicted in Fig. 2.2a. This model describes a core section of *E. coli*'s central metabolism and covers allosteric and transcriptional regulation. It simulates the reversal of carbon flow through the Emden–Meyerhoff–Pathway, which is required to switch between growth on glucolytic and gluconeogenic substrates, e.g. glucose and acetate.

The model consists of five enzymes (E_i), one transcription factor (Cra), four genes (e_i), and two metabolites (phosphoenylpyruvate, PEP, and fructose-bisphosphate, FBP). It contains 16 rates: 4 enzyme production rates, 6 compound dilution and degradation rates, 5 metabolic reaction rates, and 1 transcription factor–metabolite binding rate (of Cra to FBP). It further includes a simplified representation of the phosphotransferase system (PTS), which couples the uptake and phosphorylation of extracellular glucose to the conversion of PEP to pyruvate. The model is centered on the transcription factor Cra, whose activity controls the expression of four of the modeled enzymes and is itself controlled by the metabolite FBP. For the model equations 2.15–2.24, refer to Section 2.6.2.

2.4.1 Derivation of the complete solution space

To estimate the parameters of the kinetic equations and to analyze the complete solution space with the divide–and–conquer approach, we apply the steps presented
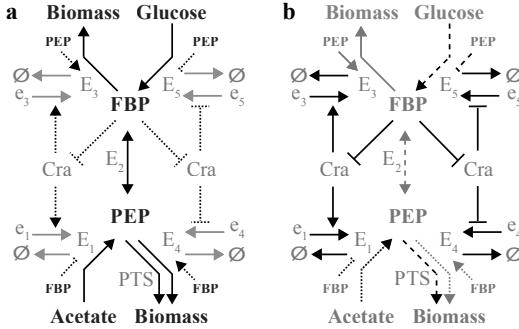
Figure 2.2: The example model. **a**, The topology of the example model. Metabolites and metabolic reactions are black; genes (e_i), proteins (Cra, E_i), and protein production and degradation rates are gray; regulatory interactions are dotted. **b**, Decomposition of the kinetic equations into six independent subproblems. Subproblem 1 (according to Table 2.1), regular black lines. 2, dotted black lines. 3, dashed black lines. 4, regular gray lines. 5, dotted gray lines. 6, dashed gray line.

in the previous section.

Step 1 is fulfilled with the complete state and rate measurement data sets listed in Table 2.2, which were obtained from literature for balanced growth on either the glucolytic substrate glucose or the gluconeogenic substrate acetate.

In step 2, the degree of decomposition is chosen. Overall, the system comprises 39 parameters. Of these, the growth rate μ and the concentrations of the carbon sources *Glucose* and *Acetate* are directly measured, literature values are assumed for ρ and k_{degr}, and n_{E_i} (the number of subunits in the quaternary structure of an enzyme) is set to 4 for all tetrameric enzymes. Therefore, $\mathbf{q} = [\mu, Glucose, Acetate, \rho, k_{degr}, n_{E_1}, n_{E_3}, n_{E_4}, n_{E_5}]$. Of the 30 parameters in $\mathbf{p}$, only $K_{Cra,FBP}$ and n_{Cra}, which describe the binding of FBP to Cra, appear in more than one rate equation. Therefore, all rate equations containing these two parameters are merged to a composite subproblem. Table 2.1 and Fig. 2.2b summarize the resulting six independent subproblems into which the global estimation problem decomposes.

In step 3, the parameters of the six subproblems are constrained by a system of algebraic equations. Each of these equations reduces a subproblem's degree of freedom by one, and can be rearranged such that one of the subproblem's parameters becomes dependent on the others. This process is described in detail in Section 2.6.4, with Table 2.1 summarizing the resulting (arbitrary) division into free and dependent parameters. The complete space of global solutions Ω comprises all parameter vectors within admissible bounds $\mathbf{p}^L \leq \mathbf{p} \leq \mathbf{p}^U$ that solve the obtained system of algebraic equations (Equations 2.27–2.33), i.e. are located on the solution manifolds. Because of the division into free and dependent parameters, a global solution can be easily generated by choosing a combination of free param-

Table 2.1: Decomposition of the global estimation problem into six independent subproblems, and the division of the parameters into free and dependent parameters. α, number of parameters. β, number of constraints. $\alpha - \beta$, degrees of freedom.

Sub-problem	α	β	$\alpha - \beta$	Free parameters	Dependent parameters
1	10	8	2	$K_{Cra,FBP}$ n_{Cra}	$v_{e_1,max}$ $v_{e_3,max}$ $v_{e_4,max}$ $v_{e_5,max}$ K_{e_1,Cra_A} K_{e_3,Cra_A} K_{e_4,Cra_A} K_{e_5,Cra_A}
2	4	1	3	L_{E_1} $K_{E_1,PEP}$ $K_{E_1,Acetate}$	k_{cat,E_1}
3	4	1	3	L_{E_5} $K_{E_5,FBP}$ $K_{E_5,Glucose}$	k_{cat,E_5}
4	4	2	2	L_{E_3} $K_{E_3,FBP}$	k_{cat,E_3} $K_{E_3,PEP}$
5	4	2	2	L_{E_4} $K_{E_4,PEP}$	k_{cat,E_4} $K_{E_4,FBP}$
6	4	2	2	$K_{E_2,PEP}$ $K_{E_2,FBP}$	$v_{E_2,f}$ $v_{E_2,r}$

eter values and calculating the dependent parameters with Equations 2.27–2.33. Note that the identification of these parameter dependencies is an active research area by itself [74, 114].

2.4.2 Analysis of the model behavior

To analyze the model behavior, we exploit the decomposition of the solution space into independent and analytically known manifolds.

First, we verify that the solution space was correctly determined and that the two measured steady states exist and are stable. To do so, we randomly generated 1,000 global solutions within admissible parameter bounds ($0.1 \leq n_{Cra} \leq 4$, $0.1 \leq K_i \leq 10$, $1 \leq L_i \leq 10^7$) by assigning random values to the free parameters and calculating the dependent parameters with Equations 2.27–2.33. We then simulated the model with initial conditions equal to the two measured conditions. As expected, we obtained perfectly level lines for both conditions and all compound concentrations (not shown). Therefore, as all of the sampled parameter vectors reproduce the steady state measurement data exactly, the solution space has been correctly determined. Furthermore, because the simulations remain in the steady states indefinitely, we can conclude that both steady states exist and are stable (at least for these 1,000 samples).

Next, we obtain a general overview of the possible system responses to a se-

quence of perturbations. This is necessary because although all global solutions reproduce the *stationary* measurement data exactly, different parameter combinations may lead to very distinct *dynamic* responses. As our example model describes the reversal of carbon flow through a metabolic pathway, we are interested in a complete picture of the possible dynamics during such flux reversals. Therefore, we chose to perturb the system by switching the carbon source from acetate to glucose at $t = 0$ h, and back to acetate at $t = 35$ h.
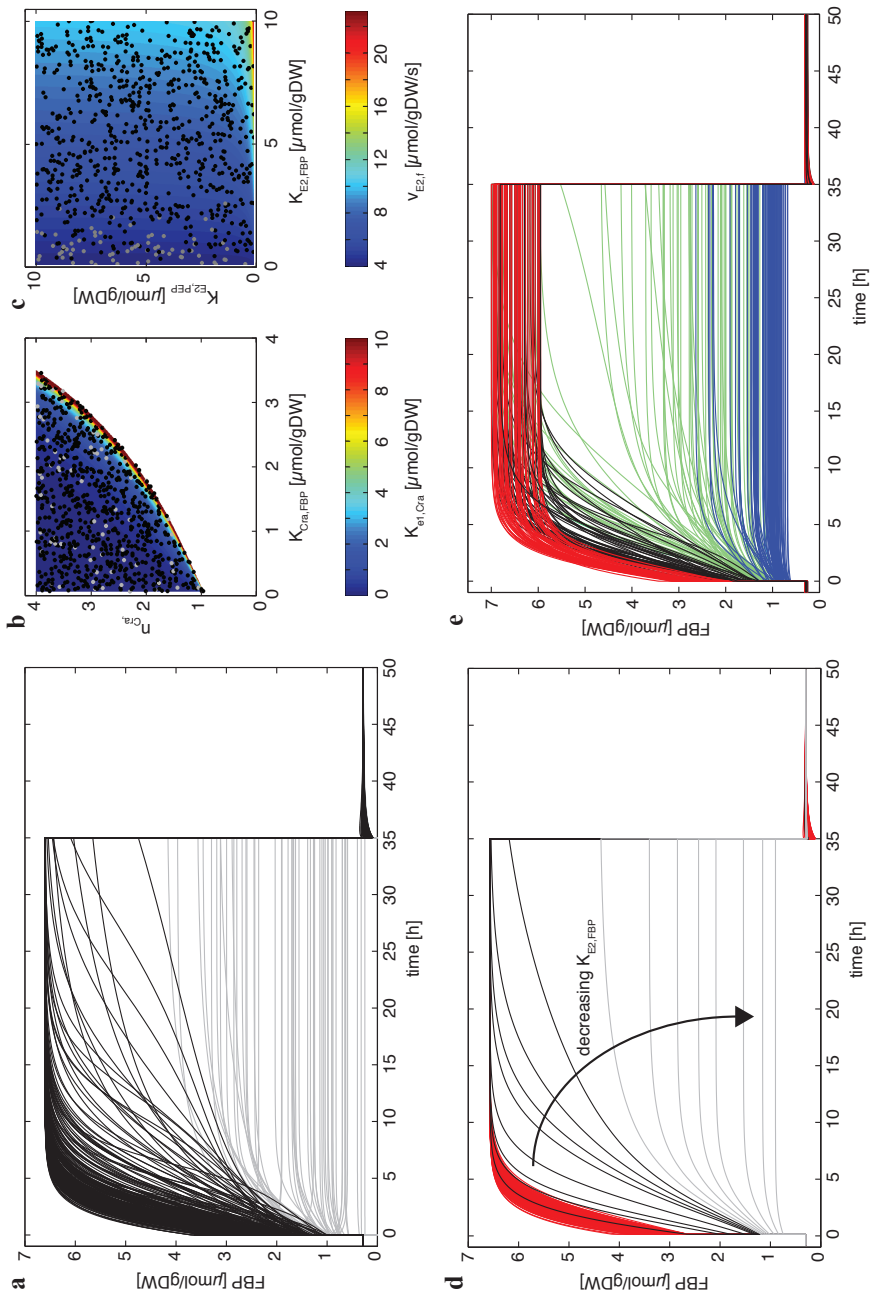
Figure 2.3a shows the simulated responses of the FBP concentration to these perturbations (with the previously sampled parameters). While all simulations successfully adapt from glucose to acetate, the dynamic behavior of the adaptation from acetate to glucose varies widely and can be categorized into two response families: Responses of family A converge to the measured steady state on glucose at 6.6 μmol/gDW, and responses of family B converge to a second steady state on glucose with a parameter–dependent concentration below 4.5 μmol/gDW. Therefore, when adapting from the measured steady state on acetate, the measured steady state on glucose — although it has been verified to exist and be stable — is only attractive for the parameter subset belonging to family A. In this context, note that the existence of the second steady state on glucose was identified by a sampling strategy and could have remained unnoticed if merely a point solution had been determined.

Next, we exploit the division into two response families to identify those parameters that most sensitively shape the dynamic response. If a parameter sensitively shapes the response, its value should determine the response family. We therefore compared the parameters' distributions in the two response families using

---

Figure 2.3 (preceding page): Analysis of the solution space. **a,** Simulated responses of the FBP concentration to carbon source shifts from acetate to glucose (at 0 h) and back to acetate (at 35 h). A random sampling of 1,000 parameter vectors $\in \Omega$ reveals that the possible system responses vary widely. The responses can be classified into two response families: 63% of the simulations converge (on different trajectories) to the measured steady state on glucose at 6.6 μmol/gDW (black lines, family A). The remaining 37% converge (on different trajectories) to a second steady state at a parameter–dependent concentration below 4.5 μmol/gDW (gray lines, family B). **b,** The manifold from subproblem 1 shows a dependent parameter as a function of the two free parameters. Black dots denote the sampled parameters that lead to responses in family A, whereas gray dots denote those that belong to family B. The locations of both the black and gray dots are evenly distributed across the entire manifold. **c,** On this manifold from subproblem 6, the gray dots cluster in a region with low values of both $K_{E_2,FBP}$ and $v_{E_2,f}$. **d,** The system response is sensitively determined by the value of a single free parameter, $K_{E_2,FBP}$. A randomization of all free parameters except $K_{E_2,FBP} = 10$ only marginally affects the trajectories (1,000 red lines), whereas a stepwise reduction of only this parameter ($K_{E_2,FBP} = 10; 4; 3; 2; 1.5; 1.4; 1.3; 1.2; 1.1; 1.0; 0.9; 0.8; 0.7; 0.5; 0.3; 0.1$) with all other free parameters constant sensitively shapes the response. **e,** The value of the free parameter $K_{E_2,FBP}$ remains the decisive factor in determining the shape of the trajectories even in the presence of 10% measurement noise ($K_{E_2,FBP}=10; 3; 1; 0.1$ for the red, black, green and blue curve sets, respectively).

the student's t–test. The parameters' p–values (see Table 2.3) span many orders of magnitude, with $K_{E_2,FBP}$ and $v_{E_2,f}$ exhibiting extremely low p–values. Therefore, the distribution of these two parameters is significantly different between the two response families. Using the derived solution manifolds, this result can be graphically illustrated. In most cases, as in Fig. 2.3b, the parameter values of either response family are evenly distributed across the manifold. However, in the case of $K_{E_2,FBP}$ and $v_{E_2,f}$ (Fig. 2.3c and Fig. 2.4), the parameter values of response family B are clustered in a particular region of the manifold. We therefore suspect that these two parameters dominantly shape the response.

To test if $K_{E_2,FBP}$ and $v_{E_2,f}$ indeed sensitively determine the system response, we first set the free parameter $K_{E_2,FBP}$ to its maximal admissible value and randomized all other free parameters as before. By fixing only this free parameter, we were able to constrain the possible responses tightly: All trajectories rapidly converge to the measured steady state on glucose, i.e. belong to family A (Fig. 2.3d). We then arbitrarily selected one of these trajectories and kept its parameters constant with the exception of $K_{E_2,FBP}$, which we decreased stepwise across its entire admissible range. By varying only this free parameter (and the two dependent parameters $v_{E_2,f}$ and $v_{E_2,r}$ with it), i.e. by moving the parameter vector in the direction of the negative $K_{E_2,FBP}$–axis of the solution manifold shown in Fig. 2.3c, we were able to move the trajectory across the entire range of the possible responses (Fig. 2.3d). Thus, in addition to having identified the two most important parameters, we also understand how their variation affects the system response.

Next, we assessed if these two important parameters retain their dominant role in determining the response in the presence of 10% measurement noise. We generated four sets of trajectories with $K_{E_2,FBP}$ (and thereby $v_{E_2,f}$) at different levels and all other free parameters and the measurement data randomized. Fig. 2.3e shows that despite of these sources of variation, $K_{E_2,FBP}$ still largely determines the response: Only for the green curve set with $K_{E_2,FBP} = 1$, which is in the transition region between the response families A and B (see Fig. 2.3d), do these sources of variation have a considerable impact on the trajectories. Therefore, the obtained understanding of how the system response is dominantly shaped by $K_{E_2,FBP}$ (and $v_{E_2,f}$) is reasonably robust with respect to measurement noise. Note that instead of assuming a flat noise magnitude of e.g. 10%, more detailed information about the uncertainties of individual data points can be used, if available.

To conclude, by exploiting the solution manifolds derived with the divide–and–conquer approach, we were capable to obtain a profound system understanding even though the parameter values were not identifiable due to limited and noisy measurement data. In general, the discovery of the most important parameter values suggests targeted experiments to measure these values and may already provide a valuable insight by itself. Before drawing biological conclusions from this particular example system, however, its predictive power should be tested, or alternatively, it should be ensured that the observed effect extends to other model variants and is thus not specific to the chosen model structure, which is merely one

among many possible mathematical representations of the available biochemical knowledge. During this process, which has already been demonstrated by previous studies [102, 104], the divide–and–conquer approach can be repeatedly applied.

2.5 Discussion

In this chapter, we presented the divide–and–conquer approach for the analysis of underdetermined biochemical models. This approach exploits a 'trivial point' at which the complete solution space of the global parameter estimation problem can be derived analytically. Using an example system, we have demonstrated how the complete solution space can be derived and subsequently analyzed. This strategy resulted in a sound system understanding and the identification of targeted experiments.

The main difficulty in applying this approach is to move a real–world estimation problem onto that 'trivial point', i.e. to fulfill Conditions 1 and 2. This can be achieved by various means. First, an incomplete measurement data set can be extended to a complete set of observables by incorporating additional biological knowledge. Second, the measurement noise can be initially neglected and the robustness of the derived insights with respect to measurement noise assessed at a later stage. Third, additional, possibly time–course measurement data that does not belong to a complete steady state data set can also be included in the divide–and–conquer approach. This can be achieved by providing a global optimizer with the derived equality and inequality constraints on the parameters (Equations 2.27–2.33). Then, the optimizer can determine that parameter combination on the solution manifolds which best reproduces the additional measurements *in addition to* exactly reproducing the complete steady state data sets.

To enable analytical solution spaces of large–dimensional parameter estimation problems, the divide–and–conquer approach decomposes the whole solution space via Condition 3 into independent subspaces, for which analytical solutions are feasible. This decomposition occurs automatically when few parameters, whose number in our experience increases only slightly with model size, are fixed at literature values and thereby excluded from the anyway underdetermined parameter estimation problem. The proposed analytical approach is thus well scalable to the often large sizes of realistic models.

Although this approach is not confined to any specific type of model, it is best suited for application areas where models are typically underdetermined, yet – omics data sets are available. Due to the many parameters of enzyme kinetics and the availability of metabolomics and fluxomics data, this approach is especially suited for models of metabolism.

The key advantage of the divide–and–conquer approach is that the global solution space of the parameters can be structured in *manageable subspaces*, which are known *completely* and *analytically*. This greatly facilitates the analysis of the possible system responses within the solution space. Of particular interest is the

identification of those few [66] parameter combinations that most sensitively shape the system response — in fact, this task is one of the major problems raised in systems biology [96]. Therefore, by focusing directly on the system responses and not on the parameter values, the divide–and–conquer approach is a practical strategy to extract valuable insights from underdetermined biochemical models.

2.6 Method details

2.6.1 Verbose derivation of the conditions triggering decomposition

The derivation of Conditions 1–3, i.e. Equations 2.8, 2.10 and 2.12, is fairly condensed in Section 2.3.1. Therefore, we here provide a more verbose derivation of these conditions, which when fulfilled decompose the global parameter estimation problem into smaller subproblems.

In Section 2.3.1, we depart from a general formulation of the global estimation problem, presented in Equations 2.1–2.5, and successively specialize this general formulation to a formulation composed of independent subproblems, Eq. 2.13. The conditions imposed during this specialization are Conditions 1–3, the necessary and sufficient conditions to trigger the decomposition.

The first thing done in Section 2.3.1 is the rewriting of Eq. 2.2 to Eq. 2.6 because the latter equation explicitly contains the observable rates v. Then, to get rid of the integral in Eq. 2.1, we take advantage of the discrete nature of biochemical measurements and replace the integral with a sum over the measurement time points, resulting in Eq. 2.7. The general formulation of the global estimation problem that we next specialize is thus given by Equations 2.3–2.7.

The first condition we impose to specialize the general formulation is to demand that at all measurement time points t_i, the measurement data set must consist of all differential state variables x and all rates v. Therefore, $\mathbf{y}_{msd}^T = \begin{pmatrix} \mathbf{x}_{msd}^T & \mathbf{v}_{msd}^T \end{pmatrix}$ (Eq. 2.8). To be able to calculate the difference between measurement and model prediction, $\mathbf{y}_{msd}(t_i) - \mathbf{y}(t_i)$, which appears in the cost function (Eq. 2.7), the model must also predict all differential state variables x and all rates v at all measurement time points t_i. Therefore, the predictor function must be $\mathbf{k}^T = \begin{pmatrix} \mathbf{x}^T & \mathbf{v}^T \end{pmatrix}$ (Eq. 2.9). Because of Eq. 2.3, this predictor function leads to the required model prediction $\mathbf{y} = \begin{pmatrix} \mathbf{x}^T & \mathbf{v}^T \end{pmatrix}$.

Next, we plug Equations 2.8 and 2.9 into the cost function, Eq. 2.7. Therefore, in Eq. 2.7, $\mathbf{y}_{msd}$ becomes $\begin{pmatrix} \mathbf{x}_{msd}^T & \mathbf{v}_{msd}^T \end{pmatrix}^T$ and $\mathbf{y}$ becomes $\begin{pmatrix} \mathbf{x}^T & \mathbf{v}^T \end{pmatrix}^T$. This formulation of the cost function J appears in Eq. 2.10. We next make the argument that $J = 0$ must be a global optimum because the cost function J is by definition non–negative. $J \geq 0$ must hold because J is a sum–of–squares, which is obviously non–negative, weighted by the scaling matrix $\mathbf{W}$, which cannot introduce a sign change due to being diagonal with non–negative elements. For $J = 0$ to occur, the model structure must allow for an exact fit to the data, which in turn essentially requires an underdetermined problem. The second condition we impose is to demand that $J = 0$ does indeed occur, as stated in Eq. 2.10.

We next exploit the fact that a sum–of–squares can only vanish when all the individual summands vanish. Because of this, Eq. 2.10 can only hold if $\mathbf{x}_{msd}(t_i) = \mathbf{x}(t_i)$ and $\mathbf{v}_{msd}(t_i) = \mathbf{v}(t_i)$ for all t_i, $i = 1 \ldots m$. From Eq. 2.6, we also know that $\mathbf{v}$ is a function of $(\mathbf{x}, \mathbf{p}, \mathbf{q}, t)$. Therefore, we know that $\mathbf{v}(\mathbf{x}, \mathbf{p}, \mathbf{q}, t_i) = \mathbf{v}(t_i)_{msd}$. Because of $\mathbf{x}_{msd}(t_i) = \mathbf{x}(t_i)$, as stated above, we arrive at $\mathbf{v}(\mathbf{x}_{msd}(t_i), \mathbf{p}, \mathbf{q}, t_i) = \mathbf{v}_{msd}(t_i)$, which is the vector formulation of Eq. 2.11.

With r the number of rates or components in $\mathbf{v}$, Eq. 2.11 comprises $m \cdot r$ equations. The only unknown in Eq. 2.11 is $\mathbf{p}$. Therefore, Eq. 2.11 is a system of $m \cdot r$ algebraic equations, which is in practice underdetermined and can be solved to obtain the complete solution space of $\mathbf{p}$. However, depending on the model size, the derived solution space can be very large–dimensional, and therefore nontransparent to the modeler. To understand the solution space better, it would be of tremendous help if the solution space consisted of pairwise independent, smaller–dimensional subspaces. To trigger such a decomposition of the solution space, we next exploit the fact that parameters in biochemical models tend to have a specific mechanistic meaning and as such appear only in one rate equation. We thus impose the third condition (Eq. 2.12), which demands that the parameter set consists of disjunct subsets such that each of these subsets fully parameterizes a subset of the rate equations. The desired extreme of this condition is that each parameter indeed appears in only one rate equation, which leads to a maximal decomposition of the solution space. The undesired extreme is that such disjunct sets do not exist at all, which does not allow for a decomposition of the solution space.

Because of the decomposition of the parameter vector $\mathbf{p}$, which is the sole unknown in the equation system of Eq. 2.11, this equation system is also decomposed into s pairwise independent subsets (Eq. 2.13). Therefore, Eq. 2.13 is a formulation of the global estimation problem that consists of s independent subproblems in the form of algebraic equation systems, which can be solved independently to obtain the complete solution spaces of the parameter vectors $\mathbf{p}_k$ associated with the subproblems. The aggregate of these solution subspaces is the complete space of global solutions to the parameter estimation problem of Equations 2.3–2.7, given that the three imposed conditions hold.

2.6.2 Model equations

This section presents the model equations of the example system illustrated in Figure 2.2a.

The system comprises the six dynamic state variables

$$\mathbf{x} = [E_1 \ \ E_3 \ \ E_4 \ \ E_5 \ \ PEP \ \ FBP]^T \ , \tag{2.15}$$

which model the four enzyme and the two metabolite concentrations.

These compounds are interconverted by the 15 rates

$$\begin{aligned} \mathbf{v} = [&v_{ex,e_1} \ \ v_{ex,e_3} \ \ v_{ex,e_4} \ \ v_{ex,e_5} \ \ v_{d,E_1} \ \ v_{d,E_3} \ \ v_{d,E_4} \ \ v_{d,E_5} \\ &v_{d,PEP} \ \ v_{d,FBP} \ \ v_{r,E_1} \ \ v_{r,E_2} \ \ v_{r,E_3} \ \ v_{r,E_4} \ \ v_{r,E_5}]^T \ , \end{aligned} \tag{2.16}$$

where the subscripts ex denote expression rates, d denote combined dilution and degradation rates, r denote metabolic reaction rates, and e_i denote the genes encoding for the modeled enzymes.

Note that we have simplified the binding process of FBP and Cra such that the binding state of Cra directly tracks the concentration of its effector metabolite FBP. Therefore,

$$Cra_A = \frac{Cra_T}{1 + \left(\frac{FBP}{K_{Cra,FBP}}\right)^{n_{Cra}}} , \tag{2.17}$$

where Cra_A is the active concentration of Cra that is not bound to FBP, Cra_T is the total Cra concentration and $K_{Cra,FBP}$ the FBP concentration required for half–saturation of Cra with FBP.

With these definitions, the time progression of these state variables is

$$\begin{aligned}
\frac{dE_i}{dt} &= v_{ex,e_i} - v_{d,E_i} \\
\frac{dPEP}{dt} &= v_{r,E_1} + v_{r,E_2} - v_{r,E_4} - v_{r,E_5} - v_{d,PEP} \\
\frac{dFBP}{dt} &= -0.5 \cdot v_{r,E_2} + v_{r,E_5} - v_{r,E_3} - v_{d,FBP} .
\end{aligned} \tag{2.18}$$

In the remainder of this section, we present the algebraic equations for all the rates v_i appearing in Eq. 2.18.

If the transcription factor Cra acts as activator on the expression of an enzyme E from its gene e, the expression rate is given

$$v_{ex,e} = \rho\, \mu\, v_{e,max} \frac{Cra_A}{Cra_A + K_{e,Cra_A}} , \tag{2.19}$$

and if Cra acts as repressor, the expression rate is given by

$$v_{ex,e} = \rho\, \mu\, v_{e,max} \left(1 - \frac{Cra_A}{Cra_A + K_{e,Cra_A}}\right) . \tag{2.20}$$

In these equations, μ is the growth rate, $v_{e,max}$ the maximal expression rate, and K_{e,Cra_A} the active Cra concentration required for half–maximal expression. ρ parameterizes the linear influence that μ exerts on the efficiency of the gene expression machinery. ρ is set to 1, because its value is anyhow later corrected with the optimization of the multiplicative parameters $v_{e,max}$.

The combined dilution and degradation rates of a compound x are given by

$$v_{d,x} = (\mu + k_{degr})\, x, \tag{2.21}$$

with a degradation rate of $k_{degr} = 0$ in the case of metabolites, and $k_{degr} = 2.8 \cdot 10^{-5} s^{-1}$ in the case of proteins [51].

The enzymes E_3 and E_4 are activated in a cooperative manner by PEP and FBP, respectively, and are modeled using a Monod-Wyman-Changeux (MWC) kinetics, such that

$$v_{r,E_i} = \frac{k_{cat,E_i}\, E_i\, \frac{S}{K_{E_i,S}} \left(1 + \frac{S}{K_{E_i,S}}\right)^{n_{E_i}-1}}{\left(\frac{S}{K_{E_i,S}}\right)^{n_{E_i}} + L_{E_i} \left/ \left(1 + \frac{M}{K_{E_i,M}}\right)^{n_{E_i}}\right.} . \tag{2.22}$$

The enzymes E_1 and E_5 are inhibited in a cooperative manner by FBP and PEP, respectively, and are also modeled with a MWC kinetics, such that

$$v_{r,E_i} = \frac{k_{cat,E_i} \, E_i \, \frac{S}{K_{E_i,S}} \left(1 + \frac{S}{K_{E_i,S}}\right)^{n_{E_i}-1}}{\left(\frac{S}{K_{E_i,S}}\right)^{n_{E_i}} + L_{E_i} \, \left(1 + \frac{M}{K_{E_i,M}}\right)^{n_{E_i}}} \; . \tag{2.23}$$

In these equations, k_{cat,E_i} denote the maximal turnover capacities, $K_{E_i,S}$ and $K_{E_i,M}$ the required respective concentrations of substrate and effector for half–saturation, n_{E_i} the number of monomers in the active enzyme complex, and L_{E_i} parameterizes the ligand binding. The enzyme E_2 is modeled with a reversible Michaelis–Menten kinetics, such that

$$v_{r,E_2} = \frac{v_{E_2,f} \, E_2 \, \frac{FBP}{K_{E_2,FBP}} - v_{E_2,r} \, E_2 \, \frac{PEP}{K_{E_2,PEP}}}{1 + \frac{FBP}{K_{E_2,FBP}} + \frac{PEP}{K_{E_2,PEP}}}, \tag{2.24}$$

where $v_{E_2,f}$ and $v_{E_2,r}$ are the maximal reaction rates in the forward and reverse direction, respectively, and $K_{E_2,FBP}$ and $K_{E_2,PEP}$ the substrate concentrations required for half–saturation. Note that this enzyme's concentration is assumed to be constant.

2.6.3 Measurement data set

Table 2.2 lists the measurement data [57, 119, 142, 233] that were used to optimize the parameters. The relative enzyme concentrations listed therein were derived from DNA microarray data, thereby assuming that enzyme concentrations scale with their mRNA concentrations.

2.6.4 Derivation of the equality constraints on the parameters

This section supplements the exemplary application of the divide–and–conquer approach to obtain the complete solution space of the model defined in Equations 2.15–2.24. In the following, we derive the equality constraints on the parameters from the measurement data shown in Table 2.2, and show how these constraints divide the parameter set $\mathbf{p}$ into free and dependent parameters. The equality constraints are derived independently for each of the six subproblems into which the global optimization problem decomposes (the six subproblems are summarized in Table 2.1 and Figure 2.2b). In the following, the subscripts $_{Glc}$ and $_{Act}$ refer to measurement data points in the glucose and acetate data sets, respectively.

First, we optimize the parameters of subproblem 1, which contains the four gene expression rates and the binding of the transcription factor Cra to the metabolite FBP. Each of the four gene expression rates v_{ex,e_i} with $i = 1, 3, 4, 5$ contains two parameters, $v_{e_i,max}$ and K_{e_i,Cra_A}. With the two measurement data sets, these

Table 2.2: Measurement data obtained for balanced growth on glucose and acetate. Extracellular metabolites in $[\frac{g}{l}]$. Growth rate in $[\frac{1}{h}]$. Intracellular metabolites in $[\frac{\mu mol}{gDW}]$. Enzyme levels in [AU]. Metabolic fluxes in $[\frac{\mu mol}{gDW\ s}]$.

Condition	*Acetate*	*Glucose*	μ		*PEP*	*FBP*
Acetate	5	0	0.20		0.59	0.28
Glucose	0	5	0.64		0.21	6.6

Condition	E_1	E_2	E_3	E_4	E_5
Acetate	10.65	1	3.3	0.61	0.59
Glucose	1	1	1	1	1

Condition	J_{E_1}	J_{E_2}	J_{E_3}	J_{E_4}	J_{E_5}
Acetate	0.198	-0.188	0.094	0.01	0
Glucose	0	3.871	0.06	1.874	1.997

parameters can be determined exactly as a function of the two parameters describing the transcription factor–metabolite binding, $K_{Cra,FBP}$ and n_{Cra}. With Cra_A from Eq. 2.15 and $v_{d,x}$ from Eq. 2.21, the explicit solutions are for $i = 1, 3$

$$
\begin{aligned}
K_{e_i,Cra_A} &= (\nu\, Cra_{A,Glc} - Cra_{A,Act})(1-\nu)^{-1} \\
v_{e_i,max} &= v_{d,E_i,Glc}\,(Cra_{A,Glc} + K_{e_i,Cra_A})(\rho\,\mu_{Glc}\,Cra_{A,Glc})^{-1}
\end{aligned}
\tag{2.25}
$$

with

$$
\nu = Cra_{A,Act}\,\mu_{Act}\,v_{d,E_i,Glc}\,(Cra_{A,Glc}\,\mu_{Glc}\,v_{d,E_i,Act})^{-1} \quad , \tag{2.26}
$$

and for $i = 4, 5$

$$
\begin{aligned}
K_{e_i,Cra_A} &= (\nu\, Cra_{A,Glc} - Cra_{A,Act})(1-\nu)^{-1} \\
v_{e_i,max} &= v_{d,E_i,Glc}\left[\rho\mu_{Glc}\left(1 - Cra_{A,Glc}(Cra_{A,Glc} + K_{e_i,Cra_A})^{-1}\right)\right]^{-1}
\end{aligned}
\tag{2.27}
$$

with

$$
\nu = \mu_{Act}\,v_{d,E_i,Glc}\left(\mu_{Glc}\,v_{d,E_i,Act}\right)^{-1} \quad . \tag{2.28}
$$

These equations reduce the ten–dimensional parameter space to a two–dimensional solution space.

Next, we optimize the parameters of the five enzyme kinetics. The kinetic equations for both uptake enzymes E_1 (subproblem 2) and E_5 (subproblem 3) each contain four uncertain parameters, with only one constraining measurement each.

We chose to express v_{E_i} with $i = 1, 5$ as a function of the remaining parameters, such that:

$$
\begin{aligned}
v_{E_1} &= \frac{\frac{J_{E_1,Act}\, K_{E_1,Acetate}}{Acetate\, \cdot E_{1,Act}}}{\left(1+Acetate/K_{E_1,Acetate}\right)^{n_1}+L_{E_1}\left(1+PEP_{Act}/K_{E_1,PEP}\right)} \\[2pt]
&\qquad\qquad\qquad \left(1+Acetate/K_{E_1,Acetate}\right)^{n_1-1} \\[6pt]
v_{E_5} &= \frac{\frac{J_{E_5,Glc}\, K_{E_5,Glucose}}{Glucose\, \cdot E_{5,Glc}}}{\left(1+Glucose/K_{E_5,Glucose}\right)^{n_5}+L_{E_5}\left(1+FBP_{Glc}/K_{E_5,FBP}\right)} \\[2pt]
&\qquad\qquad\qquad \left(1+Glucose/K_{E_5,Glucose}\right)^{n_5-1}
\end{aligned}
\tag{2.29}
$$

These equations reduce the four–dimensional parameter spaces to three–dimensional solution spaces.

The kinetic equations for the enzymes E_3 (subproblem 4) and E_4 (subproblem 5) each contain four uncertain parameters, with two measurements to constrain them. With $i = 3, 4$, we chose to express v_{E_i} and the $K_{E_i,S}$-values of the respective substrates, $K_{E_3,PEP}$ and $K_{E_4,FBP}$, as a function of the remaining parameters. This gives two functions,

$$
\begin{aligned}
f_1 &= \frac{J_{E_i,Glc}\left[\left(1+S_{Glc}/K_{E_i,S}\right)^{n_i}+L_i/\left(\left(1+A_{Glc}/K_{E_i,A}\right)^{n_i}\right)\right]}{E_{i,Glc}\, S_{Glc}/K_{E_i,S}\left(1+S_{Glc}/K_{E_i,S}\right)^{n_i-1}} \\[6pt]
f_2 &= \frac{J_{E_i,Act}\left[\left(1+S_{Act}/K_{E_i,S}\right)^{n_i}+L_i/\left(\left(1+A_{Act}/K_{E_i,A}\right)^{n_i}\right)\right]}{E_{i,Act}\, S_{Act}/K_{E_i,S}\left(1+S_{Act}/K_{E_i,S}\right)^{n_i-1}} \quad,
\end{aligned}
\tag{2.30}
$$

with substrate $S = PEP$ and activator $A = FBP$ for $i = 3$, and substrate $S = FBP$ and activator $A = PEP$ for $i = 4$. Then, $K_{E_i,S}$ is given implicitly by

$$
K_{E_i,S} : \quad 0 = f_1 - f_2 \quad,
\tag{2.31}
$$

and v_{E_i} is given by

$$
v_{E_i} = f_1 \quad.
\tag{2.32}
$$

These equations reduce the four–dimensional parameter spaces to two–dimensional solution spaces.

Finally, the kinetic equation for the enzyme E_2 (subproblem 6) contains four uncertain parameters, with two measurements to constrain them. We chose to express $v_{E_2,f}$ and $v_{E_2,r}$ as functions of $K_{E_2,PEP}$ and $K_{E_2,FBP}$, such that

$$
\begin{aligned}
v_{E_2,f} &= \left[\frac{E_{2,Act}\, PEP_{Act}}{E_{2,Glc}\, PEP_{Glc}} J_{E_2,Glc}\left(1+\frac{FBP_{Glc}}{K_{E_2,FBP}}+\frac{PEP_{Glc}}{K_{E_2,PEP}}\right)\right. \\
&\qquad \left. +J_{E_2,Act}\left(1+\frac{FBP_{Act}}{K_{E_2,FBP}}+\frac{PEP_{Act}}{K_{E_2,PEP}}\right)\right] \\
&\qquad \left[\frac{E_{2,Act}}{K_{E_2,FBP}}\left(\frac{PEP_{Act}\, FBP_{Glc}}{PEP_{Glc}}-FBP_{Act}\right)\right]^{-1} \\
v_{E_2,r} &= \frac{K_{E_2,PEP}}{PEP_{Glc}}\left[v_{E_2,f}\,\frac{FBP_{Glc}}{K_{E_2,FBP}}-\frac{J_{E_2,Glc}}{E_{2,Glc}}\left(1+\frac{FBP_{Glc}}{K_{E_2,FBP}}+\frac{PEP_{Glc}}{K_{E_2,PEP}}\right)\right].
\end{aligned}
\tag{2.33}
$$

These equations reduce the four–dimensional parameter space to a two–dimensional solution space.

2.6.5 Statistical analysis of the solution space

In this section, we discuss the statistical analysis we used to identify the parameters that most critically determine the system response. To identify these parameters, we take advantage of the two response families introduced in Section 2.4.2 and compare the parameter combinations that belong to response family A with those that belong to response family B. If a parameter does not critically determine the system response, its value should be equally distributed in both response families. To test whether a parameter is equally distributed in both response families, we compare the distribution of the sampled parameter values between the two families using the two–tailed student's t-test with equal variances. This test results in one p–value per parameter. The p–value is the probability, under the null hypothesis that both parameter sets A and B are drawn from the same distribution, of observing a value as extreme or more extreme of the test statistic

$$ t = \frac{\bar{a}\,\bar{b}}{\sqrt{\frac{\sigma}{n_a} + \frac{\sigma}{n_b}}} \quad , \tag{2.34} $$

with $\bar{a}$ and $\bar{b}$ the means of the sampled parameter values in both response families, σ the pooled standard deviation, and n_a and n_b the sample sizes. The lower a parameter's p–value is, the more statistically different is the distribution of the sampled parameter values between the two response families A and B.

Table 2.3 lists the p–values of all parameters. Most parameters exhibit a high p–value and are therefore not suspected to critically shape the system response. Only few parameters exhibit low p–values, and among these, the free parameter $K_{E_2,FBP}$ and the dependent parameter $v_{E_2,f}$ exhibit extremely low p–values. Therefore, the values of these two parameters critically determine to which response family a parameter combination belongs.

In addition, Table 2.3 lists the mean and standard deviations for each parameter. Most parameters, such as L_{E_i}, exhibit large standard deviations, which implies that these parameter values are only poorly determined. However, some parameters, such as v_{E_3}, v_{E_4} and k_{cat,E_3} exhibit very narrow standard deviations, which implies that the values of these parameters are fairly well determined. The distributions of the parameters with low p–values are significantly different between the two response families.

Table 2.3: Statistical data (mean values, standard deviations, and p–values) of the distributions of the sampled parameter values in response families A and B.

Free parameter	mean±std family A	mean±std family B	p–value
n_{Cra}	1.37±0.84	1.43±0.96	0.55
$K_{Cra,FBP}$	2.87±0.75	3.00±0.76	0.17
L_{E_1}	$5 \cdot 10^5 \pm 2 \cdot 10^6$	$8 \cdot 10^5 \pm 2 \cdot 10^6$	0.28
$K_{E_1,PEP}$	5.09±2.90	5.18±3.18	0.82
$K_{E_1,Acetate}$	5.14±2.84	4.77±2.70	0.30
L_{E_5}	$6 \cdot 10^5 \pm 2 \cdot 10^6$	$6 \cdot 10^5 \pm 2 \cdot 10^6$	0.88
$K_{E_5,PEP}$	7.07±1.68	6.13±1.49	$1 \cdot 10^{-5}$
$K_{E_5,Glucose}$	5.11±2.90	5.18±2.84	0.86
L_{E_3}	$7 \cdot 10^4 \pm 2 \cdot 10^5$	$8 \cdot 10^4 \pm 2 \cdot 10^5$	0.72
$K_{E_3,PEP}$	5.09±2.79	4.73±2.72	0.32
L_{E_4}	$2 \cdot 10^6 \pm 2 \cdot 10^6$	$2 \cdot 10^6 \pm 2 \cdot 10^6$	0.52
$K_{E_4,FBP}$	0.61±0.26	0.69±0.25	0.01
$K_{E_2,PEP}$	4.91±2.85	5.62±2.65	0.05
$K_{E_2,FBP}$	5.48±2.59	0.85±0.56	$3 \cdot 10^{-145}$

Dependent parameter	mean±std A family A	mean±std B family B	p–value
v_{E_1}	52.5±133	53.2±126	0.96
v_{E_3}	5.43±0.50	5.42±0.51	0.93
v_{E_4}	$1.16 \pm 7 \cdot 10^{-3}$	$1.16 \pm 8 \cdot 10^{-3}$	0.88
v_{E_5}	$1.17 \pm 9 \cdot 10^{-3}$	$1.17 \pm 9 \cdot 10^{-3}$	0.88
$K_{E_1,Cra}$	2.17±8.18	2.23±7.67	0.96
$K_{E_3,Cra}$	0.09±0.10	0.09±0.10	0.95
$K_{E_4,Cra}$	3.25±0.80	3.12±1.01	0.22
$K_{E_5,Cra}$	2.79±0.69	2.68±0.87	0.22
k_{cat,E_1}	$2 \cdot 10^4 \pm 2 \cdot 10^5$	$1 \cdot 10^4 \pm 6 \cdot 10^4$	0.95
k_{cat,E_5}	$3 \cdot 10^5 \pm 1 \cdot 10^5$	$2 \cdot 10^5 \pm 7 \cdot 10^5$	0.70
k_{cat,E_3}	$0.06 \pm 9 \cdot 10^{-4}$	$0.06 \pm 9 \cdot 10^{-4}$	0.65
$K_{E_3,PEP}$	0.11±0.10	0.11±0.10	0.65
k_{cat,E_4}	184±677	254±626	0.41
$K_{E_4,PEP}$	0.30±0.15	0.32±0.14	0.20
$v_{E_2,f}$	7.59±1.85	4.48±0.36	$3 \cdot 10^{-151}$
$v_{E_2,r}$	5.86±4.61	24.2±22.1	$3 \cdot 10^{-9}$

2.6.6 Further graphical illustration of the results of the statistical analysis

Figure 2.4 shows Fig. 2.3c viewed along the $K_{E_2,PEP}$–axis. From this perspective, it can be seen that a combination of roughly $K_{E_2,FBP} < 3$ and $v_{E_2,f} < 6$ is necessary for the emergence of an attractive second steady state (gray dots).
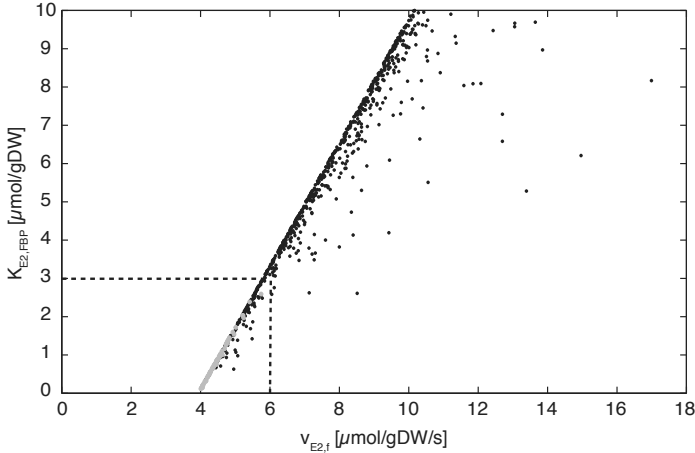
Figure 2.4: This figure plots the sampled parameters shown in Fig. 2.3c viewed along the $K_{E_2,PEP}$–axis. Black dots denote parameter combinations that belong to response family A; gray dots denote those that belong to response family B. The emergence of an attractive second steady state, i.e. a response belonging to family B, is only possible if the parameters $K_{E_2,FBP}$ and $v_{E_2,f}$ lie within a region that is roughly delimited with dashed lines.

2.7 Acknowledgements

Chapter 3

Bacterial adaptation through distributed sensing of metabolic fluxes

Oliver Kotte, Judith Zaugg and Matthias Heinemann

3.1 Summary

The recognition of carbon sources and the adaptation to recognized changes are of particular importance for bacterial survival in fluctuating environments. Despite a thorough knowledge base of *E. coli*'s central metabolism and its regulation, fundamental aspects of the employed sensing and adaptation mechanisms remain unclear. For instance, for many of *E. coli*'s substrates such as acetate, no sensing mechanism has as yet been identified. Also, it remains on the molecular level unclear how the known local regulations work together to accomplish the particularly complex, system-wide adaptations between glucose and acetate. In this chapter, we present a differential equation model of *E. coli*'s central metabolism that includes molecular regulation on both enzymatic and genetic layers. With this model, we show that the interplay of the known interactions in *E. coli*'s central metabolism explains in molecular-level detail the system-wide adaptations of metabolic operation between glycolytic and gluconeogenic carbon sources (such as glucose and acetate). We show that these adaptations are enabled by an indirect recognition of carbon sources through a mechanism we termed distributed sensing of intracellular metabolic fluxes. This mechanism uses two distinct general motifs to establish flux-signalling metabolites, whose bindings to target transcription factors form flux sensors. These sensors are embedded in global feedback loop architectures that orchestrate the regulatory adjustments to recognized changes in carbon source availability. By connecting system-level understanding to molecular-level knowledge, these general principles improve our understanding of bacterial metabolism's operation in environments with fluctuating carbon sources.

3.2 Introduction

Adaptations to fluctuating carbon sources are of particular importance for bacteria. These adaptations are realized by molecular *systems*, which (i) *recognize* carbon sources and (ii) *regulate* the adjustment of metabolic operation — on both enzymatic and genetic regulatory layers — to the recognized changes. To understand system behavior, molecular knowledge alone is often not sufficient [96]. Instead, it needs to be understood how a system's behavior emerges from the interactions between the characterized molecules [95]. To attain such a system understanding of bacterial adaptations to carbon sources, the *coupling* between the recognition and regulation aspects and between the enzymatic and genetic regulatory layers must be understood.

Recent studies in *E. coli* have focused on these couplings to improve our understanding of such adaptations in terms of general, topological motifs [13, 84, 103, 128, 171, 189, 226]. However, these studies do not link these topological motifs to the molecular-level details of specific adaptations; therefore, the molecular-level interplay of enzymatic and genetic regulation is to date only understood for comparatively simple carbon source adaptations, such as the adaptation to lactose [42, 144]. For more complex adaptations involving many operons, such as

the adaptations between the glycolytic substrate glucose and the gluconeogenic substrate acetate that require an extensive remodeling of central metabolism, it remains unclear how recognition and regulation function in molecular-level detail, and how these processes are coupled to a coordinated adaptation.

Bacteria typically employ two major means to recognize carbon sources. Some sugars (e.g. fructose, galactitol, mannitol, mannose, sorbitol) are recognized via the phosphotransferase sugar uptake system (PTS) (e.g. [25, 26, 83, 155, 184]); other sugars (e.g. arabinose, glycerol, galactose, lactose, maltose, melibiose, fucose) are recognized intracellularly via regulatory proteins (transcription factors, in short TFs) (e.g. [42, 128, 144, 189]). However, for organic acids such as succinate, malate or the frequently encountered acetate as well as for many other carbon sources, neither transmembrane sensors nor regulatory proteins with sensing function have been identified. It thus remains unclear how these carbon sources are recognized.

Concerning the regulated adjustment of metabolic operation between growth on glycolytic and gluconeogenic carbon sources, only local aspects are understood in molecular-level detail. Examples of such local aspects are the branch point effect at the diversion of carbon flux through the glyoxylate shunt [109], the PEP-pyruvate-oxaloacetate node as the switch point for carbon flux distribution [181], or the regulation of cAMP levels by the PTS [25]. What remains unclear is how these local regulations work together to accomplish a coordinated adaptation on the systems level.

In this chapter, we show that (i) the interplay of the known interactions in *E. coli*'s central metabolism is capable to *indirectly* recognize carbon sources through a mechanism we termed *distributed sensing of intracellular metabolic fluxes*, and that (ii) these molecular-level interactions can regulate *E. coli*'s adjustment of metabolic operation between growth on glycolytic and gluconeogenic carbon sources, and that (iii) this adaptation is governed by general principles. We derived these results with a simulation-based approach that rests on a differential equation model of *E. coli*'s central metabolism covering both enzymatic and transcriptional regulation.

3.3 Model

When we screened the available molecular knowledge of central metabolism to understand *E. coli*'s adaptations from growth on glucose to acetate and vice versa, we noted that:

- Four TFs have been identified that regulate the expression of central metabolic enzymes *and* whose activities are modulated by binding of central metabolites:
  - Cra–fructose-1,6-bisphosphate (Cra-FBP) [162]
  - Crp–cyclic AMP (Crp-cAMP) [30]
  - IclR–glyoxylate (IclR-GLX) and IclR–pyruvate (IclR-PYR) [118]

  − PdhR–pyruvate (PdhR-PYR) [157].

- Each of the four involved metabolites assumes distinct levels during glycolytic and gluconeogenic growth across available published experimental datasets [19, 25, 30, 119].

- The levels of enzymes in central metabolic pathways regulated by these TFs is markedly distinct for growth on glycolytic and gluconeogenic carbon sources [142].

- The directions into which these enzyme levels change are consistent with their transcriptional regulation by the four TFs, assuming that these TF activities are distinct during glycolytic and gluconeogenic growth due to distinct levels of their activating/inhibiting effectors [92].

The fact that this set of differentially expressed pathways covers most of central metabolism, and the promising results of previous work on the role of TF-metabolite interactions in cellular recognition and regulation [128, 171, 189] let us to hypothesize that the current knowledge of *E. coli*'s metabolism can already explain the molecular adaptations between glycolytic and gluconeogenic carbon sources. Specifically, we hypothesized that these adaptations are accomplished by a system-wide regulation architecture that emerges when the known enzymatic and transcriptional regulations become coupled through TF-metabolite interactions. To (i) assess whether the such coupled molecular interactions can indeed work together to adapt metabolic operation, and if yes, (ii) to understand this system-level adaptation in molecular-level detail, we constructed a comprehensive differential equation model that is centered on the *coupling* of enzymatic and transcriptional regulation, which is accomplished by the above listed TF-metabolite interactions.

3.3.1 Model topology

The model topology, shown in Figure 3.6, comprises the Embden-Meyerhoff pathway, the tricarboxylic acid (TCA) cycle, the glyoxylate shunt, the anaplerotic reactions, the diversion of carbon flux to the glyoxylate shunt through phosphorylation of isocitrate dehydrogenase, the uptake of glucose through phosphorylation of PTS proteins, the uptake and excretion of acetate, the allosteric regulation of enzymes in the mentioned pathways and their transcriptional regulation by the above listed TFs, and the regulation of these TFs' activities through the above listed TF-metabolite interactions. Overall, the model comprises two compartments, the cell and its environment. The cellular compartment contains twelve metabolites, 22 enzymes and two PTS proteins, four TFs, 17 transcriptional regulations, 28 enzymatic regulations, 26 metabolic reactions, two kinase and two phosphatase reactions, five TF-metabolite interactions, the expression of 16 genes and the degradation of the produced proteins as well as their dilution due to cell

growth. The environmental compartment contains two carbon sources. The cell and the environment are coupled through three exchange reactions.

This model topology is centered on the above listed TF-metabolite-interactions, as illustrated in Figure 3.1, and includes the known molecular interactions in *E. coli*'s central metabolism, retrieved from the EcoCyc database [92]. The followed strategy for the systematic assembly of these interactions (see Section 3.6.1.1) (i) ensures the inclusion of all known interactions between modeled compounds, and (ii) minimizes the number of omitted interactions across the system boundary, which occur, for instance, when a modeled reaction is regulated by a metabolite outside the system boundary. The two measures establish a system boundary that cuts out the studied adaptation mechanism from the rest of the cell. To reduce complexity, we did not model energy, cofactor, oxygen, and proton balances.
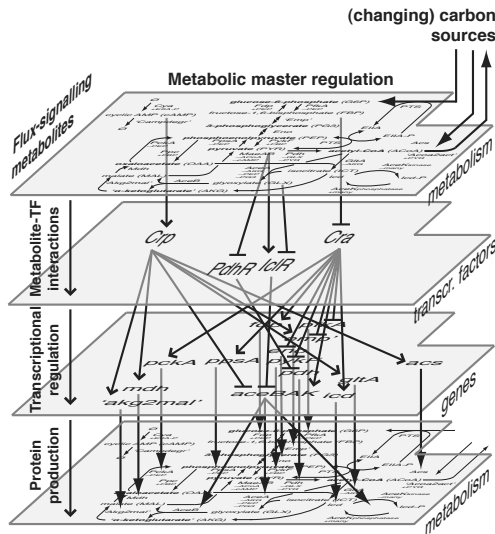
Figure 3.1: The model topology is centered around the known TF-metabolite interactions and establishes a feedback loop from the metabolic layer via the transcriptional regulatory layer and the gene expression layer back to the metabolic layer. The metabolic layer is directly upstream of the transcriptional regulatory layer, enabling it to perform the coordinating function of metabolic master regulation.

3.3.2 Kinetic equations

We translated the topology into differential equations by assigning the most appropriate rate law to each interaction (see Tables 3.1 and 3.2). To formulate the rate equation for the biomass-generating reaction, we back-calculated the require-

ments for metabolites outside the system boundary to their respective precursors inside the system boundary. We included the growth-rate dependency of the gene expression rates due to growth rate-dependent levels of DNA polymerases and ribosomes [33].

The resulting 47 ordinary differential equations contain 193 parameters and are of the form

$$\dot{\mathbf{x}} = \mathbf{S} \cdot \mathbf{f}(\mathbf{x}, \mathbf{p})$$

with $\mathbf{x}$ the vector of dynamic state variables (the concentrations of intracellular compounds and extracellular carbon sources), $\mathbf{f}$ the vector of kinetic rate equations, $\mathbf{p}$ the vector of parameters, and $\mathbf{S}$ the stoichiometric matrix. Refer to Section 3.6.1 for the full model equations.

Parameter values for the rate equations were estimated through application of the 'divide-and-conquer approach' [101] on published experimental steady state -omics data sets for balanced growth on either glucose or acetate (see Section 3.6.2).

3.4 Results

Table 3.5 lists the values of a single parameter vector that excellently reproduces the steady state data (see Tables 3.3 and 3.4) for balanced growth on either glucose or acetate with the coefficient of determination $R^2 \approx 0.999$, and with the sensitivities shown in Figure 3.8. These parameter values are used for all simulations presented in this chapter.

3.4.1 Reproduction of known physiological behavior

To test if the model is indeed capable to adapt to changing carbon sources, we subjected it to a sequence of three consecutive environments. The first environment contains glucose as the sole carbon source and is 'inoculated' with glucose-adapted in silico cells, meaning that the initial conditions of the in silico cells were set to the steady state values of glucose growth. Figure 3.2 shows that the in silico cells grow on glucose, produce acetate from glucose, and after glucose depletion commence their adaptation to acetate, which they re-consume until it is depleted. The in silico cells are then 'transferred' to a second environment that contains acetate as the sole carbon source, which the cells consume to successfully complete their adaptation to the acetate-adapted steady state. Finally, the in silico cells are 'transferred' to a third environment that contains both glucose and acetate as carbon sources. Although at this point the cells are adapted to the present acetate, they quickly adapt to glucose instead, produce acetate from glucose, and then re-adapt to acetate following glucose depletion.

These simulations show that the model reproduces *E.coli*'s known physiological behavior of glucose repression, which is the preferential uptake of glucose over acetate, and of an overflow metabolism, which is the production of acetate from glucose. Figures 3.9–3.13 show that (i) throughout the transitions, all intracellular
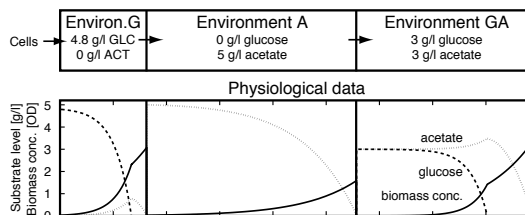
Figure 3.2: The model reproduces *E. coli*'s known physiological behaviors of preferential glucose uptake and of acetate production from glucose.

metabolite and enzyme levels remain within physiologically reasonable bounds and (ii) all compound levels and reaction rates approach their measured steady state values for balanced growth.

Whereas the measured steady states and the physiological behavior are correctly reproduced, the simulated trajectories of the intracellular compound levels during the transitions from one of the two steady states to the other are uncertain. The uncertainty of the simulated trajectories arises from (i) uncertainty in the parameter values, (ii) uncertainty in the model structure due to ambiguity in the selection of rate laws, and (iii) possible effects of not modeled cellular regulations, i.e. those ensuring the here omitted energy, cofactor etc. balances, onto the simulated trajectories. It is important to note, however, that despite of the time profiles during transitions being uncertain, the model is *capable* to successfully adapt between growth on glucose and acetate. This capability can only arise if it is supported by the model structure — which in systems biology models tightly constrains the possible responses [34, 66]. We can thus use the model as a tool to derive hypotheses for the emergence of the studied adaptations from the interplay of the modeled, molecular-level interactions.

3.4.2 Recognition of extracellular carbon sources

As the in silico cells successfully adapt to fluctuating levels of glucose and acetate, they must have a mechanism to recognize these carbon sources. But how does an in silico cells recognize acetate without a transmembrane sensor for extracellular acetate or a TF binding to intracellular acetate? Further, is the regulation of Crp activity by the glucose transporter the exclusive mechanism to recognize glucose, or is this sensing function of the PTS integrated into a larger sensing architecture?
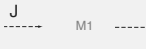
Because a sensing mechanism translates environmental information into TF activity, either through phosphorylation or effector binding, the sensing mechanism of the in silico cell is identified once it is understood how an extracellular carbon source affects intracellular TF activities. Since the TFs Cra, Crp, IclR and PdhR are not phosphorylated, the in silico cell can only modulate its TF activities by modulating the concentrations of the TFs' effectors. Our model simulations

revealed that the TFs' effectors indeed respond to changes in the availability of extracellular carbon sources with a concentration change that modulates the TFs' activities (see Figure 3.14). But what links the levels of the intracellular effectors to the presence of extracellular carbon sources? The only entity capable to provide this missing link is metabolic flux, which must therefore form an integral part of the sensing mechanism. With this constraint, we deduced the following tripartite sensing mechanism.

First, the cell needs to ensure at least a basal uptake of the recognized carbon source, e.g. through a basal expression of the relevant transporter. Hence, when that carbon source enters the cell's environment, it is taken up at least at a basal rate. In the in silico cell, such basal uptake is realized by a constitutive expression of the glucose-transporting PTS, and by a basal expression of the acetate-transporting 'super-enzyme' Acs, which in the model lumps the acetate transport reaction and the subsequent conversion to acetyl-CoA (note that on glucose, acetate is both produced *and* re-consumed at a lower rate). Because the uptake of a carbon source propagates as intracellular flux through downstream metabolic pathways, the *ensured* uptake of a carbon source whenever present implies that these fluxes are affected by the *presence* of this carbon source.

Second, the allosteric enzyme regulation in pathways affected by the uptake flux is such that a metabolic intermediate in these pathways responds to the uptake flux in a defined way. The in silico cell uses two distinct motifs to establish such *flux-signalling metabolites.*

1. The motif *pathway usage*, illustrated in the left column of Figure 3.3, places the flux-signalling metabolite in a pathway that is only used in one growth condition but not in the other; hence, synthesis of the metabolite, i.e. a high metabolite concentration, signals pathway usage. The in silico cell uses this motif to establish distinct levels of cyclic AMP (cAMP) and glyoxylate (GLX). In both cases, the differential pathway induction is realized through fast changes in protein phosphorylation states. The phosphorylation of the PTS protein EIIA, which induces the formation of cAMP, is directly coupled to glucose uptake; the coupling of acetate uptake to flux through the glyoxylate shunt is achieved by the regulation of the phosphorylation of isocitrate dehydrogenase, which diverts flux from the TCA cycle to the glyoxylate shunt.

2. The motif *flux direction*, illustrated in the right column of Figure 3.3, places the flux-signalling metabolite in a reversible pathway that is operated in different directions depending on the growth condition. Since different flux directions mean that different enzymes 'consume' the metabolite, these enzymes' kinetics can have different 'flux resistances' such that the metabolite assumes high levels for one flux direction and low levels for the other; hence, the metabolite level signals flux direction. The in silico cell uses this motif to create distinct levels of the flux-signalling metabolites fructose-1,6-bisphosphate (FBP) and pyruvate (PYR).

Figure 3.3: Two general motifs to establish flux-signalling metabolites. The first motif, pathway usage, places the flux-signalling metabolite in a pathway that is used in one growth condition but not in the other. The second motif, flux direction, places the flux-signalling metabolite in a reversible pathway that is used in different directions depending on the growth conditions.

Third, the cell *senses* the *signal* provided by the flux-signalling metabolites through interactions of these metabolites with target TFs. These interactions realize *sensors for intracellular metabolic flux*. The in silico cell is equipped with four such flux sensors. Three of these sensors ultimately measure at different positions the flux direction (and possibly its magnitude) through the Embden-Meyerhoff pathway (EMP). Cra-FBP measures the flux through the upper EMP where FBP is located. On glucose, PdhR-PYR measures the lower glycolytic flux that is fed into the TCA cycle via PYR, whereas on acetate, PdhR-PYR measures the flux through the malic enzymes that is fed into the lower EMP. When glucose is taken up, the PTS-coupled Crp-cAMP sensor reports the glucose uptake and thus the glycolytic flux into the upper EMP; on acetate, the PTS-coupled cAMP level equilibrates with the PEP/PYR ratio that results from gluconeogenic flux into the lower EMP. The fourth of these flux sensors, IclR-GLX, senses the flux through the glyoxylate shunt and is further modulated by the PYR signal (through IclR-PYR).

Figure 3.4 shows that the levels of the in silico cell's four flux-signalling metabolites as well as the activities of the four target TFs are distinct not only for balanced growth on either glucose or acetate, but also for lower uptake rates of these substrates. Thus, the property of distinct TF activities is robust to changes in the substrate uptake rates, which may arise e.g. from fluctuating concentrations of the carbon sources. Because of this robustness, the TFs reliably sense the *presence* of extracellular glucose and acetate through binding 'endogenous' [128, 189] metabolite signals. Note that in some cases (e.g. Cra-FBP on glucose, see the fourth column in Figure 3.4), the gradual differences in TF activity resulting from variations in substrate uptake rates vary sensitively and monotonously with the

magnitude of the sensed flux, which conceptually enables the sensing of intracellular flux *magnitudes*.
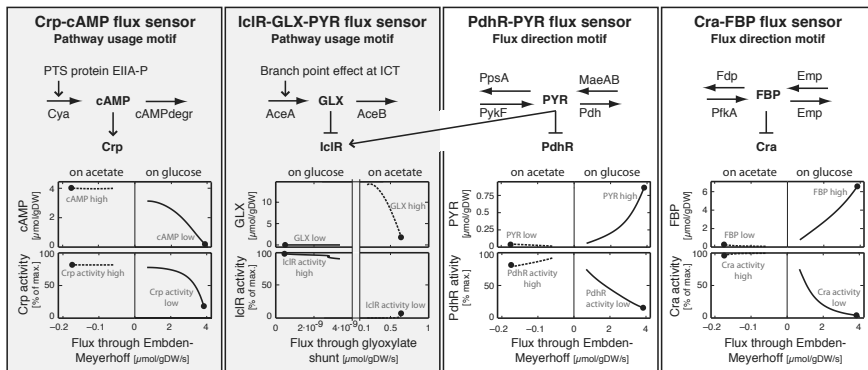
Figure 3.4: The four flux sensors of the in silico cell. The four flux-signalling metabolites are established through regulatory circuits that realize the motifs 'pathway usage' and 'flux direction'. The simulated dependencies of the flux-signalling metabolites on the fluxes they measure show that these metabolites exhibit markedly distinct levels not only for balanced growth (filled dots), but also for up to 80% lower glucose (regular lines) or acetate (dashed lines) uptake rates. The levels of these metabolites thus report extracellular glucose and acetate. These distinct metabolite levels propagate into distinct activities of their target TFs, which adjusts the transcriptional regulation exerted by these TFs to the present carbon sources.

Because the in silico cell establishes sensors for several intracellular fluxes, the overall sensing architecture infers the present carbon sources from a *pattern* of intracellular fluxes and is as such of a *distributed* nature. We therefore termed this architecture *distributed sensing of intracellular metabolic fluxes*. The core of this sensing architecture is formed not by transmembrane sensors but by the four circuits shown in Figure 3.4, which establish flux-signalling metabolites according to the two presented general motifs. These circuits employ intracellular metabolic flux as a mediator to couple the presence of extracellular carbon sources to the levels of intracellular metabolites. The recognition of glucose through the PTS transmembrane complex is embedded as one circuit in this distributed sensing architecture; the other three circuits function without the help of transmembrane complexes.

3.4.3 Coupling of recognition and regulation

Having identified the sensing mechanism, it still needs to be understood how the recognition is coupled to the regulation such that the in silico cell responds to recognized changes in carbon source availability with a regulated adjustment of metabolic operation.

The in silico cell achieves the coupling between recognition and regulation through its TFs, whose activities respond to the available carbon sources and at the same time regulate the expression of target genes. This combined recognition and regulation function of the four TFs closes four *global feedback loops* that overarch the metabolic and genetic layers as illustrated in Figure 3.5. The first half of these four loops forms the recognition function and is established by the flux-signalling metabolites binding to their target TFs, creating flux sensors (the four columns in Figure 3.4, and the upper of the three layers of arrows in Figure 3.1). The second half of these four loops forms the regulation function and is established by (i) the transcriptional regulation of the four flux-sensing TFs (the middle layer of arrows in Figure 3.1), which causes the regulated enzymes to approach their measured steady state values, and by (ii) the impact of this transcriptional regulation on metabolic operation (the lower layer of arrows in Figure 3.1), which, together with allosteric enzyme regulation, adjusts the carried metabolic fluxes to produce biomass precursors from the present carbon sources.
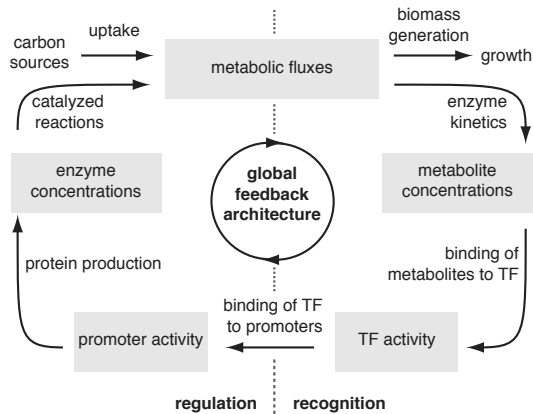
Figure 3.5: Global feedback loop architecture. This architecture, which overarches the metabolic and genetic layers, ties the recognition of carbon sources and the regulation of the adjustments to the recognized changes together.

To sum up, the adaptation of the in silico cell arises from the global feedback loop-embedded, flux sensor-adjusted transcriptional regulation of their four TFs, with each TF performing one part of the overall adaptation. This adaptation incorporates both the influence of the metabolic on the genetic layer, achieved through TF-metabolite interactions, and of the genetic on the metabolic layer, achieved through the impact of adjusted enzyme levels on metabolic fluxes. Remarkably, the thus formed global feedback loop follows the general logic of the recently proposed consumer motif, which has been suggested to be ideal for the regulation of carbon source uptake [103].

3.4.4 Metabolic master regulation

We have shown that the transcriptional adaptation of the in silico cell is realized by four global feedback loops. As the four involved TFs are not regulated by a common transcriptional master regulator, the question arises how the regulatory actions of these TFs are coordinated to a coherent overall response.

Because the in silico cell can modulate its four TF activities only through the levels of the TFs' effectors, a coordination of the TF activities requires a coordination of the levels of the (flux-signalling) effectors. This coordination must occur on the metabolic layer, which is one step upstream of the TF regulatory layer (see Figure 3.1). We identified two means that contribute to such *metabolic master regulation.*

First, flux-signalling metabolite levels are coupled because the metabolic fluxes to which they respond are connected to each other through the flux balances at network nodes. Therefore, when the local flux signalled by one of the flux-signalling metabolites (and thus this metabolite's level) changes, then the local fluxes signalled by the other metabolites (and thus these metabolites' levels) are also likely to change.

Second, flux-signalling metabolite levels are also coupled because they mutually regulate each other's adjacent enzymes. Therefore, a change in one metabolite levels propagates into changes of the regulated levels. In detail:

- The levels of cAMP, PYR, PEP and the phosphorylation state of the EIIA protein are coupled through the phosphorylations of the PTS transporter [25].

- The levels of FBP and PEP are coupled through a dual-time switch, a motif that has been shown to be rapidly inducible yet robust to noise [31]. This switch ensures a high level of FBP and a low level of the downstream PEP on glucose, and the reverse behavior on acetate. The fast switch is formed by PEP and FBP mutually activating each other's consuming enzymes (Fdp, PykF, Ppc) through feed-forward loops, a motif that enables a high level of the upstream metabolite to lower the level of the downstream metabolite [102]. The slow switch amplifies this coupling through the transcriptional regulation of FBP-inhibited Cra on the expression of these metabolites' producing and consuming enzymes Fdp, PfkA, PckA, PykF and PpsA.

- The levels of PYR and GLX are coupled through GLX acting as inhibitor of the PYR-consuming Pdh reaction, and PYR acting as corepressor of the glyoxylate shunt operon aceBAK and as effector regulating the phoshorylation of isocitrate dehydrogenase (which diverts flux from the TCA through the glyoxylate shunt).

These two means endow the metabolic layer with the function of metabolic master regulation, which coordinates the transcriptional regulations of the four TFs to a coherent overall response.

3.5 Discussion

In this chapter, we presented a differential equation model of *E. coli*'s central metabolism. This model includes transcriptional, posttranslational and enzymatic regulation and is centered on the coupling of the genetic and metabolic layers, which is accomplished by TF-metabolite interactions. The model offers a consistent explanation of how a multitude of known molecular interactions fit into a coherent systems picture; the interactions work together like gear wheels that mesh with one another to adapt central metabolism between growth on the glycolytic substrate glucose and the gluconeogenic substrate acetate. Although the experimental validation of the here computationally derived principles is challenging and certainly beyond the scope of this work, two facts strongly suggest that these principles do operate in vivo.

First, the extent *and* simplicity to which the studied systems-level adaptation is explained through the known molecular interactions is stunning: *E. coli*'s known enzymatic and transcriptional regulation, when coupled through the bindings of effectors to only *four* TFs, can explain the recognition of glycolytic and gluconeogenic carbon sources *and* the regulated adjustments of the Embden-Meyerhoff pathway, of the glyoxylate shunt, of the anaplerotic reactions and of the pyruvate dehydrogenase reaction between growth on glucose and acetate, incorporating the branch point effect through phosphorylation of isocitrate dehydrogenase, the uptake of glucose through PTS protein phosphorylations, the uptake and excretion of acetate, while reproducing the preferential uptake of glucose and the production of acetate from glucose.

Second, the general principles deduced from the presented model also describe the well-studied (and not modelled) adaptation of *E. coli* from glucose to lactose [42, 144]. In detail, the basal uptake of extracellular lactose is ensured through a basal expression of the lactose transporter LacI. Intracellular lactose is a flux-signalling metabolite for lactose uptake flux, established through the motif of pathway usage. A flux sensor is created through the binding of this flux signal to the lactose repressor protein LacI. A 'global' feedback loop is closed by the LacI-regulated induction of the lac operon. Hence, the here deduced *distributed* adaptation mechanism reduces to correctly describe the (not distributed) adaptation of one operon by one TF. As *E. coli*'s adaptation between glucose and acetate and its adaptation to lactose can both be explained by the general principles deduced in this chapter, these adaptations may differ only in complexity but not in their nature of functioning according to the here presented principles.

The general principles presented in this chapter fall under the umbrella of *distributed flux sensing*, which is realized through the binding of *flux-signalling metabolites* to target TFs, created through the motifs *signalling of pathway usage* and *signalling of flux direction*, embedded in *global feedback loop architectures*, and coordinated by *metabolic master regulation*. These principles arise from the coupling of recognition and regulation as well as of enzymatic and genetic regulation; they provide the missing link to understand system-level adaptations to carbon

sources in molecular-level detail. A fascinating implication of these principles is that they place metabolism on the top of the cellular regulatory hierarchy — a position recently argued for in higher cells [93, 191].

3.6 Model details

3.6.1 Model structure

This section presents in detail the mechanistic model of *E. coli*'s central carbon metabolism and its regulation.

3.6.1.1 Strategy for the derivation of the model topology

The topology of the model is shown in Figure 3.6. This topology comprises two compartments. The first compartment represents the cells' closed environment (as in batch experiments). The environment contains a population of identical cells and the two carbon sources glucose and acetate. The second compartment represents the cells.

The topology of the cellular compartment is centered on the five known transcription factor (TF)-metabolite interactions in *E. coli*'s central metabolism (Cra-fructose-1,6-bisphosphate, Crp-cyclic AMP, IclR-glyoxylate, IclR-pyruvate, and PdhR-pyruvate). The topology was derived with the following strategy:

1. The topology is seeded with the five TF-metabolite-interactions and the two carbon sources glucose and acetate.

2. The seeded metabolites are connected to each other through the inclusion of metabolic pathways.

3. The enzymes catalyzing the metabolic reactions are added. In the case of isozymes, only the dominant enzyme is chosen.

4. Regulations of enzyme activity through metabolites appearing in the model, and through phosphorylations, are added.

5. The expression of the modeled enzymes is added, along with their known transcriptional regulations by the four TFs. The concentrations of the TFs themselves and of the phosphotransferase system (PTS) proteins are modeled as constant. The intermediate mRNA is ignored.

6. The model is simplified by merging linear pathways to single reactions whenever the eliminated metabolic intermediates do not appear as effectors elsewhere in the model, and the merged enzymes are unregulated or co-regulated by the same TF.

7. The production of biomass from the precursor metabolites and the calculation of the growth rate are added.
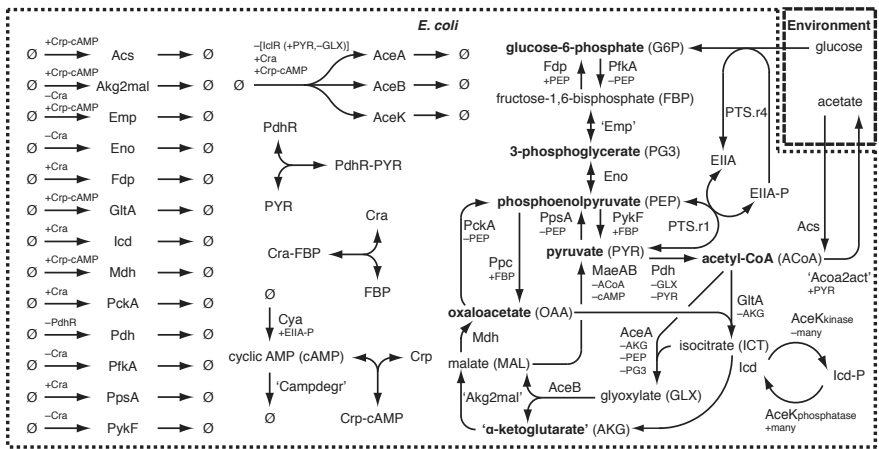
Figure 3.6: Topology of the model. The model comprises two compartments, the cell and its environment. Bold metabolite names are biomass precursors and as such substrates for first-order reactions to void. Regulation of enzyme activity through small molecule effectors is indicated below the enzyme name, transcriptional regulation above the protein-producing reaction.

3.6.1.2 Overview of the model structure

When translated into mathematical equations, the model comprises 47 state variables $\mathbf{x}$, 193 parameters $\mathbf{p}$, and 109 rate equations $\mathbf{f}(\mathbf{x}, \mathbf{p})$. Given an initial condition $\mathbf{x}(t = 0)$, the progression of the state variables over the independent variable, the time t, is given by $\dot{\mathbf{x}} = \mathbf{S} \cdot \mathbf{f}(\mathbf{x}, \mathbf{p})$, with $\mathbf{S}$ the constant stoichiometric matrix of dimension 47 x 109. Hence, $\mathbf{S} \cdot \mathbf{f}$ calculates the differential change of the state variables from the rates occurring in the system. The model is thus fully described by $(\mathbf{x}, \mathbf{p}, \mathbf{f}, \mathbf{S})$.

The 47 state variables, arranged in the vector $\mathbf{x}$, are shown in Table 3.3. This table lists the full names of the molecular compounds represented by the 47 state variables x_{abbrv}, with $_{abbrv}$ the abbreviations for the represented compounds. Table 3.3 also lists the measured or derived values during balanced growth on glucose and acetate, which were used for parameter estimation (see Section 3.6.2) and as initial conditions for the simulations.

The 109 rates, arranged in the vector $\mathbf{f}$, are shown in Table 3.4. This table lists the full names of the molecular rates represented by the 109 rates f_{abbrv}, with $_{abbrv}$ the abbreviations for the represented rates. Table 3.4 also lists the measured or derived values during balanced growth on glucose and acetate, which are used for parameter estimation (see Section 3.6.2).

The 193 parameters, which appear in the rate equations $\mathbf{f}(\mathbf{x}, \mathbf{p})$, are arranged in the vector $\mathbf{p}$ and shown in Table 3.5. This table lists the mechanistic meaning of the parameters and their estimated values (see Section 3.6.2).

The stoichiometric matrix $\mathbf{S}$ is mostly zero, except at few entries. Rather than presenting this sparse 47 x 109 matrix directly, it is far more informative to present the 47 equations that arise when $\mathbf{S} \cdot \mathbf{f}$ is multiplied out (see next Section). If needed, the stoichiometric matrix $\mathbf{S}$ can be deduced from these equations.

3.6.1.3 Balance equations

The 47 differential equations, which describe the time progression of the 47 state variables over time as a function of the system's rates, balance the

- biomass of the cell population

- concentrations of extracellular carbon sources

- concentrations of metabolites

- concentrations and phosphorylation states of enzymes and PTS proteins

- binding states of transcription factors.

The balance equations are:

Biomass

$$\dot{x}_{OD} = f_{ENV,growth}$$

Extracellular carbon sources

$$\dot{x}_{ACT} = f_{ENV,ACTex} - f_{ENV,ACTup}$$
$$\dot{x}_{GLC} = -f_{ENV,GLCup}$$

Metabolites

$$\dot{x}_{ACoA} = f_{E,Acs} + f_{E,Pdh} - f_{E,Acoa2act} - f_{E,GltA} - f_{E,AceB}$$
$$- f_{D,ACoA} - f_{BM,ACoA}$$
$$\dot{x}_{AKG} = f_{E,AceA} + f_{E,Icd} - f_{E,Akg2mal} - f_{D,AKG} - f_{BM,AKG}$$
$$\dot{x}_{cAMP} = f_{E,Cya} - f_{E,CAMPdegr} - f_{D,cAMP}$$
$$\dot{x}_{FBP} = f_{E,PfkA} - 0.5 \cdot f_{E,Emp} - f_{E,Fdp} - f_{D,FBP}$$
$$\dot{x}_{G6P} = f_{E,Fdp} - f_{E,PfkA} + f_{PTS,r_4} - f_{D,FBP} - f_{BM,FBP}$$
$$\dot{x}_{GLX} = f_{E,AceA} - f_{E,AceB} - f_{D,GLX}$$
$$\dot{x}_{ICT} = f_{E,GltA} - f_{E,AceA} - f_{E,Icd} - f_{D,ICT}$$
$$\dot{x}_{MAL} = f_{E,AceB} + f_{E,Akg2mal} - f_{E,MaeAB} - f_{E,Mdh} - f_{D,MAL}$$
$$\dot{x}_{OAA} = f_{E,Ppc} + f_{E,Mdh} - f_{E,PckA} - f_{E,GltA} - f_{D,OAA} - f_{BM,OAA}$$
$$\dot{x}_{PEP} = f_{E,PckA} + f_{E,PpsA} + f_{E,Eno} - f_{E,Ppc} - f_{E,PykF} - f_{PTS,r_1}$$
$$- f_{D,PEP} - f_{BM,PEP}$$
$$\dot{x}_{PG3} = f_{E,Emp} - f_{E,Eno} - f_{D,PG3} - f_{BM,PG3}$$
$$\dot{x}_{PYR} = f_{E,MaeAB} + f_{E,PykF} - f_{E,Pdh} - f_{E,PpsA} + f_{PTS,r_1}$$
$$- f_{D,PYR} - f_{BM,PYR}$$

Enzymes and PTS proteins

$$\dot{x}_{AceA} = f_{G,aceA} - f_{D,AceA}$$
$$\dot{x}_{AceB} = f_{G,aceB} - f_{D,AceB}$$
$$\dot{x}_{AceK} = f_{G,aceK} - f_{D,AceK}$$
$$\dot{x}_{Acoa2act} = f_{G,acoa2act} - f_{D,Acoa2act}$$
$$\dot{x}_{Acs} = f_{G,acs} - f_{D,Acs}$$
$$\dot{x}_{Akg2mal} = f_{G,akg2mal} - f_{D,Akg2mal}$$
$$\dot{x}_{CAMPdegr} = f_{G,campdegr} - f_{D,CAMPdegr}$$
$$\dot{x}_{Cya} = f_{G,cya} - f_{D,Cya}$$

$$\dot{x}_{Emp} = f_{G,emp} - f_{D,Emp}$$
$$\dot{x}_{Eno} = f_{G,eno} - f_{D,Eno}$$
$$\dot{x}_{Fdp} = f_{G,fdp} - f_{D,Fdp}$$
$$\dot{x}_{GltA} = f_{G,gltA} - f_{D,GltA}$$
$$\dot{x}_{Icd} = f_{G,icd} - f_{D,Icd} - f_{E,AceK-Ki} + f_{E,AceK-Ph}$$
$$\dot{x}_{Icd-P} = -f_{D,Icd-P} + f_{E,AceK-Ki} - f_{E,AceK-Ph}$$
$$\dot{x}_{Mdh} = f_{G,mdh} - f_{D,Mdh}$$
$$\dot{x}_{MaeAB} = f_{G,maeAB} - f_{D,MaeAB}$$
$$\dot{x}_{PckA} = f_{G,pckA} - f_{D,PckA}$$
$$\dot{x}_{Pdh} = f_{G,pdh} - f_{D,Pdh}$$
$$\dot{x}_{PfkA} = f_{G,pfkA} - f_{D,PfkA}$$
$$\dot{x}_{Ppc} = f_{G,ppc} - f_{D,Ppc}$$
$$\dot{x}_{PpsA} = f_{G,ppsA} - f_{D,PpsA}$$
$$\dot{x}_{PykF} = f_{G,pykF} - f_{D,PykF}$$
$$\dot{x}_{EIIA} = f_{G,eiia} - f_{D,EIIA} - f_{PTS,r_1} + f_{PTS,r_4}$$
$$\dot{x}_{EIIA-P} = -f_{D,EIIA-P} + f_{PTS,r_1} - f_{PTS,r_4}$$
$$\dot{x}_{EIICB} = f_{G,eiicb} - f_{D,EIICB}$$

Transcription factors
$$\dot{x}_{Cra} = f_{G,cra} - f_{D,Cra} - f_{TF,Cra}$$
$$\dot{x}_{CraFBP} = -f_{D,CraFBP} + f_{TF,Cra}$$
$$\dot{x}_{Crp} = f_{G,crp} - f_{D,Crp} - f_{TF,Crp}$$
$$\dot{x}_{CrpcAMP} = -f_{D,CrpcAMP} + f_{TF,Crp}$$
$$\dot{x}_{IclR} = f_{G,iclR} - f_{D,IclR}$$
$$\dot{x}_{PdhR} = f_{G,pdhR} - f_{D,PdhR} - f_{TF,PdhR}$$
$$\dot{x}_{PdhRPYR} = -f_{D,PdhRPYR} + f_{TF,PdhR}$$

The 109 individual rates $f(\mathbf{x}, \mathbf{p})$ that appear in these differential equations are structured in the six units

1. Cell growth and carbon source dynamics

2. Metabolic reactions and protein phosphorylations

3. Transcription factor- metabolite bindings

4. Gene expression

5. Dilution and degradation of compounds

6. Biomass generation and growth rate calculation.

The following six sections describe in detail the modeling of these units and present the mechanistic rate equations $\mathbf{f}(\mathbf{x}, \mathbf{p})$ belonging to these units.

3.6.1.4 Cell growth and carbon source dynamics

The closed environment is fully described by the extracellular concentrations of the two carbon sources glucose (x_{GLC}) and acetate (x_{ACT}), both of the unit $g\,l^{-1}$. The size of the *E. coli* cell population within this environment is given by its biomass concentration (x_{OD}).

To model the interaction of the *E. coli* cells with their environment, the substrate exchange rates between the two compartments need to be quantified. Rates in the compartment containing the intracellular processes are normalized to cell dry weight and are of the unit $\mu mol\,(g_{DW}\,s)^{-1}$; rates in the compartment containing the environment scale with the biomass concentration of the cell population and are of the unit $g_{SUBSTRATE}\,(l \cdot s)^{-1}$. Therefore, to quantify the substrate exchange rates of the whole cell population, the substrate exchange rates of dry weight-normalized cells, calculated in Section 3.6.1.5, need to be scaled with the biomass concentration and converted to the proper units. The parameters needed to quantify this conversion are the molar mass of glucose $p_{ENV,M_{GLC}}$, the molar mass of acetate $p_{ENV,M_{ACT}}$, and the remaining unit conversions subsumed in $p_{ENV,UC}$.

The rate equations describing cell growth and the production and consumption of environmental carbon sources are:

$$f_{ENV,growth} = \mu\,x_{OD}$$
$$f_{ENV,GLCup} = p_{ENV,M_{GLC}}\,p_{ENV,UC}\,x_{OD}\,f_{PTS,r4}$$
$$f_{ENV,ACTup} = p_{ENV,M_{ACT}}\,p_{ENV,UC}\,x_{OD}\,f_{E,Acs}$$
$$f_{ENV,ACTex} = p_{ENV,M_{ACT}}\,p_{ENV,UC}\,x_{OD}\,f_{E,Acoa2act}$$

3.6.1.5 Metabolic reactions and protein phosphorylations

All metabolic reactions occurring in the model are catalyzed by enzymes. Also, the phosphorylation and dephosphorylation of the enzyme Icd is catalyzed by the two enzymatic reactions of the enzyme AceK. Further, the uptake and phosphorylation of glucose is achieved with the help of the protein phosphorylations occurring in the phosphotransferase system (PTS). Table 3.1 lists the chosen types of kinetic equations describing these reactions. The full equations follow this table.

The metabolic reactions that produce biomass from the seven modeled precursor metabolites are presented in Section 3.6.1.9.

Table 3.1: Overview of the chosen types of kinetic rate equations for the metabolic reactions and protein phosphorylation processes. The following abbreviations are used: MM – Michaelis-Menten kinetics, revMM – reversible Michaelis-Menten kinetics, 2S-MM – Two-substrate Michaelis-Menten kinetics, MWC – Monod-Wyman-Changeux kinetics, rev2S-1st – Reversible two-substrate first order kinetics, (A) – Activator, (I) – Inhibitor.

Rate	Type	Substrate(s)	Effectors
$f_{E,AceA}$	MWC	ICT	AKG (I), PEP (I), PG3 (I)
$f_{E,AceB}$	2S-MM	GLX, ACoA	—
$f_{E,AceK-Ki}$	MWC	Icd	AKG (I), GLX (I), ICT (I), OAA (I), PEP(I), PG3 (I), PYR (I)
$f_{E,AceK-Ph}$	MWC	Icd-P	AKG (A), OAA (A), PEP (A), PG3 (A), PYR (A)
$f_{E,Acoa2act}$	MWC	ACoA	PYR (A)
$f_{E,Acs}$	MM	ACT	—
$f_{E,Akg2mal}$	MM	AKG	—
$f_{E,CAMPdegr}$	MM	cAMP	—
$f_{E,Cya}$	MM	EIIA-P *	— *
$f_{E,Emp}$	revMM	FBP, PG3	—
$f_{E,Eno}$	revMM	PG3, PEP	—
$f_{E,Fdp}$	MWC	FBP	PEP (A)
$f_{E,GltA}$	2S-MM	ACoA, OAA	AKG (I, competitive)
$f_{E,Icd}$	MWC	ICT	PEP (I)
$f_{E,MaeAB}$	MWC	MAL	ACoA (I), cAMP (I)
$f_{E,Mdh}$	Hill	MAL	—
$f_{E,PckA}$	MM	OAA	PEP (I, competitive)
$f_{E,Pdh}$	MWC	PYR	GLX (I), PYR (I)
$f_{E,PfkA}$	MWC	G6P	PEP (I)
$f_{E,Ppc}$	MWC	PEP	FBP (A)
$f_{E,PpsA}$	MWC	PYR	PEP (I)
$f_{E,PykF}$	MWC	PEP	FBP (A)
f_{PTS,r_1}	rev2S-1st	PEP, EIIA; PYR, EIIA-P	—
f_{PTS,r_4}	MM	EIIA-P, GLC	—

* The Cya reaction is activated by EIIA-P and produces cAMP from void. Because 'void' cannot be a substrate in a mechanistic equation, the Cya reaction is modeled with the activator EIIA-P as substrate. This approach follows the one chosen by [24].

$$f_{E,AceA} = x_{AceA}\, p_{AceA,k_{cat}} \frac{x_{ICT}}{p_{AceA,K_{ICT}}} \left(1 + \frac{x_{ICT}}{p_{AceA,K_{ICT}}}\right)^{p_{AceA,n}-1}$$

$$\left[\left(1 + \frac{x_{ICT}}{p_{AceA,K_{ICT}}}\right)^{p_{AceA,n}} + p_{AceA,L}\left(1 + \frac{x_{PEP}}{p_{AceA,K_{PEP}}}\right.\right.$$

$$\left.\left. + \frac{x_{PG3}}{p_{AceA,K_{PG3}}} + \frac{x_{AKG}}{p_{AceA,K_{AKG}}}\right)^{p_{AceA,n}}\right]^{-1}$$

$$f_{E,AceB} = x_{AceB}\, p_{AceB,k_{cat}}\, x_{GLX}\, x_{ACoA} \left(p_{AceB,K_{GLX}ACoA}\, p_{AceB,K_{ACoA}}\right.$$

$$\left. + p_{AceB,K_{ACoA}}\, x_{GLX} + p_{AceB,K_{GLX}}\, x_{ACoA} + x_{GLX}\, x_{ACoA}\right)^{-1}$$

$$f_{E,AceK-Ki} = x_{AceK}\, p_{AceK,k_{cat,ki}} \frac{x_{Icd}}{p_{AceK,K_{Icd}}} \left(1 + \frac{x_{Icd}}{p_{AceK,K_{Icd}}}\right)^{p_{AceK,n}-1}$$

$$\left[\left(1 + \frac{x_{Icd}}{p_{AceK,K_{Icd}}}\right)^{p_{AceK,n}} + p_{AceK,L}\left(1 + \frac{x_{ICT}}{p_{AceK,K_{ICT}}}\right.\right.$$

$$+ \frac{x_{GLX}}{p_{AceK,K_{GLX}}} + \frac{x_{OAA}}{p_{AceK,K_{OAA}}} + \frac{x_{AKG}}{p_{AceK,K_{AKG}}} + \frac{x_{PEP}}{p_{AceK,K_{PEP}}}$$

$$\left.\left. + \frac{x_{PG3}}{p_{AceK,K_{PG3}}} + \frac{x_{PYR}}{p_{AceK,K_{PYR}}}\right)^{p_{AceK,n}}\right]^{-1}$$

$$f_{E,AceK-Ph} = x_{AceK}\, p_{AceK,k_{cat,ph}} \frac{x_{Icd-P}}{p_{AceK,K_{Icd-P}}}$$

$$\left(1 + \frac{x_{Icd-P}}{p_{AceK,K_{Icd-P}}}\right)^{p_{AceK,n}-1} \left[\left(1 + \frac{x_{Icd-P}}{p_{AceK,K_{Icd-P}}}\right)^{p_{AceK,n}}\right.$$

$$+ p_{AceK,L}\left(1 + \frac{x_{OAA}}{p_{AceK,K_{OAA}}} + \frac{x_{AKG}}{p_{AceK,K_{AKG}}} + \frac{x_{PEP}}{p_{AceK,K_{PEP}}}\right.$$

$$\left.\left. + \frac{x_{PG3}}{p_{AceK,K_{PG3}}} + \frac{x_{PYR}}{p_{AceK,K_{PYR}}}\right)^{-p_{AceK,n}}\right]^{-1}$$

$$f_{E,Acoa2act} = x_{Acoa2act}\, p_{Acoa2act,k_{cat}} \frac{x_{ACoA}}{p_{Acoa2act,K_{ACoA}}}$$

$$\left(1 + \frac{x_{ACoA}}{p_{Acoa2act,K_{ACoA}}}\right)^{p_{Acoa2act,n}-1}$$

$$\left[\left(1 + \frac{x_{ACoA}}{p_{Acoa2act,K_{ACoA}}}\right)^{p_{Acoa2act,n}}\right.$$

$$+ p_{Acoa2act,L} \left(1 + \frac{x_{PYR}}{p_{Acoa2act,K_{PYR}}}\right)^{-p_{Acoa2act,n}} \Bigg]^{-1}$$

$$f_{E,Acs} = \frac{x_{Acs}\, p_{Acs,k_{cat}}\, x_{ACT}}{x_{ACT} + p_{Acs,K_{ACT}}}$$

$$f_{E,Akg2mal} = \frac{x_{Akg2mal}\, p_{Akg2mal,k_{cat}}\, x_{AKG}}{x_{AKG} + p_{Akg2mal,K_{AKG}}}$$

$$f_{E,CAMPdegr} = \frac{p_{CAMPdegr,k_{cat}}\, x_{CAMPdegr}\, x_{cAMP}}{x_{cAMP} + p_{CAMPdegr,K_{cAMP}}}$$

$$f_{E,Cya} = \frac{p_{Cya,k_{cat}}\, x_{Cya}\, x_{EIIA-P}}{x_{EIIA-P} + p_{Cya,K_{EIIA-P}}}$$

$$f_{E,Emp} = \frac{x_{Emp}\left(p_{Emp,k_{cat,f}} \frac{x_{FBP}}{p_{Emp,K_{FBP}}} - p_{Emp,k_{cat,r}} \frac{x_{PG3}}{p_{Emp,K_{PG3}}}\right)}{1 + \frac{x_{FBP}}{p_{Emp,K_{FBP}}} + \frac{x_{PG3}}{p_{Emp,K_{PG3}}}}$$

$$f_{E,Eno} = \frac{x_{Eno}\left(p_{Eno,k_{cat,f}} \frac{x_{PG3}}{p_{Eno,K_{PG3}}} - p_{Eno,k_{cat,r}} \frac{x_{PEP}}{p_{Eno,K_{PEP}}}\right)}{1 + \frac{x_{PG3}}{p_{Eno,K_{PG3}}} + \frac{x_{PEP}}{p_{Eno,K_{PEP}}}}$$

$$f_{E,Fdp} = \frac{x_{Fdp}\, p_{Fdp,k_{cat}} \frac{x_{FBP}}{p_{Fdp,K_{FBP}}}\left(1 + \frac{x_{FBP}}{p_{Fdp,K_{FBP}}}\right)^{p_{Fdp,n}-1}}{\left(1 + \frac{x_{FBP}}{p_{Fdp,K_{FBP}}}\right)^{p_{Fdp,n}} + p_{Fdp,L}\left(1 + \frac{x_{PEP}}{p_{Fdp,K_{PEP}}}\right)^{-p_{Fdp,n}}}$$

$$f_{E,GltA} = x_{GltA}\, p_{GltA,k_{cat}}\, x_{OAA}\, x_{ACoA}\left[\left(1 + \frac{x_{AKG}}{p_{GltA,K_{AKG}}}\right)\right.$$

$$p_{GltA,K_{OAAACoA}}\, p_{GltA,K_{ACoA}} + p_{GltA,K_{ACoA}}\, x_{OAA}$$

$$+ \left(1 + \frac{x_{AKG}}{p_{GltA,K_{AKG}}}\right) p_{GltA,K_{OAA}}\, x_{ACoA} + x_{OAA}\, x_{ACoA} \Bigg]^{-1}$$

$$f_{E,Icd} = \frac{x_{Icd}\, p_{Icd,k_{cat}} \frac{x_{ICT}}{p_{Icd,K_{ICT}}}\left(1 + \frac{x_{ICT}}{p_{Icd,K_{ICT}}}\right)^{p_{Icd,n}-1}}{\left(1 + \frac{x_{ICT}}{p_{Icd,K_{ICT}}}\right)^{p_{Icd,n}} + p_{Icd,L}\left(1 + \frac{x_{PEP}}{p_{Icd,K_{PEP}}}\right)^{p_{Icd,n}}}$$

$$f_{E,MaeAB} = x_{MaeAB}\, p_{MaeAB,k_{cat}} \frac{x_{MAL}}{p_{MaeAB,K_{MAL}}}$$

$$\left(1 + \frac{x_{MAL}}{p_{MaeAB,K_{MAL}}}\right)^{p_{MaeAB,n}-1}$$

$$\left[\left(1 + \frac{x_{MAL}}{p_{MaeAB,K_{MAL}}}\right)^{p_{MaeAB,n}} + p_{MaeAB,L}\right.$$

$$\left(1 + \frac{x_{ACoA}}{p_{MaeAB,K_{ACoA}}} + \frac{x_{cAMP}}{p_{MaeAB,K_{cAMP}}}\right)^{p_{MaeAB,n}} \Bigg]^{-1}$$

$$f_{E,Mdh} = \frac{x_{Mdh}\, p_{Mdh,k_{cat}}\, x_{MAL}{}^{p_{Mdh,n}}}{x_{MAL}{}^{p_{Mdh,n}} + p_{Mdh,K_{MAL}}{}^{p_{Mdh,n}}}$$

$$f_{E,PckA} = \frac{x_{PckA}\, p_{PckA,k_{cat}}\, x_{OAA}}{x_{OAA} + p_{PckA,K_{OAA}}\left(1 + \frac{x_{PEP}}{p_{PckA,K_{PEP}}}\right)}$$

$$f_{E,Pdh} = x_{Pdh}\, p_{Pdh,k_{cat}} \frac{x_{PYR}}{p_{Pdh,K_{PYR}}} \left(1 + \frac{x_{PYR}}{p_{Pdh,K_{PYR}}}\right)^{p_{Pdh,n}-1}$$

$$\left[\left(1 + \frac{x_{PYR}}{p_{Pdh,K_{PYR}}}\right)^{p_{Pdh,n}}\right.$$

$$\left.+ p_{Pdh,L}\left(1 + \frac{x_{GLX}}{p_{Pdh,K_{GLX}}} + \frac{x_{PYR}}{p_{Pdh,K_{I,PYR}}}\right)^{p_{Pdh,n}}\right]^{-1}$$

$$f_{E,PfkA} = x_{PfkA}\, p_{PfkA,k_{cat}} \frac{x_{G6P}}{p_{PfkA,K_{G6P}}} \left(1 + \frac{x_{G6P}}{p_{PfkA,K_{G6P}}}\right)^{p_{PfkA,n}-1}$$

$$\left[\left(1 + \frac{x_{G6P}}{p_{PfkA,K_{G6P}}}\right)^{p_{PfkA,n}}\right.$$

$$\left.+ p_{PfkA,L}\left(1 + \frac{x_{PEP}}{p_{PfkA,K_{PEP}}}\right)^{p_{PfkA,n}}\right]^{-1}$$

$$f_{E,Ppc} = \frac{x_{Ppc}\, p_{Ppc,k_{cat}} \frac{x_{PEP}}{p_{Ppc,K_{PEP}}} \left(1 + \frac{x_{PEP}}{p_{Ppc,K_{PEP}}}\right)^{p_{Ppc,n}-1}}{\left(1 + \frac{x_{PEP}}{p_{Ppc,K_{PEP}}}\right)^{p_{Ppc,n}} + p_{Ppc,L}\left(1 + \frac{x_{FBP}}{p_{Ppc,K_{FBP}}}\right)^{-p_{Ppc,n}}}$$

$$f_{E,PpsA} = x_{PpsA}\, p_{PpsA,k_{cat}} \frac{x_{PYR}}{p_{PpsA,K_{PYR}}} \left(1 + \frac{x_{PYR}}{p_{PpsA,K_{PYR}}}\right)^{p_{PpsA,n}-1}$$

$$\left[\left(1 + \frac{x_{PYR}}{p_{PpsA,K_{PYR}}}\right)^{p_{PpsA,n}}\right.$$

$$\left.+ p_{PpsA,L}\left(1 + \frac{x_{PEP}}{p_{PpsA,K_{PEP}}}\right)^{p_{PpsA,n}}\right]^{-1}$$

$$f_{E,PykF} = x_{PykF}\, p_{PykF,k_{cat}} \frac{x_{PEP}}{p_{PykF,K_{PEP}}} \left(1 + \frac{x_{PEP}}{p_{PykF,K_{PEP}}}\right)^{p_{PykF,n}-1}$$

$$\left[\left(1 + \frac{x_{PEP}}{p_{PykF,K_{PEP}}}\right)^{p_{PykF,n}}\right.$$

$$\left.+ p_{PykF,L}\left(1 + \frac{x_{FBP}}{p_{PykF,K_{FBP}}}\right)^{-p_{PykF,n}}\right]^{-1}$$

$$f_{PTS,r_1} = p_{PTS,k_1}\, x_{PEP}\, x_{EIIA} - p_{PTS,km_1}\, x_{PYR}\, x_{EIIA-P}$$

$$f_{PTS,r_4} = \frac{p_{PTS,k_4}\, x_{EIICB}\, x_{EIIA_P}\, x_{GLC}}{(p_{PTS,K_{EIIA}} + x_{EIIA-P})(p_{PTS,K_{GLC}} + x_{GLC})}$$

3.6.1.6 Transcription factor–metabolite interactions

The interactions of the transcription factors Cra, Crp and PdhR with the respective metabolites FBP, cAMP, and PYR are modeled with single rate equations each, which combine the association and dissociation rates into one rate. A positive net rate results when the association rate is higher than the dissociation rate, a negative net rate when the reverse is the case. The net rate can be determined by first calculating the association and dissociation rates separately and then joining them to the net rate. This is essentially what the following equation do; however,

the calculation of the net rate in these equations is rearranged such that the net rate is determined through scaling (and unit conversion of) the deviation of the actual binding state from the steady state level. The resulting rate equations are:

$$f_{TF,Cra} = p_{Cra,scale} \left[\frac{(x_{Cra} + x_{CraFBP})\, x_{FBP}{}^{p_{Cra,n}}}{x_{FBP}{}^{p_{Cra,n}} + p_{Cra,K_{FBP}}{}^{p_{Cra,n}}} - x_{CraFBP} \right]$$

$$f_{TF,Crp} = p_{Crp,scale} \left[\frac{(x_{Crp} + x_{CrpcAMP})\, x_{cAMP}{}^{p_{Crp,n}}}{x_{cAMP}{}^{p_{Crp,n}} + p_{Crp,K_{cAMP}}{}^{p_{Crp,n}}} - x_{CrpcAMP} \right]$$

$$f_{TF,PdhR} = p_{PdhR,scale} \left[\frac{(x_{PdhR} + x_{PdhRPYR})\, x_{PYR}{}^{p_{PdhR,n}}}{x_{PYR}{}^{p_{PdhR,n}} + p_{PdhR,K_{PYR}}{}^{p_{PdhR,n}}} - x_{PdhRPYR} \right].$$

The binding of Cra to FBP is assumed to be cooperative with degree $n = 2$, because Cra is structurally similar to LacI (both proteins belong to the GalR/LacI family of transcriptional regulators), which binds to lactose with $n \approx 2$ [231]. The bindings of the other two interactions are assumed to be non-cooperative ($n = 1$), reducing their Hill-type kinetics to Michaelis-Menten-type.

The binding of the transcription factor IclR to the metabolites GLX and PYR is modeled jointly with the binding of IclR to the promoter region of the *aceBAK* operon, and is presented in Section 3.6.1.7. This modeling avoids the introduction of three additional states (IclR bound to either GLX or PYR or both) and the rates between these, and is straightforward because IclR represses only that one operon.

3.6.1.7 Gene expression

This section first describes the modeling of regulated gene expression, and then of unregulated gene expression. It continues with explaining how the growth rate dependency of gene expression was modeled, and, finally, lists the resulting rate equations.

Expression of regulated genes The production rates of most modeled proteins are regulated by at least one of the four modeled transcription factors (known transcriptional regulations exerted by other transcription factors have been ignored, because these are outside the chosen system boundary, see Section 3.6.1.1).

With two exceptions, the expression of a gene is regulated by only one transcription factor. In this case, the impact of transcriptional regulation on the protein production rate is modeled as a weighed sum of two production rates. These two production rates are quantified by the parameters $p_{<gene>,v_{<TF>},unbound}$ and $p_{<gene>,v_{<TF>},bound}$, respectively. The parameter $p_{<gene>,v_{<TF>},unbound}$ quantifies the production rate when the promoter is not occupied by the regulating transcription factor; the parameter $p_{<gene>,v_{<TF>},bound}$ quantifies the production rate

when the promoter is occupied by the regulating transcription factor ('bound'). These two production rates are weighed by the occupancy of the promoter with the regulating transcription factor. To calculate the occupancy and thus the weighting factor, either a Michaelis-Menten or a Hill kinetics was used, with the regulating transcription factor's active form, i.e. Cra, Crp-cAMP, or PdhR (as opposed to their inactive forms Cra-FBP, Crp, and PdhR-PYR), as substrate. A Michaelis-Menten kinetics was preferred unless it was unable to reproduce the steady state rates on glucose and acetate; in these cases, we chose a Hill kinetics instead.

One of the two promoters regulated by more than one TF is that of the super-enzyme 'Emp'. The production of this enzyme, which represents the section of the Emden-Meyerhoff pathway between FBP and PG3, is transcriptionally activated by Crp and repressed by Cra. The contributions of these two transcriptions factors to the overall expression rate are modeled as additive (as opposed to multiplicative).

The second promoter regulated by more than one TF is that of the *aceBAK* operon, which controls the production of the three enzymes AceA, AceB and AceK. The ratio of these three enzymes' production rates $f_{G,aceA} : f_{G,aceB} : f_{G,aceK}$ is approximately equal to $1 : 0.3 : 0.03$ [38]. In the model, this ratio is ensured by first calculating the gene expression rate of *aceA*, and then scaling that rate with 0.3 and 0.03 to obtain the expression rates of the *aceB* and *aceK* genes, respectively. The expression of the *aceA* gene is jointly regulated by the three transcription factors Cra, Crp and IclR. The impact of these three individual regulations on the overall expression of the *aceA* gene is modeled as additive (as opposed to multiplicative).

The third of the three additive contributions to the overall expression rate of the *aceA* gene is its repression by the transcription factor IclR. The two ligands recognized by IclR act as corepressor (PYR) or activator (GLX) on the transcription of the *aceBAK* operon [118], in the manner depicted in Figure 3.7. Thus, if one monitors the binding of DNA to the tetramer (T) in the absence and presence of pyruvate (P), which is opposed by glyoxylate (G), which binds to the dimer (D), the observed response Y is

$$
Y = \frac{X\left[\frac{[DNA]}{K_{DNA}}\left(1 + \frac{[P]}{K_P}\right)\right]}{\left(1 + \frac{1}{L}\left[\frac{[G]}{K_G}\left(1 + \frac{[G]}{K_G}\right)\right] + \frac{[DNA]}{K_{DNA}} + \frac{[P]}{K_P} + \frac{[DNA][P]}{K_{DNA}K_P'}\right)} \quad ,
$$

with X the maximum response and the equilibria defined in the caption of Figure 3.7.

Unlike the interaction of the other transcription factors with their respective ligands (see Section 3.6.1.6), the mechanistic description of IclR activity is directly incorporated into the expression rate equation of the *aceA* gene.

Expression of unregulated genes Constitutively expressed enzymes, PTS proteins and transcription factors are, with two exceptions, modeled with a constant
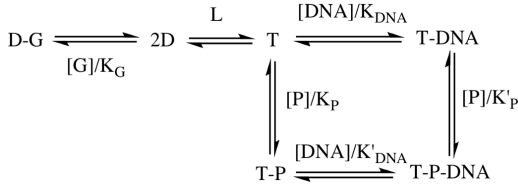
Figure 3.7: Graphical representation of the interactions described by [118], from which the kinetic equation describing these interactions is derived. IcIR, as dimer (D) or tetramer (T), is being titrated by DNA, glyoxylate (G), and/or pyruvate (P). The equilibria needed for the derivation are $K_{DNA} = \frac{[T][DNA]}{[T-DNA]}$, $K_P = \frac{[T][P]}{[T-P]}$, $K'_P = \frac{[T-DNA][P]}{[T-P-DNA]}$, $K_G = \frac{[G][D]}{[G-D]}$ and $L = \frac{[T]}{[D]^2}$. As only three out of the four equilibria around a cycle are needed to define it, the remaining equilibrium $K'_{DNA} = \frac{[T-P][DNA]}{[T-P-DNA]}$ is not needed for the derivation of the kinetic equation.

concentration. The production (see the Equations below), dilution and degradation rates (see Section 3.6.1.8) of these proteins are set to zero; therefore, the constant protein concentrations are determined by the initial conditions.

The two exceptions are the enzymes Ppc and MaeAB (which lumps the two isoenzymes MaeA and MaeB). Neither of the genes *ppc*, *maeA* nor *maeB* is known to be regulated [92], yet, the respective mRNA concentrations were found to be markedly distinct for growth on glucose and acetate [142]. As it has been observed that protein and mRNA abundances in *E. coli* cells are significantly correlated [80], the respective protein concentrations are very likely also markedly distinct. Accordingly, the measured steady state fluxes on glucose and acetate could not be simultaneously reproduced with constant Ppc and MaeAB concentrations.

Therefore, we set up the model equations in such a way that the differences in the Ppc and MaeAB concentrations on glucose and acetate are proportional to the measured differences in the respective mRNA concentrations. To realize such concentration changes as a response to changes in the availability of glucose and acetate, we chose the following implementation. First, for each of the two enzymes Ppc and MaeAB, the model calculates the sum of dilution and degradation rates that would occur if the Ppc and MaeAB concentrations were in steady state (for the calculation of the carbon source-dependent steady state concentrations SS_{Ppc} and SS_{MaeAB} of Ppc and MaeAB, respectively, see Section 3.6.1.9). Then, the actual production rates of Ppc and MaeAB are set to equal their steady state degradation+dilution rates. The effect of this modeling is that the actual, possibly out-of-steady state protein concentrations x_{MaeAB} and x_{Ppc} approach their calculated, carbon source-dependent steady state concentrations. With this workaround, the model is capable to reproduce the measured steady state fluxes for growth on glucose and acetate.

Growth rate–dependency of gene expression Gene expression is growth rate–dependent due to growth-rate dependent concentrations of DNA polymerases and ribosomes [33]. The growth-rate dependent efficiency of the gene expression machinery is modeled as a linear function of the growth rate ($p_{BM,k_{expr}} \cdot \mu$), and the expression rates of all regulated genes are multiplied by this function.

We found that when this growth rate-dependency is neglected, the measured steady state concentrations of the proteins cannot be reproduced on both glucose and acetate simultaneously. The reason for such failed reproduction is that to reproduce the measured steady states where the protein production and dilution+degradation rates are equal, the production rates must be capable to balance the dilution rates for both high and low growth rates. However, the model covers a wide range of growth rates (the growth rate on glucose is approximately three times higher than the growth rate on acetate); therefore, the inherently growth-rate dependent dilution rates can vary over a wide range. To make the production rates capable to balance the dilution rates for both high and low growth rates, the growth-rate dependency of gene expression must be considered as well.

Rate equations for gene expression Table 3.2 lists the chosen types of rate equations to describe the process of gene expression. The full equations follow this table.

Table 3.2: Overview of the chosen types of rate equations to describe the process of gene expression. The following abbreviations are used: MM – Michaelis-Menten kinetics plus basal expression term, Hill – Hill kinetics plus basal expression term, (A) – transcriptional activator, (R) – transcriptional repressor.

Rate	Type	Regulators
$f_{G,aceA}$	special, see text	Cra (A), Crp-cAMP (R), IclR (R)
$f_{G,aceB}$	$= 0.3 \cdot f_{G,aceA}$	—
$f_{G,aceK}$	$= 0.03 \cdot f_{G,aceA}$	—
$f_{G,acoa2act}$	$= 0$	—
$f_{G,acs}$	Hill	Crp-cAMP (A)
$f_{G,akg2mal}$	Hill	Crp-cAMP (A)
$f_{G,campdegr}$	$= 0$	—
$f_{G,cra}$	$= 0$	—
$f_{G,crp}$	$= 0$	—
$f_{G,cya}$	$= 0$	—
$f_{G,emp}$	MM	Cra (R), Crp-cAMP (A)
$f_{G,eno}$	MM	Cra (R)
$f_{G,fdp}$	MM	Cra (A)
$f_{G,gltA}$	Hill	Crp-cAMP (A)
$f_{G,icd}$	MM	Cra (A)

...Table 3.2 continued.

Rate	Type	Regulators
$f_{G,iclr}$	$= 0$	—
$f_{G,maeAB}$	special, see text	—
$f_{G,mdh}$	MM	Crp-cAMP (A)
$f_{G,pckA}$	MM	Cra (A)
$f_{G,pdh}$	MM	PdhR (R)
$f_{G,pdhr}$	$= 0$	—
$f_{G,pfkA}$	MM	Cra (R)
$f_{G,ppc}$	special, see text	—
$f_{G,ppsA}$	MM	Cra (A)
$f_{G,pykF}$	MM	Cra (R)
$f_{G,EIIA}$	$= 0$	—
$f_{G,EIICB}$	$= 0$	—

$$
f_{G,aceA} = \mu \, p_{BM,k_{expr}} \left\{ \left(1 - \frac{x_{Cra}}{x_{Cra} + p_{aceBAK,K_{Cra}}} \right) \right.
$$

$$
p_{aceBAK,v_{Cra,unbound}}
$$

$$
+ \frac{x_{Cra}}{x_{Cra} + p_{aceBAK,K_{Cra}}} \, p_{aceBAK,v_{Cra,bound}}
$$

$$
+ \left(1 - \frac{x_{CrpcAMP}}{x_{CrpcAMP} + p_{aceBAK,K_{Crp}}} \right) p_{aceBAK,v_{Crp,unbound}}
$$

$$
+ \frac{x_{CrpcAMP}}{x_{CrpcAMP} + p_{aceBAK,K_{Crp}}} \, p_{aceBAK,v_{Crp,bound}}
$$

$$
+ \left[1 - \frac{p_{aceBAK,DNA}}{p_{aceBAK,K_{DNA}}} \left(1 + \frac{x_{PYR}}{p_{aceBAK,K_{PYRprime}}} \right) \right.
$$

$$
\left(1 + \frac{1}{L} \left(\frac{x_{GLX}}{p_{aceBAK,K_{GLX}}} \right) \right) \left(1 + \frac{x_{GLX}}{p_{aceBAK,K_{GLX}}} \right)
$$

$$
+ \frac{p_{aceBAK,DNA}}{p_{aceBAK,K_{DNA}}} + \frac{x_{PYR}}{p_{aceBAK,K_{PYR}}} + \frac{p_{aceBAK,DNA}}{p_{aceBAK,K_{DNA}}}
$$

$$
\left. \left. \frac{x_{PYR}}{p_{aceBAK,K_{PYRprime}}} \right)^{-1} \right] p_{aceBAK,k_{cat,IclR}} \, x_{IclR} \right\}
$$

$$
f_{G,aceB} = p_{aceBAK,aceBfactor} \, f_{G,aceA}
$$

$$
f_{G,aceK} = p_{aceBAK,aceKfactor} \, f_{G,aceA}
$$

$$
f_{G,acoa2act} = 0
$$

$$f_{G,acs} = \mu\, p_{BM,k_{expr}} \left[\left(1 - \frac{x_{CrpcAMP}{}^{p_{acs,n}}}{x_{CrpcAMP}{}^{p_{acs,n}} + p_{acs,K_{Crp}}{}^{p_{acs,n}}} \right) \right.$$

$$p_{acs,v_{Crp,unbound}}$$

$$\left. + \frac{x_{CrpcAMP}{}^{p_{acs,n}}}{x_{CrpcAMP}{}^{p_{acs,n}} + p_{acs,K_{Crp}}{}^{p_{acs,n}}}\, p_{acs,v_{Crp,bound}} \right]$$

$$f_{G,akg2mal} = \mu\, p_{BM,k_{expr}} \left\{ p_{akg2mal,v_{Crp,unbound}} \left[1 - x_{CrpcAMP}{}^{p_{akg2mal,n}} \right. \right.$$

$$\left. \left(x_{CrpcAMP}{}^{p_{akg2mal,n}} + p_{akg2mal,K_{Crp}}{}^{p_{akg2mal,n}} \right)^{-1} \right]$$

$$+ \frac{x_{CrpcAMP}{}^{p_{akg2mal,n}}}{x_{CrpcAMP}{}^{p_{akg2mal,n}} + p_{akg2mal,K_{Crp}}{}^{p_{akg2mal,n}}}$$

$$p_{akg2mal,v_{Crp,bound}} \Big\}$$

$$f_{G,campdegr} = 0$$

$$f_{G,cra} = 0$$

$$f_{G,crp} = 0$$

$$f_{G,cya} = 0$$

$$f_{G,emp} = \mu\, p_{BM,k_{expr}} \left[\left(1 - \frac{x_{Cra}}{x_{Cra} + p_{emp,K_{Cra}}} \right) p_{emp,v_{Cra,unbound}} \right.$$

$$+ \frac{x_{Cra}}{x_{Cra} + p_{emp,K_{Cra}}}\, p_{emp,v_{Cra,bound}}$$

$$+ \left(1 - \frac{x_{CrpcAMP}}{x_{CrpcAMP} + p_{emp,K_{Crp}}} \right) p_{emp,v_{Crp,unbound}}$$

$$\left. + \frac{x_{CrpcAMP}}{x_{CrpcAMP} + p_{emp,K_{Crp}}}\, p_{emp,v_{Crp,bound}} \right]$$

$$f_{G,eno} = \mu\, p_{BM,k_{expr}} \left[\left(1 - \frac{x_{Cra}}{x_{Cra} + p_{eno,K_{Cra}}} \right) p_{eno,v_{Cra,unbound}} \right.$$

$$\left. + \frac{x_{Cra}}{x_{Cra} + p_{eno,K_{Cra}}}\, p_{eno,v_{Cra,bound}} \right]$$

$$f_{G,fdp} = \mu\, p_{BM,k_{expr}} \left[\left(1 - \frac{x_{Cra}}{x_{Cra} + p_{fdp,K_{Cra}}} \right) p_{fdp,v_{Cra,unbound}} \right.$$

$$\left. + \frac{x_{Cra}}{x_{Cra} + p_{fdp,K_{Cra}}}\, p_{fdp,v_{Cra,bound}} \right]$$

$$f_{G,gltA} = \mu\, p_{BM,k_{expr}} \left[\left(1 - \frac{x_{CrpcAMP}{}^{p_{gltA,n}}}{x_{CrpcAMP}{}^{p_{gltA,n}} + p_{gltA,K_{Crp}}{}^{p_{gltA,n}}}\right) \right.$$

$$\left. p_{gltA,v_{Crp,unbound}} + \frac{x_{CrpcAMP}{}^{p_{gltA,n}}}{x_{CrpcAMP}{}^{p_{gltA,n}} + p_{gltA,K_{Crp}}{}^{p_{gltA,n}}}\, p_{gltA,v_{Crp,bound}} \right]$$

$$f_{G,icd} = \mu\, p_{BM,k_{expr}} \left[\left(1 - \frac{x_{Cra}}{x_{Cra} + p_{icd,K_{Cra}}}\right) p_{icd,v_{Cra,unbound}} \right.$$

$$\left. + \frac{x_{Cra}}{x_{Cra} + p_{icd,K_{Cra}}}\, p_{icd,v_{Cra,bound}} \right]$$

$$f_{G,iclr} = 0$$

$$f_{G,maeAB} = (\mu + p_{D,k_{degr}})\, SS_{x_{MaeAB}}$$

$$f_{G,mdh} = \mu\, p_{BM,k_{expr}} \left[\left(1 - \frac{x_{CrpcAMP}}{x_{CrpcAMP} + p_{mdh,K_{Crp}}}\right) p_{mdh,v_{Crp,unbound}} \right.$$

$$\left. + \frac{x_{CrpcAMP}}{x_{CrpcAMP} + p_{mdh,K_{Crp}}}\, p_{mdh,v_{Crp,bound}} \right]$$

$$f_{G,pckA} = \mu\, p_{BM,k_{expr}} \left[\left(1 - \frac{x_{Cra}}{x_{Cra} + p_{pckA,K_{Cra}}}\right) p_{pckA,v_{Cra,unbound}} \right.$$

$$\left. + \frac{x_{Cra}}{x_{Cra} + p_{pckA,K_{Cra}}}\, p_{pckA,v_{Cra,bound}} \right]$$

$$f_{G,pdh} = \mu\, p_{BM,k_{expr}} \left[\left(1 - \frac{x_{PdhR}}{x_{PdhR} + p_{pdh,K_{PdhR}}}\right) p_{pdh,v_{PdhR,unbound}} \right.$$

$$\left. + \frac{x_{PdhR}}{x_{PdhR} + p_{pdh,K_{PdhR}}}\, p_{pdh,v_{PdhR,bound}} \right]$$

$$f_{G,pdhr} = 0$$

$$f_{G,pfkA} = \mu\, p_{BM,k_{expr}} \left[\left(1 - \frac{x_{Cra}}{x_{Cra} + p_{pfkA,K_{Cra}}}\right) p_{pfkA,v_{Cra,unbound}} \right.$$

$$\left. + \frac{x_{Cra}}{x_{Cra} + p_{pfkA,K_{Cra}}}\, p_{pfkA,v_{Cra,bound}} \right]$$

$$f_{G,ppc} = (\mu + p_{D,k_{degr}})\, SS_{x_{Ppc}}$$

$$f_{G,ppsA} = \mu\, p_{BM,k_{expr}} \left[\left(1 - \frac{x_{Cra}}{x_{Cra} + p_{ppsA,K_{Cra}}}\right) p_{ppsA,v_{Cra,unbound}} \right.$$

$$+ \frac{x_{Cra}}{x_{Cra} + p_{ppsA,K_{Cra}}} \, p_{ppsA,v_{Cra,bound}} \Bigg]$$

$$f_{G,pykF} = \mu \, p_{BM,k_{expr}} \left[\left(1 - \frac{x_{Cra}}{x_{Cra} + p_{pykF,K_{Cra}}} \right) p_{pykF,v_{Cra,unbound}} \right.$$

$$\left. + \frac{x_{Cra}}{x_{Cra} + p_{pykF,K_{Cra}}} \, p_{pykF,v_{Cra,bound}} \right]$$

$$f_{G,EIIA} = 0$$

$$f_{G,EIICB} = 0$$

3.6.1.8 Dilution and degradation of compounds

All metabolites dilute with the growth rate due to the expanding volume of the cell. Similarly, proteins dilute in the same manner, and additionally degrade with a here assumed 'universal' protein degradation rate k_{degr}. The concentrations of some proteins such as the transcription factors, however, are assumed as constant; these proteins are in the model neither produced (see Section 3.6.1.7), nor diluted or degraded.

Metabolite dilution

$$f_{D,ACoA} = \mu \, x_{ACoA}$$

$$f_{D,cAMP} = \mu \, x_{cAMP}$$

$$f_{D,FBP} = \mu \, x_{FBP}$$

$$f_{D,G6P} = \mu \, x_{G6P}$$

$$f_{D,GLX} = \mu \, x_{GLX}$$

$$f_{D,ICT} = \mu \, x_{ICT}$$

$$f_{D,MAL} = \mu \, x_{MAL}$$

$$f_{D,OAA} = \mu \, x_{OAA}$$

$$f_{D,PEP} = \mu \, x_{PEP}$$

$$f_{D,PG3} = \mu \, x_{PG3}$$

$$f_{D,PYR} = \mu \, x_{PYR}$$

$$f_{D,AKG} = \mu \, x_{AKG}$$

Protein degradation and dilution

$$f_{D,AceA} = (\mu + p_{D,k_{degr}}) \, x_{AceA}$$

$$f_{D,AceB} = (\mu + p_{D,k_{degr}}) \, x_{AceB}$$

$$f_{D,AceK} = (\mu + p_{D,k_{degr}}) \, x_{AceK}$$

$$f_{D,Acoa2act} = 0$$
$$f_{D,Acs} = (\mu + p_{D,k_{degr}})\, x_{Acs}$$
$$f_{D,CAMPdegr} = 0$$
$$f_{D,Cra} = 0$$
$$f_{D,CraFBP} = 0$$
$$f_{D,Crp} = 0$$
$$f_{D,CrpcAMP} = 0$$
$$f_{D,Cya} = 0$$
$$f_{D,Emp} = (\mu + p_{D,k_{degr}})\, x_{Emp}$$
$$f_{D,Eno} = (\mu + p_{D,k_{degr}})\, x_{Eno}$$
$$f_{D,Fdp} = (\mu + p_{D,k_{degr}})\, x_{Fdp}$$
$$f_{D,GltA} = (\mu + p_{D,k_{degr}})\, x_{GltA}$$
$$f_{D,Icd} = (\mu + p_{D,k_{degr}})\, x_{Icd}$$
$$f_{D,Icd-P} = (\mu + p_{D,k_{degr}})\, x_{Icd-P}$$
$$f_{D,IclR} = 0$$
$$f_{D,MaeAB} = (\mu + p_{D,k_{degr}})\, x_{MaeAB}$$
$$f_{D,Mdh} = (\mu + p_{D,k_{degr}})\, x_{Mdh}$$
$$f_{D,PckA} = (\mu + p_{D,k_{degr}})\, x_{PckA}$$
$$f_{D,Pdh} = (\mu + p_{D,k_{degr}})\, x_{Pdh}$$
$$f_{D,PdhR} = 0$$
$$f_{D,PdhRPYR} = 0$$
$$f_{D,PfkA} = (\mu + p_{D,k_{degr}})\, x_{PfkA}$$
$$f_{D,Ppc} = (\mu + p_{D,k_{degr}})\, x_{Ppc}$$
$$f_{D,PpsA} = (\mu + p_{D,k_{degr}})\, x_{PpsA}$$
$$f_{D,PykF} = (\mu + p_{D,k_{degr}})\, x_{PykF}$$
$$f_{D,Akg2mal} = (\mu + p_{D,k_{degr}})\, x_{Akg2mal}$$
$$f_{D,EIIA} = 0$$
$$f_{D,EIIA_P} = 0$$
$$f_{D,EIICB} = 0$$

3.6.1.9 Biomass production rates and growth rate calculation

Calculation of growth rates from simulated intracellular concentrations and reaction rates (such as of the biomass–producing metabolic reactions) is still an unresolved problem in the field of kinetic models. A previous study was unable to find a function that reproduces the measured growth rates from the simulated compound concentrations and reaction rates [24]. Instead, the authors of this work used the measured time course of the growth rate as input to the model. The same group has later established a correlation between the growth rate on glycolytic substrates and the phosphorylation of a PTS protein [25], which the here presented model unfortunately cannot exploit since this correlation only holds for glycolytic growth but not for growth on the gluconeogenic substrate acetate.

In this study, we chose a different approach to calculate the growth rate. Instead of calculating the growth rate as a function of the intracellular states and rates, we exploited the fact that the growth rate is determined by the quality of the growth medium [204] and calculated the growth rate as a function of the available carbon sources. This function uses two weights that depend on the available carbon sources to interpolate between the steady state growth rates on either glucose or acetate as the sole carbon sources.

The two weights used to calculate the carbon source-dependent growth rate are defined as

$$\alpha_{GLC} = \frac{x_{GLC}}{x_{GLC} + p_{PTS,K_{GLC}}}$$
$$\alpha_{ACT} = \frac{x_{ACT}}{x_{ACT} + p_{Acs,K_{ACT}}} \left(1 - \alpha_{GLC}\right),$$

with $p_{PTS,K_{GLC}}$ the Monod constant for glucose and $p_{Acs,K_{ACT}}$ the Monod constant for acetate. With this definition, α_{GLC} vanishes when glucose is absent and approaches 1 for increasing glucose concentrations, α_{ACT} vanishes when acetate is absent and approaches 1 for increasing acetate concentrations only when glucose is absent, and $0 \leq \alpha_{GLC} + \alpha_{ACT} < 1$. These two weigths are then used to calculate the growth rate from the available carbon sources through

$$\mu = \alpha_{GLC}\, p_{BM,\mu_{GLC}} + \alpha_{ACT}\, p_{BM,\mu_{ACT}}, \tag{3.1}$$

which interpolates between the known growth rates on glucose ($p_{BM,\mu_{GLC}}$) and acetate ($p_{BM,\mu_{ACT}}$).

The two weights α_{GLC} and α_{ACT} are further used to calculate the carbon source-dependent steady state concentrations $SS_{x_{Ppc}}$ and $SS_{x_{MaeAB}}$ of the enzymes Ppc and MaeAB (see Section 3.6.1.7) through

$$SS_{x_{MaeAB}} = \alpha_{GLC} \cdot 1.00 \cdot 10^{-3} \frac{g_{Prot}}{g_{DW}} + \alpha_{ACT} \cdot 3.40 \cdot 10^{-3} \frac{g_{Prot}}{g_{DW}}$$
$$SS_{x_{Ppc}} = \alpha_{GLC} \cdot 1.00 \cdot 10^{-3} \frac{g_{Prot}}{g_{DW}} + \alpha_{ACT} \cdot 2.80 \cdot 10^{-4} \frac{g_{Prot}}{g_{DW}},$$

which interpolate between the steady state data on glucose and acetate.

The reactions for the production of biomass from the seven precursor metabolites ACoA, AKG, G6P, OAA, PEP, PG3 and PYR are modeled with first order kinetics,

$$f_{BM,ACoA} = k_{BM,ACoA}\, x_{ACoA}$$
$$f_{BM,AKG} = k_{BM,AKG}\, x_{AKG}$$
$$f_{BM,G6P} = k_{BM,G6P}\, x_{G6P}$$
$$f_{BM,OAA} = k_{BM,OAA}\, x_{OAA}$$
$$f_{BM,PEP} = k_{BM,PEP}\, x_{PEP}$$
$$f_{BM,PG3} = k_{BM,PG3}\, x_{PG3}$$
$$f_{BM,PYR} = k_{BM,PYR}\, x_{PYR}\,,$$

with $k_{BM,M}$ the seven first order reaction rate constants. These rate constants were determined as follows. At steady state, the seven biomass production rates $f_{BM,M,SS}$ and the seven metabolite concentrations $x_{M,SS}$ are known for growth on either glucose or acetate (see Section 3.6.2). Thus, the carbon source-dependent first order rate constants are given by $k_{BM,M} = f_{BM,M,SS}/x_{M,SS}$. As these rate constants differ for growth on glucose or acetate, their actual values depend on the available carbon sources and are determined through

$$k_{BM,ACoA} = \alpha_{GLC}\, p_{BM,GLC_{ACoA}} + \alpha_{ACT}\, p_{BM,ACT_{ACoA}}$$
$$k_{BM,AKG} = \alpha_{GLC}\, p_{BM,GLC_{AKG}} + \alpha_{ACT}\, p_{BM,ACT_{AKG}}$$
$$k_{BM,G6P} = \alpha_{GLC}\, p_{BM,GLC_{G6P}} + \alpha_{ACT}\, p_{BM,ACT_{G6P}}$$
$$k_{BM,OAA} = \alpha_{GLC}\, p_{BM,GLC_{OAA}} + \alpha_{ACT}\, p_{BM,ACT_{OAA}}$$
$$k_{BM,PEP} = \alpha_{GLC}\, p_{BM,GLC_{PEP}} + \alpha_{ACT}\, p_{BM,ACT_{PEP}}$$
$$k_{BM,PG3} = \alpha_{GLC}\, p_{BM,GLC_{PG3}} + \alpha_{ACT}\, p_{BM,ACT_{PG3}}$$
$$k_{BM,PYR} = \alpha_{GLC}\, p_{BM,GLC_{PYR}} + \alpha_{ACT}\, p_{BM,ACT_{PYR}}\,,$$

which interpolate between the carbon source-dependent first order rate constants on glucose p_{BM,GLC_M} and acetate p_{BM,ACT_M}.

3.6.2 Parameter estimation

This section first outlines the followed parameter estimation strategy. Then, it describes the derivation of complete state and rate data sets, which need to be determined in order to apply the chosen parameter estimation strategy. Finally, it lists the values of the estimated parameters.

3.6.2.1 Application of the divide–and–conquer approach

We applied the divide–and–conquer approach [101] to determine the parameters of the presented model from experimental data. This approach decomposes the

global estimation problem into multiple independent subproblems of much smaller dimension. This property effectively removes the burden of providing a global optimizer with sufficiently good initial guesses of the parameter values, which is a significant relief given the size of this estimation problem.

In the first step of the divide–and–conquer approach, complete steady state -omics data sets of all the model's states and rates must be obtained. However, the available -omics measurement data sets are still incomplete. Therefore, we needed to incorporate additional biological knowledge in order to 'extrapolate' these incomplete measurement data sets to complete data sets of observables. Based on published measurement data, we derived two complete data sets for balanced growth on glucose and acetate, respectively. For details on this derivation, refer to Sections 3.6.2.2 and 3.6.2.3. The two complete data sets comprise 153 data points each, or 306 in total — 2 x 44 data points of states (see Table 3.3), and 2 x 109 data points of rates (see Table 3.4).

The second step in the application of the divide–and–conquer approach is the decomposition of the global estimation problem into multiple independent subproblems of smaller dimension. The degree of this decomposition, and thus the advantage of using the approach, can be increased by ensuring that each *estimated* parameter appears in exactly one rate equation. In the model, all parameters appear in exactly one rate equation, with the exception of two: The universal protein degradation rate $p_{D,k_{degr}}$ appears in all dynamic protein degradation rate equations, and the scaling factor of the growth-rate dependent efficiency of the gene expression machinery $p_{BM,k_{expr}}$ appears in all dynamic gene expression rate equations. If these two parameters are excluded from the estimation problem, then the estimation problems of all rates become independent of each other. In order to trigger such a decomposition, we excluded these two parameter from the estimation problem. Instead, we used a literature value for the universal protein degradation rate $p_{D,k_{degr}}$ and arbitrarily set the scaling factor of the growth-rate dependent efficiency of the gene expression machinery $p_{D,k_{degr}}$ to 1. This arbitrary value is justified because in the model, $p_{D,k_{degr}}$ is always multiplied with parameters that describe the maximal gene expression rates, such that the later estimation of these maximal gene expression rates can correct for the arbitrary value of $p_{D,k_{degr}}$. With the exclusion of these two parameters from the estimation problem, the estimation problems of the individual rate equations are decoupled from each other and can thus be solved independently.

The third and final fourth steps of the divide–and–conquer approach concern the derivation of the complete, analytical solution spaces of the exactly determined or underdetermined parameter estimation subproblems, and the systematic analysis of these spaces to obtain a sound system understanding in the face of non-identifiable parameter values [101]. When an estimation subproblem was underdetermined, we screened the literature and two previous models [24, 36] for values of parameters that appear in the underdetermined subproblem. The solutions to the small dimensional estimation subproblems were then joined to the solution of the original parent estimation problem. The resulting parameter values

are listed in Table 3.5.

3.6.2.2 Derivation of complete data sets of state variables

Table 3.3 contains the steady state values of the state variables $\mathbf{x}$ on glucose and acetate. These values are either experimental data, or, where experimental data was not available, estimated from experimental data through the inclusion of additional biological knowledge. Specifically, to obtain two complete metabolomes, we used NET analysis [105] to integrate and consolidate quantitative metabolite measurements from multiple sources [36, 44, 48, 52, 119, 153, 158] measured during growth on either glucose or acetate, into two thermodynamically consistent data sets. To complete these data sets, we again used NET analysis to infer unknown metabolite concentrations from measured concentrations.

To obtain a complete proteome for balanced growth on glucose, we used measured concentrations where available [3, 80] and set all unmeasured protein concentrations to the arbitrary value $1.0 \cdot 10^{-3} g_{Prot}/g_{DW}$. These arbitrary protein concentrations on glucose are justified because in the model, a protein concentration always appears paired with a multiplicative parameter k_{cat} such that the later estimation of this parameter value can correct for an arbitrarily chosen absolute concentration of the protein. Unfortunately, we could not find any published proteome data for growth on acetate. However, it has been observed that protein and mRNA abundances in *E. coli* cells are significantly correlated [80]. Therefore, we used the data of a microarray study that determined the ratios of mRNA concentrations between growth on glucose and acetate [141, 142] to estimate the protein concentrations on acetate from those on glucose.

In five cases (EIIA, Icd, Cra, Crp, and PdhR), the concentration of a physical compound is distributed over two state variables. In these cases, the total concentration of the physical compound is the sum of these two state variables. Two of these cases arise because the PTS protein EIIA and the enzyme Icd both exist in a phosphorylated and an unphosphorylated form. As the degree of phosphorylation ρ with $0 \leq \rho \leq 1$ of these compounds is known for balanced growth on glucose and acetate [24, 221], the steady state concentrations of the phosphorylated and unphosphorylated forms, x_p and x_u respectively, can be calculated from the total protein concentrations x_{tot} through $x_p = \rho \, x_{tot}$ and $x_u = (1 - \rho) \, x_{tot}$. The remaining three cases arise because the TFs Cra, Crp, and PdhR appear either bound or unbound to their respective metabolite effectors FBP, cAMP, and PYR. Because the concentrations of the metabolite-bound and free forms on glucose and acetate are unknown, we needed a rationale to set these concentrations. As the activity of these TFs is believed to be markedly distinct during glycolytic and gluconeogenic growth, and the TF activities are modulated by their small molecule effectors, we chose the parameters quantifying the TF-metabolite bindings such that the difference between the TFs' metabolite-bound concentrations on glucose and acetate is maximized.

Table 3.3: Comprehensive list of the dynamic state variables **x**, the full names of the represented compounds, and the steady state values on glucose and acetate, which are based on experimental data. The units of the states are $g\,l^{-1}$ for carbon sources, $\frac{\mu mol}{g_{DW}}$ for metabolites, $\frac{g_{Prot}}{g_{DW}}$ for proteins, and $[OD]$ for the biomass concentration.

Name of state	Description	Data on glucose	Data on acetate
x_{OD}	Biomass concentration	-	-
x_{ACT}	Extracellular acetate	-	-
x_{GLC}	Extracellular glucose	-	-
x_{ACoA}	Acetyl-CoA	0.35	1.9
x_{AKG}	α-Ketoglutarate	0.2	1.1
x_{cAMP}	Cyclic AMP	0.2	4
x_{FBP}	Fructose-1,6-bisphosphate	6.6	0.28
x_{G6P}	Glucose-6-phosphate	1.85	1.17
x_{GLX}	Glyoxylate	1.00E-08	1.35
x_{ICT}	Isocitrate	1.35E-03	1.54
x_{MAL}	Malate	3.6	6.65
x_{OAA}	Oxaloacetate	0.05	0.07
x_{PEP}	Phosphoenolpyruvate	0.21	0.59
x_{PG3}	3-Phosphoglycerate	5.75	1.35
x_{PYR}	Pyruvate	0.9	0.03
x_{AceA}	Isocitrate lyase	4.68E-03	1.03E-01
x_{AceB}	Malate synthase A	1.40E-03	3.09E-02
x_{AceK}	Isocitrate dehydrogenase phosphatase/kinase	1.40E-04	3.09E-03
$x_{Acoa2act}$	Enzyme for the reaction from ACoA to ACT	1.00E-03	3.00E-04
x_{Acs}	Acetyl-CoA synthetase	3.62E-05	3.35E-04
$x_{Akg2mal}$	Enzyme for the reaction from AKG to MAL	1.00E-03	2.10E-03
$x_{CAMPdegr}$	Degradation of cAMP	1.00E-03	1.00E-03
x_{Cya}	Adenylate cyclase	1.00E-03	1.00E-03
x_{Emp}	Enzyme for the reversible reaction between FBP and PG3	1.14E-02	9.64E-03
x_{Eno}	Enolase	1.14E-02	6.21E-03
x_{Fdp}	Fructose-1,6-bisphosphatase I	7.48E-05	2.44E-04
x_{GltA}	Citrate synthase	2.93E-04	1.01E-03
x_{Icd}	Unphosphorylated isocitrate dehydrogenase	4.28E-03	2.47E-03
x_{Icd-P}	Phosphorylated isocitrate	1.78E-04	7.41E-03

. . . Table 3.3 continued.

Name of state	Description	Data on glucose	Data on acetate
	dehydrogenase		
x_{MaeAB}	Malic enzymes MaeAB	1.00E-03	3.40E-03
x_{Mdh}	Malate dehydrogenase	4.91E-04	1.56E-03
x_{PckA}	Phosphoenolpyruvate carboxykinase	3.37E-04	2.78E-03
x_{Pdh}	Pyruvate dehydrogenase	1.00E-03	3.79E-04
x_{PfkA}	6-phosphofructokinase I	2.42E-04	1.50E-04
x_{Ppc}	Phosphoenolpyruvate carboxylase	3.78E-04	1.06E-04
x_{PpsA}	Phosphoenolpyruvate synthase	1.00E-03	1.30E-02
x_{PykF}	Pyruvate kinase I	2.50E-03	5.47E-04
x_{EIIA}	Unphosphorylated PTS protein EIIA	9.65E-02	1.99E-03
x_{EIIA-P}	Phosphorylated PTS protein EIIA	3.48E-03	9.80E-02
x_{EIICB}	PTS protein EIICB	3.00E-03	3.00E-03
x_{Cra}	Free Cra	2.97E-04	6.99E-03
x_{CraFBP}	Cra bound to fructose-1,6-bisphosphate	6.99E-03	2.97E-04
x_{Crp}	Free Crp	5.96E-03	1.33E-03
$x_{CrpcAMP}$	Crp bound to cyclic AMP	1.33E-03	5.96E-03
x_{IclR}	IclR	7.29E-03	7.29E-03
x_{PdhR}	free PdhR	1.13E-03	6.17E-03
$x_{PdhRPYR}$	PdhR bound to pyruvate	6.17E-03	1.13E-03

3.6.2.3 Derivation of complete data sets of rates

Table 3.4 lists the values of the steady state rates $\mathbf{f}$ on glucose and acetate. These values are either experimental data, or, where experimental data was not available, estimated from experimental data through the inclusion of additional biological knowledge.

Two complete sets of metabolic reaction rates (covering the rates $\mathbf{f_E}$, $\mathbf{f_{PTS}}$ and $\mathbf{f_{BM}}$) were provided by the results of ^{13}C tracer experiments on glucose [57] and acetate [233].

To obtain two complete data sets of dilution rates $\mathbf{f_{dil}}$ of the compounds $\mathbf{x}$ due to cell growth, we used the model $\mathbf{f_{dil}} = \mu \mathbf{x}$ to calculate the dilution rates from the steady state concentrations of the proteins and metabolites on glucose and acetate, and from the known growth rate μ on these substrates. Similarly, to obtain two complete data sets of degradation rates $\mathbf{f_{degr}}$ of the proteins $\mathbf{x}$, we used the model $\mathbf{f_{degr}} = p_{D,k_{degr}} \mathbf{x}$ to calculate the degradation rates from the steady state protein concentrations on glucose and acetate, assuming the same degradation rate $p_{D,k_{degr}}$ for all proteins. The protein dilution and degradation rates, together with the steady state assumption $\mathbf{f_G} = \mathbf{f_{dil}} + \mathbf{f_{degr}}$, are used to

calculate the steady state gene expression rates $\mathbf{f_G}$ on glucose and acetate.

The concentrations of the proteins EIIA and Icd are distributed over two state variables; one of these variables denotes the phosphorylated form, and the other the unphosphorylated form. Because the degrees of phosphorylation of the proteins EIIA and Icd have been determined experimentally [24, 221], the *ratio* of the these proteins' steady state phosphorylation and dephosphorylation rates *between growth on glucose and acetate* is also known. However, the *magnitudes* of these rates are uncertain. We observed in preliminary simulations that the phosphorylation and dephosphorylation rates need to be sufficiently fast to not introduce oscillations into the metabolic network. Therefore, we set the steady state phosphorylation and desphosphorylation rates of the proteins EIIA and Icd to sufficiently high magnitudes in order to avoid such oscillations.

Similarly, as the steady state concentrations of the effector-bound and free forms of the transcription factors Cra, Crp, and PdhR have been estimated in Section 3.6.2.2, the ratios of these two forms' concentrations between growth on glucose and acetate are also known. Therefore, the ratios of both the association and dissociation rates between growth on glucose and acetate are known; however, the magnitudes of these rates are uncertain. In order to allow the binding states of the transcription factors to track the metabolite concentrations, the ratio of association and dissociation rates was scaled to the metabolic time scale, which is significantly faster than the slow time scale on which the TF's regulation of gene expression operates.

Table 3.4: Comprehensive list of the steady state rates $\mathbf{f}$. The units of the rates are h^{-1} for the growth rates; $\frac{g}{l \cdot s}$ for the substrate uptake and excretion rates of the cell population; $\frac{\mu mol}{g_{DW} \cdot s}$ for the enzymatic reaction rates $f_{E,\cdot}$ (except the AceK kinase and phosphatase reactions), for the PTS reactions $f_{PTS,\cdot}$, for the dilution rates of metabolites $f_{D,\cdot}$, for the association and dissociation rates of transcription factors to metabolites $f_{TF,\cdot}$, and for the biomass production rates $f_{BM,\cdot}$; $\frac{g_{Prot}}{g_{DW} \cdot s}$ for the AceK kinase and phosphatase reactions, for the gene expression rates $f_{G,\cdot}$ and for the protein dilution+degradation rates $f_{D,\cdot}$.

Name of rate	Description	Data on glucose	Data on acetate
$f_{ENV,growth}$	Growth rate	0.64	0.20
$f_{ENV,GLCup}$	Glucose uptake rate of the population	-	-
$f_{ENV,ACTup}$	Acetate uptake rate of the population	-	-
$f_{ENV,ACTex}$	Acetate excretion rate of the population	-	-
$f_{E,AceA}$	Metabolic flux through AceA	3.52E-04	0.666
$f_{E,AceB}$	Metabolic flux through AceB	3.52E-04	0.666

. . . Table 3.4 continued.

Name of state	Description	Data on glucose	Data on acetate
$f_{E,AceK-Ki}$	Rate of the AceK-kinase reaction	6.61E-04	1.91E-02
$f_{E,AceK-Ph}$	Rate of the AceK-phosphatase reaction	6.61E-04	1.91E-02
$f_{E,Acoa2act}$	Metabolic flux through Acoa2act, Cell dry weight-normalized acetate excretion rate	1.33	0.25
$f_{E,Acs}$	Metabolic flux through Acs, Cell dry weight-normalized acetate uptake rate	0	3.45
$f_{E,Akg2mal}$	Conversion rate of AKG to MAL	0.443	2.32
$f_{E,CAMPdegr}$	Degradation rate of cAMP	0.667	0.976
$f_{E,Cya}$	Production rate of cAMP	0.667	0.976
$f_{E,Emp}$	Conversion rate between FBP and PG3	3.87	-0.188
$f_{E,Eno}$	Metabolic flux through Eno	3.59	-0.277
$f_{E,Fdp}$	Metabolic flux through Fdp	3.00E-02	9.89E-02
$f_{E,GltA}$	Metabolic flux through GltA	0.570	2.38
$f_{E,Icd}$	Metabolic flux through Icd	0.569	1.717
$f_{E,MaeAB}$	Metabolic flux through MaeAB	6.88E-02	0.227
$f_{E,Mdh}$	Metabolic flux through Mdh	0.443	2.76
$f_{E,PckA}$	Metabolic flux through PckA	2.58E-02	0.282
$f_{E,Pdh}$	Metabolic flux through Pdh	2.484	5.50E-02
$f_{E,PfkA}$	Metabolic flux through PfkA	1.966	1.00E-02
$f_{E,Ppc}$	Metabolic flux through Ppc	0.610	3.20E-03
$f_{E,PpsA}$	Metabolic flux through PpsA	1.30E-03	1.61E-02
$f_{E,PykF}$	Metabolic flux through PykF	0.695	1.00E-04
f_{PTS,r_1}	Metabolic flux through the PTS reaction r_1	2.22	0
f_{PTS,r_4}	Metabolic flux through the PTS reaction r_4, Cell dry weight-normalized glucose uptake rate	2.22	0
$f_{TF,Cra}$	Combined association & dissociation rates between Cra and FBP	0	0
$f_{TF,Crp}$	Combined association & dissociation rates between Crp and cAMP	0	0

... Table 3.4 continued.

Name of state	Description	Data on glucose	Data on acetate
$f_{TF,PdhR}$	Combined association & dissociation rates between PdhR and PYR	0	0
$f_{G,aceA}$	aceA expression rate	9.63E-07	8.61E-06
$f_{G,aceB}$	aceB expression rate	2.89E-07	2.58E-06
$f_{G,aceK}$	aceK expression rate	2.89E-08	2.58E-07
$f_{G,acoa2act}$	acoa2act expression rate	0	0
$f_{G,acs}$	acs expression rate	7.53E-09	2.81E-08
$f_{G,akg2mal}$	akg2mal expression rate	2.06E-07	1.75E-07
$f_{G,campdegr}$	campdegr expression rate	0	0
$f_{G,cra}$	cra expression rate	0	0
$f_{G,crp}$	crp axpression rate	0	0
$f_{G,cya}$	cya expression rate	0	0
$f_{G,EIIA}$	eiia expression rate	0	0
$f_{G,EIICB}$	eiicb expression rate	0	0
$f_{G,emp}$	emp expression rate	2.37E-06	8.10E-07
$f_{G,eno}$	eno expression rate	2.37E-06	5.22E-07
$f_{G,fdp}$	fdp expression rate	1.56E-08	2.05E-08
$f_{G,gltA}$	gltA expression rate	6.09E-08	8.45E-08
$f_{G,icd}$	icd expression rate	9.17E-07	8.26E-07
$f_{G,maeAB}$	maeAB expression rate	2.06E-07	2.84E-07
$f_{G,iclr}$	iclr expression rate	0	0
$f_{G,mdh}$	mdh expression rate	1.02E-07	1.31E-07
$f_{G,pckA}$	pckA expression rate	7.01E-08	2.34E-07
$f_{G,pdh}$	pdh expression rate	2.08E-07	3.18E-08
$f_{G,pdhr}$	pdhR expression rate	0	0
$f_{G,pfkA}$	pfkA expression rate	5.04E-08	1.26E-08
$f_{G,ppc}$	ppc expression rate	7.86E-08	8.89E-09
$f_{G,ppsA}$	ppsA expression rate	2.06E-07	1.09E-06
$f_{G,pykF}$	pykF expression rate	5.20E-07	4.59E-08
$f_{D,ACoA}$	Dilution rate of ACoA	6.22E-05	1.06E-04
$f_{D,AKG}$	Dilution rate of AKG	3.56E-05	6.11E-05
$f_{D,cAMP}$	Dilution rate of cAMP	3.56E-05	2.22E-04
$f_{D,FBP}$	Dilution rate of FBP	1.17E-03	1.56E-05
$f_{D,G6P}$	Dilution rate of G6P	3.29E-04	6.50E-05
$f_{D,GLX}$	Dilution rate of GLX	1.78E-12	7.50E-05
$f_{D,ICT}$	Dilution rate of ICT	2.40E-07	8.56E-05
$f_{D,MAL}$	Dilution rate of MAL	6.40E-04	3.69E-04

... Table 3.4 continued.

Name of state	Description	Data on glucose	Data on acetate
$f_{D,OAA}$	Dilution rate of OAA	8.89E-06	3.89E-06
$f_{D,PEP}$	Dilution rate of PEP	3.73E-05	3.28E-05
$f_{D,PG3}$	Dilution rate of PG3	1.02E-03	7.50E-05
$f_{D,PYR}$	Dilution rate of PYR	1.60E-04	1.67E-06
$f_{D,AceA}$	Degr.&dilution rate of AceA	9.63E-07	8.61E-06
$f_{D,AceB}$	Degr.&dilution rate of AceB	2.89E-07	2.58E-06
$f_{D,AceK}$	Degr.&dilution rate of AceK	2.89E-08	2.58E-07
$f_{D,Acoa2act}$	Degr.&dilution rate of Acoa2act	0	0
$f_{D,Acs}$	Degr.&dilution rate of Acs	7.53E-09	2.81E-08
$f_{D,Akg2mal}$	Degr.&dilution rate of Akg2mal	2.06E-07	1.75E-07
$f_{D,CAMPdegr}$	Degr.&dilution rate of CAMPdegr	0	0
$f_{D,Cra}$	Degr.&dilution rate of Cra	0	0
$f_{D,CraFBP}$	Degr.&dilution rate of CraFBP	0	0
$f_{D,Crp}$	Degr.&dilution rate of Crp	0	0
$f_{D,CrpcAMP}$	Degr.&dilution rate of CrpcAMP	0	0
$f_{D,Cya}$	Degr.&dilution rate of Cya	0	0
$f_{D,EIIA}$	Degr.&dilution rate of EIIA	0	0
$f_{D,EIIA-P}$	Degr.&dilution rate of EIIA-P	0	0
$f_{D,EIICB}$	Degr.&dilution rate of EIICB	0	0
$f_{D,Emp}$	Degr.&dilution rate of Emp	2.37E-06	8.10E-07
$f_{D,Eno}$	Degr.&dilution rate of Eno	2.37E-06	5.22E-07
$f_{D,Fdp}$	Degr.&dilution rate of Fdp	1.56E-08	2.05E-08
$f_{D,GltA}$	Degr.&dilution rate of GltA	6.09E-08	8.45E-08
$f_{D,Icd}$	Degr.&dilution rate of Icd	8.80E-07	2.06E-07
$f_{D,Icd-P}$	Degr.&dilution rate of Icd-P	3.67E-08	6.19E-07
$f_{D,IclR}$	Degr.&dilution rate of IclR	0	0
$f_{D,MaeAB}$	Degr.&dilution rate of MaeAB	2.06E-07	2.84E-07
$f_{D,Mdh}$	Degr.&dilution rate of Mdh	1.02E-07	1.31E-07
$f_{D,PckA}$	Degr.&dilution rate of PckA	7.01E-08	2.34E-07
$f_{D,Pdh}$	Degr.&dilution rate of Pdh	2.08E-07	3.18E-08
$f_{D,PdhR}$	Degr.&dilution rate of PdhR	0	0
$f_{D,PdhRPYR}$	Degr.&dilution rate of PdhRPYR	0	0
$f_{D,PfkA}$	Degr.&dilution rate of PfkA	5.04E-08	1.26E-08
$f_{D,Ppc}$	Degr.&dilution rate of Ppc	7.86E-08	8.89E-09
$f_{D,PpsA}$	Degr.&dilution rate of PpsA	2.06E-07	1.09E-06
$f_{D,PykF}$	Degr.&dilution rate of PykF	5.20E-07	4.59E-08
$f_{BM,ACoA}$	Biomass flux from ACoA	0.658	0.206
$f_{BM,AKG}$	Biomass flux from AKG	0.196	6.11E-02

... Table 3.4 continued.

Name of state	Description	Data on glucose	Data on acetate
$f_{BM,G6P}$	Biomass flux from G6P	0.284	8.89E-02
$f_{BM,OAA}$	Biomass flux from OAA	0.320	0.100
$f_{BM,PEP}$	Biomass flux from PEP	8.89E-02	2.78E-02
$f_{BM,PG3}$	Biomass flux from PG3	0.284	8.89E-02
$f_{BM,PYR}$	Biomass flux from PYR	0.498	0.156

3.6.2.4 Parameter values

This section contains Table 3.5, which lists the parameter values determined through application of the divide–and–conquer approach, described in Section 3.6.2.1, on the derived data sets, presented in Sections 3.6.2.2 and 3.6.2.3.

Table 3.5: Comprehensive list of the parameters **p**, their mechanistic meanings, and their values as determined by the divide–and–conquer approach (see Section 3.6.2) on the data presented in Tables 3.3 and 3.4.

Parameter	Description	Value
$p_{ENV,M_{ACT}}$	Molar mass of acetate	180.2 $g_{ACT}\,mol^{-1}$
$p_{ENV,M_{GLC}}$	Molar mass of glucose	60 $g_{GLC}\,mol^{-1}$
$p_{ENV,UC}$	Unit conversion	9.5E-07 $g_{DW}\,(\mu[OD])^{-1}$
$p_{AceA,k_{cat}}$	Specific activity	1.03E+04 $\mu mol\,(g_{Prot}\,s)^{-1}$
$p_{AceA,n}$	Number of subunits	4
$p_{AceA,L}$	Allosteric constant	5.01E+04
$p_{AceA,K_{ICT}}$	Affinity constant	0.022 $\mu mol\,g_{DW}^{-1}$
$p_{AceA,K_{PEP}}$	Affinity constant	0.055 $\mu mol\,g_{DW}^{-1}$
$p_{AceA,K_{PG3}}$	Affinity constant	0.72 $\mu mol\,g_{DW}^{-1}$
$p_{AceA,K_{AKG}}$	Affinity constant	0.827 $\mu mol\,g_{DW}^{-1}$
$p_{AceB,k_{cat}}$	Specific activity	47.8 $\mu mol\,(g_{Prot}\,s)^{-1}$
$p_{AceB,K_{GLX}}$	Affinity constant	0.95 $\mu mol\,g_{DW}^{-1}$
$p_{AceB,K_{ACoA}}$	Affinity constant	0.755 $\mu mol\,g_{DW}^{-1}$
$p_{AceB,K_{GLXACoA}}$	Affinity constant	0.719 $\mu mol\,g_{DW}^{-1}$
$p_{AceK,k_{cat,ki}}$	Specific activity	3.4E+12 s^{-1}
$p_{AceK,k_{cat,ph}}$	Specific activity	1.7E+09 s^{-1}
$p_{AceK,n}$	Number of subunits	2
$p_{AceK,L}$	Allosteric constant	1.0E+08
$p_{AceK,K_{Icd}}$	Affinity constant	0.043 $g_{Prot}\,g_{DW}^{-1}$
$p_{AceK,K_{Icd-P}}$	Affinity constant	0.643 $g_{Prot}\,g_{DW}^{-1}$

. . . Table 3.5 continued.

Parameter	Description	Value
$p_{AceK,K_{PEP}}$	Affinity constant	0.539 $\mu mol \, g_{DW}{}^{-1}$
$p_{AceK,K_{PYR}}$	Affinity constant	0.038 $\mu mol \, g_{DW}{}^{-1}$
$p_{AceK,K_{OAA}}$	Affinity constant	0.173 $\mu mol \, g_{DW}{}^{-1}$
$p_{AceK,K_{GLX}}$	Affinity constant	0.866 $\mu mol \, g_{DW}{}^{-1}$
$p_{AceK,K_{AKG}}$	Affinity constant	0.82 $\mu mol \, g_{DW}{}^{-1}$
$p_{AceK,K_{PG3}}$	Affinity constant	1.57 $\mu mol \, g_{DW}{}^{-1}$
$p_{AceK,K_{ICT}}$	Affinity constant	0.137 $\mu mol \, g_{DW}{}^{-1}$
$p_{Acoa2act,k_{cat}}$	Specific activity	3079 $\mu mol \, (g_{Prot} \, s)^{-1}$
$p_{Acoa2act,n}$	Number of subunits	2
$p_{Acoa2act,L}$	Allosteric constant	6.39E+05
$p_{Acoa2act,K_{ACoA}}$	Affinity constant	0.022 $\mu mol \, g_{DW}{}^{-1}$
$p_{Acoa2act,K_{PYR}}$	Affinity constant	0.022 $\mu mol \, g_{DW}{}^{-1}$
$p_{Acs,k_{cat}}$	Specific activity	340 $\mu mol \, (g_{Prot} \, s)^{-1}$
$p_{Acs,K_{ACT}}$	Affinity constant	1.0E-03 $g_{ACT} \, l^{-1}$
$p_{Akg2mal,k_{cat}}$	Specific activity	1530 $\mu mol \, (g_{Prot} \, s)^{-1}$
$p_{Akg2mal,K_{AKG}}$	Affinity constant	0.548 $\mu mol \, g_{DW}{}^{-1}$
$p_{CAMPdegr,k_{cat}}$	Specific activity	1 $\mu mol \, (g_{Prot} \, s)^{-1}$
$p_{CAMPdegr,K_{cAMP}}$	Affinity constant	0.1 $\mu mol \, g_{DW}{}^{-1}$
$p_{Cya,k_{cat}}$	Specific activity	993 $\mu mol \, (g_{Prot} \, s)^{-1}$
$p_{Cya,K_{EIIA-P}}$	Affinity constant	1.7E-03 $g_{Prot} \, g_{DW}{}^{-1}$
$p_{Emp,k_{cat,f}}$	Specific activity of forward reaction	1011 $\mu mol \, (g_{Prot} \, s)^{-1}$
$p_{Emp,k_{cat,r}}$	Specific activity of of reverse reaction	857 $\mu mol \, (g_{Prot} \, s)^{-1}$
$p_{Emp,K_{FBP}}$	Affinity constant	5.92 $\mu mol \, g_{DW}{}^{-1}$
$p_{Emp,K_{PG3}}$	Affinity constant	16.6 $\mu mol \, g_{DW}{}^{-1}$
$p_{Eno,k_{cat,f}}$	Specific activity of forward reaction	705 $\mu mol \, (g_{Prot} \, s)^{-1}$
$p_{Eno,k_{cat,r}}$	Specific activity of reverse reaction	530 $\mu mol \, (g_{Prot} \, s)^{-1}$
$p_{Eno,K_{PG3}}$	Affinity constant	4.76 $\mu mol \, g_{DW}{}^{-1}$
$p_{Eno,K_{PEP}}$	Affinity constant	1.11 $\mu mol \, g_{DW}{}^{-1}$
$p_{Fdp,k_{cat}}$	Specific activity	5676 $\mu mol \, (g_{Prot} \, s)^{-1}$
$p_{Fdp,n}$	Number of subunits	4
$p_{Fdp,L}$	Allosteric constant	4.0E+06
$p_{Fdp,K_{FBP}}$	Affinity constant	3.0E-03 $\mu mol \, g_{DW}{}^{-1}$
$p_{Fdp,K_{PEP}}$	Affinity constant	0.3 $\mu mol \, g_{DW}{}^{-1}$
$p_{GltA,k_{cat}}$	Specific activity	1614 $\mu mol \, (g_{Prot} \, s)^{-1}$
$p_{GltA,K_{OAA}}$	Affinity constant	0.029 $\mu mol \, g_{DW}{}^{-1}$

... Table 3.5 continued.

Parameter	Description	Value
$p_{GltA,K_{ACoA}}$	Affinity constant	0.212 $\mu mol\, g_{DW}{}^{-1}$
$p_{GltA,K_{OAAACoA}}$	Affinity constant	0.029 $\mu mol\, g_{DW}{}^{-1}$
$p_{GltA,K_{AKG}}$	Affinity constant	0.63 $\mu mol\, g_{DW}{}^{-1}$
$p_{Icd,k_{cat}}$	Specific activity	695 $\mu mol\, (g_{Prot}\, s)^{-1}$
$p_{Icd,n}$	Number of subunits	2
$p_{Icd,L}$	Allosteric constant	127
$p_{Icd,K_{ICT}}$	Affinity constant	1.6E-04 $\mu mol\, g_{DW}{}^{-1}$
$p_{Icd,K_{PEP}}$	Affinity constant	0.334 $\mu mol\, g_{DW}{}^{-1}$
$p_{MaeAB,k_{cat}}$	Specific activity	1879 $\mu mol\, (g_{Prot}\, s)^{-1}$
$p_{MaeAB,n}$	Number of subunits	1.33
$p_{MaeAB,L}$	Allosteric constant	1.04E+05
$p_{MaeAB,K_{MAL}}$	Affinity constant	6.24E-03 $\mu mol\, g_{DW}{}^{-1}$
$p_{MaeAB,K_{ACoA}}$	Affinity constant	3.64 $\mu mol\, g_{DW}{}^{-1}$
$p_{MaeAB,K_{cAMP}}$	Affinity constant	6.54 $\mu mol\, g_{DW}{}^{-1}$
$p_{Mdh,k_{cat}}$	Specific activity	5437 $\mu mol\, (g_{Prot}\, s)^{-1}$
$p_{Mdh,n}$	Hill coefficient	1.7
$p_{Mdh,K_{MAL}}$	Affinity constant	10.1 $\mu mol\, g_{DW}{}^{-1}$
$p_{PckA,k_{cat}}$	Specific activity	377 $\mu mol\, (g_{Prot}\, s)^{-1}$
$p_{PckA,K_{OAA}}$	Affinity constant	0.184 $\mu mol\, g_{DW}{}^{-1}$
$p_{PckA,K_{PEP}}$	Affinity constant	1000 $\mu mol\, g_{DW}{}^{-1}$
$p_{Pdh,k_{cat}}$	Specific activity	5479 $\mu mol\, (g_{Prot}\, s)^{-1}$
$p_{Pdh,n}$	Number of subunits	2.65
$p_{Pdh,L}$	Allosteric constant	3.4
$p_{Pdh,K_{PYR}}$	Affinity constant	0.128 $\mu mol\, g_{DW}{}^{-1}$
$p_{Pdh,K_{I,PYR}}$	Affinity constant	0.231 $\mu mol\, g_{DW}{}^{-1}$
$p_{Pdh,K_{GLX}}$	Affinity constant	0.218 $\mu mol\, g_{DW}{}^{-1}$
$p_{PfkA,k_{cat}}$	Specific activity	5.39E+05 $\mu mol\, (g_{Prot}\, s)^{-1}$
$p_{PfkA,n}$	Number of subunits	4
$p_{PfkA,L}$	Allosteric constant	9.5E+07
$p_{PfkA,K_{G6P}}$	Affinity constant	0.022 $\mu mol\, g_{DW}{}^{-1}$
$p_{PfkA,K_{PEP}}$	Affinity constant	0.138 $\mu mol\, g_{DW}{}^{-1}$
$p_{Ppc,k_{cat}}$	Specific activity	1.49E+04 $\mu mol\, (g_{Prot}\, s)^{-1}$
$p_{Ppc,n}$	Number of subunits	3
$p_{Ppc,L}$	Allosteric constant	5.2E+06
$p_{Ppc,K_{PEP}}$	Affinity constant	0.048 $\mu mol\, g_{DW}{}^{-1}$
$p_{Ppc,K_{FBP}}$	Affinity constant	0.408 $\mu mol\, g_{DW}{}^{-1}$
$p_{PpsA,k_{cat}}$	Specific activity	1.32 $\mu mol\, (g_{Prot}\, s)^{-1}$
$p_{PpsA,n}$	Number of subunits	2
$p_{PpsA,L}$	Allosteric constant	1.0E-79

... Table 3.5 continued.

Parameter	Description	Value
$p_{PpsA,K_{PYR}}$	Affinity constant	1.77E-03 $\mu mol\, g_{DW}{}^{-1}$
$p_{PpsA,K_{PEP}}$	Affinity constant	1.0E-03 $\mu mol\, g_{DW}{}^{-1}$
$p_{PykF,k_{cat}}$	Specific activity	1.37E+04 $\mu mol\, (g_{Prot}\, s)^{-1}$
$p_{PykF,n}$	Number of subunits	4
$p_{PykF,L}$	Allosteric constant	1.0E+05
$p_{PykF,K_{PEP}}$	Affinity constant	5 $\mu mol\, g_{DW}{}^{-1}$
$p_{PykF,K_{FBP}}$	Affinity constant	0.413 $\mu mol\, g_{DW}{}^{-1}$
p_{PTS,k_1}	Specific activity	116 $g_{DW}\, (g_{Prot}\, s)^{-1}$
p_{PTS,km_1}	Specific activity	46.3 $g_{DW}\, (g_{Prot}\, s)^{-1}$
p_{PTS,k_4}	Specific activity	2520 $\mu mol\, (g_{Prot}\, s)^{-1}$
$p_{PTS,K_{EIIA}}$	Affinity constant	8.5E-03 $g_{Prot}\, g_{DW}{}^{-1}$
$p_{PTS,K_{GLC}}$	Affinity constant	1.2E-03 $g_{GLC}\, l^{-1}$
$p_{Cra,scale}$	Specific activity	100 $g_{Prot}\, (\mu mol\, s)^{-1}$
$p_{Cra,K_{FBP}}$	Affinity constant	1.36 $\mu mol\, g_{DW}{}^{-1}$
$p_{Cra,n}$	Hill coefficient	2
$p_{Crp,scale}$	Specific activity	1.0E+08 $g_{Prot}\, (\mu mol\, s)^{-1}$
$p_{Crp,K_{cAMP}}$	Affinity constant	0.895 $\mu mol\, g_{DW}{}^{-1}$
$p_{Crp,n}$	Hill coefficient	1
$p_{PdhR,scale}$	Specific activity	100 $g_{Prot}\, (\mu mol\, s)^{-1}$
$p_{PdhR,K_{PYR}}$	Affinity constant	0.164 $\mu mol\, g_{DW}{}^{-1}$
$p_{PdhR,n}$	Hill coefficient	1
$p_{aceBAK,aceBfactor}$	Scaling factor	0.3
$p_{aceBAK,aceKfactor}$	Scaling factor	0.03
$p_{aceBAK,k_{cat,IclR}}$	Specific activity	9.3E-04 s^{-1}
$p_{aceBAK,K_{DNA}}$	Affinity constant	2.19 $[AU]\, g_{DW}{}^{-1}$
$p_{aceBAK,DNA}$	DNA concentration	1 $[AU]\, g_{DW}{}^{-1}$
$p_{aceBAK,K_{PYR}}$	Affinity constant	0.897 $\mu mol\, g_{DW}{}^{-1}$
$p_{aceBAK,K_{PYRprime}}$	Affinity constant	3.01E-03 $\mu mol\, g_{DW}{}^{-1}$
$p_{aceBAK,K_{GLX}}$	Affinity constant	4.88E-03 $\mu mol\, g_{DW}{}^{-1}$
$p_{aceBAK,L}$	Allosteric constant	923
$p_{aceBAK,v_{Cra,unbound}}$	Basal expression rate	1.9E-09 $g_{Prot}(g_{DW}\, s)^{-1}$
$p_{aceBAK,v_{Cra,bound}}$	Max. expression rate	2.0E-06 $g_{Prot}(g_{DW}\, s)^{-1}$
$p_{aceBAK,K_{Cra}}$	Affinity constant	3.65E-03 $g_{Prot}\, g_{DW}{}^{-1}$
$p_{aceBAK,v_{Crp,unbound}}$	Max. expression rate	2.0E-08 $g_{Prot}(g_{DW}\, s)^{-1}$
$p_{aceBAK,v_{Crp,bound}}$	Basal expression rate	2.3E-10 $g_{Prot}(g_{DW}\, s)^{-1}$
$p_{aceBAK,K_{Crp}}$	Affinity constant	0.341 $g_{Prot}\, g_{DW}{}^{-1}$
$p_{acs,v_{Crp,unbound}}$	Basal expression rate	0 $g_{Prot}(g_{DW}\, s)^{-1}$
$p_{acs,v_{Crp,bound}}$	Max. expression rate	4.0E-08 $g_{Prot}(g_{DW}\, s)^{-1}$
$p_{acs,n}$	Hill coefficient	2.31

... Table 3.5 continued.

Parameter	Description	Value
$p_{acs,K_{Crp}}$	Affinity constant	4.7E-03 $g_{Prot}\,g_{DW}{}^{-1}$
$p_{akg2mal,v_{Crp,unbound}}$	Basal expression rate	0 $g_{Prot}(g_{DW}\,s)^{-1}$
$p_{akg2mal,v_{Crp,bound}}$	Max. expression rate	1.4E-06 $g_{Prot}(g_{DW}\,s)^{-1}$
$p_{akg2mal,K_{Crp}}$	Affinity constant	0.091 $g_{Prot}\,g_{DW}{}^{-1}$
$p_{akg2mal,n}$	Hill coefficient	0.74
$p_{emp,v_{Cra,unbound}}$	Max. expression rate	6.1 E-07 $g_{Prot}(g_{DW}\,s)^{-1}$
$p_{emp,v_{Cra,bound}}$	Basal expression rate	0 $g_{Prot}(g_{DW}\,s)^{-1}$
$p_{emp,K_{Cra}}$	Affinity constant	0.09 $g_{Prot}\,g_{DW}{}^{-1}$
$p_{emp,v_{Crp,unbound}}$	Basal expression rate	0 $g_{Prot}(g_{DW}\,s)^{-1}$
$p_{emp,v_{Crp,bound}}$	Max. expression rate	4.7 E-07 $g_{Prot}(g_{DW}\,s)^{-1}$
$p_{emp,K_{Crp}}$	Affinity constant	0.012 $g_{Prot}\,g_{DW}{}^{-1}$
$p_{eno,v_{Cra,unbound}}$	Max. expression rate	6.7 E-07 $g_{Prot}(g_{DW}\,s)^{-1}$
$p_{eno,v_{Cra,bound}}$	Basal expression rate	0 $g_{Prot}(g_{DW}\,s)^{-1}$
$p_{eno,K_{Cra}}$	Affinity constant	0.016 $g_{Prot}\,g_{DW}{}^{-1}$
$p_{fdp,v_{Cra,unbound}}$	Basal expression rate	0 $g_{Prot}(g_{DW}\,s)^{-1}$
$p_{fdp,v_{Cra,bound}}$	Max. expression rate	2.1E-08 $g_{Prot}(g_{DW}\,s)^{-1}$
$p_{fdp,K_{Cra}}$	Affinity constant	1.18E-03 $g_{Prot}\,g_{DW}{}^{-1}$
$p_{gltA,v_{Crp,unbound}}$	Basal expression rate	0 $g_{Prot}(g_{DW}\,s)^{-1}$
$p_{gltA,v_{Crp,bound}}$	Max. expression rate	6.5E-07 $g_{Prot}(g_{DW}\,s)^{-1}$
$p_{gltA,K_{Crp}}$	Affinity constant	0.04 $g_{Prot}\,g_{DW}{}^{-1}$
$p_{gltA,n}$	Hill coefficient	1.07
$p_{icd,v_{Cra,unbound}}$	Basal expression rate	1.1E-07 $g_{Prot}(g_{DW}\,s)^{-1}$
$p_{icd,v_{Cra,bound}}$	Max. expression rate	8.5E-07 $g_{Prot}(g_{DW}\,s)^{-1}$
$p_{icd,K_{Cra}}$	Affinity constant	1.17E-03 $g_{Prot}\,g_{DW}{}^{-1}$
$p_{mdh,v_{Crp,unbound}}$	Basal expression rate	0 $g_{Prot}(g_{DW}\,s)^{-1}$
$p_{mdh,v_{Crp,bound}}$	Max. expression rate	1.3 E-06 $g_{Prot}(g_{DW}\,s)^{-1}$
$p_{mdh,K_{Crp}}$	Affinity constant	0.06 $g_{Prot}\,g_{DW}{}^{-1}$
$p_{pckA,v_{Cra,unbound}}$	Basal expression rate	0 $g_{Prot}(g_{DW}\,s)^{-1}$
$p_{pckA,v_{Cra,bound}}$	Max. expression rate	3.7 E-07 $g_{Prot}(g_{DW}\,s)^{-1}$
$p_{pckA,K_{Cra}}$	Affinity constant	5.35E-03 $g_{Prot}\,g_{DW}{}^{-1}$
$p_{pdh,v_{PdhR,unbound}}$	Max. expression rate	7.7 E-08 $g_{Prot}(g_{DW}\,s)^{-1}$
$p_{pdh,v_{PdhR,bound}}$	Basal expression rate	2.8 E-10 $g_{Prot}(g_{DW}\,s)^{-1}$
$p_{pdh,K_{PdhR}}$	Affinity constant	3.4E-03 $g_{Prot}\,g_{DW}{}^{-1}$
$p_{pfkA,v_{Cra,unbound}}$	Max. expression rate	1.4 E-06 $g_{Prot}(g_{DW}\,s)^{-1}$
$p_{pfkA,v_{Cra,bound}}$	Basal expression rate	1.1 E-08 $g_{Prot}(g_{DW}\,s)^{-1}$
$p_{pfkA,K_{Cra}}$	Affinity constant	6.3E-07 $g_{Prot}\,g_{DW}{}^{-1}$
$p_{ppsA,v_{Cra,unbound}}$	Basal expression rate	0 $g_{Prot}(g_{DW}\,s)^{-1}$
$p_{ppsA,v_{Cra,bound}}$	Max. expression rate	3.3E-06 $g_{Prot}(g_{DW}\,s)^{-1}$

... Table 3.5 continued.

Parameter	Description	Value
$p_{ppsA,K_{Cra}}$	Affinity constant	0.017 $g_{Prot}\,g_{DW}{}^{-1}$
$p_{pykF,v_{Cra,unbound}}$	Max. expression rate	1.6 E-07 $g_{Prot}(g_{DW}\,s)^{-1}$
$p_{pykF,v_{Cra,bound}}$	Basal expression rate	8.8 E-10 $g_{Prot}(g_{DW}\,s)^{-1}$
$p_{pykF,K_{Cra}}$	Affinity constant	2.3E-03 $g_{Prot}\,g_{DW}{}^{-1}$
$p_{D,k_{degr}}$	Universal protein degradation rate	2.8E-05 s^{-1}
$p_{BM,k_{expr}}$	Gene expression rate constant	2.0E+04 s
$p_{BM,\mu_{ACT}}$	Growth rate on acetate	5.6E-05 s^{-1}
$p_{BM,\mu_{GLC}}$	Growth rate on glucose	1.8E-04 s^{-1}
$p_{BM,GLC_{ACoA}}$	1st order rate constant	1.88 s^{-1}
$p_{BM,GLC_{AKG}}$	1st order rate constant	0.978 s^{-1}
$p_{BM,GLC_{G6P}}$	1st order rate constant	0.154 s^{-1}
$p_{BM,GLC_{OAA}}$	1st order rate constant	6.4 s^{-1}
$p_{BM,GLC_{PEP}}$	1st order rate constant	0.423 s^{-1}
$p_{BM,GLC_{PG3}}$	1st order rate constant	0.049 s^{-1}
$p_{BM,GLC_{PYR}}$	1st order rate constant	0.553 s^{-1}
$p_{BM,ACT_{ACoA}}$	1st order rate constant	0.108 s^{-1}
$p_{BM,ACT_{AKG}}$	1st order rate constant	0.056 s^{-1}
$p_{BM,ACT_{G6P}}$	1st order rate constant	0.076 s^{-1}
$p_{BM,ACT_{OAA}}$	1st order rate constant	1.43 s^{-1}
$p_{BM,ACT_{PEP}}$	1st order rate constant	0.047 s^{-1}
$p_{BM,ACT_{PG3}}$	1st order rate constant	0.066 s^{-1}
$p_{BM,ACT_{PYR}}$	1st order rate constant	5.185 s^{-1}

3.6.3 Sensitivity analysis

The sensitivities of each of the two reproduced steady states with respect to small parameter perturbations were approximated through

$$S_{i,j} = \frac{\Delta x_i / x_i}{\Delta p_j / p_j} \approx \frac{(x_i^+ - x_i^-)/x_i}{(1.01\,p_j - 0.99\,p_j)/p_j} = \frac{x_i^+ - x_i^-}{0.02\,x_i} \quad,$$

with x_i the nominal steady state value and x_i^+ and x_i^- the values resulting from a 1% increase or decrease, respectively, of the parameter value p_j. As a heatmap of the resulting 47x193 sensitivity matrix S is too large for convenient inspection, we reduced the size of S by grouping the sensitivities of those parameters that appear in the same rate equation f_k. For this grouping, we used the size-independent overall variability
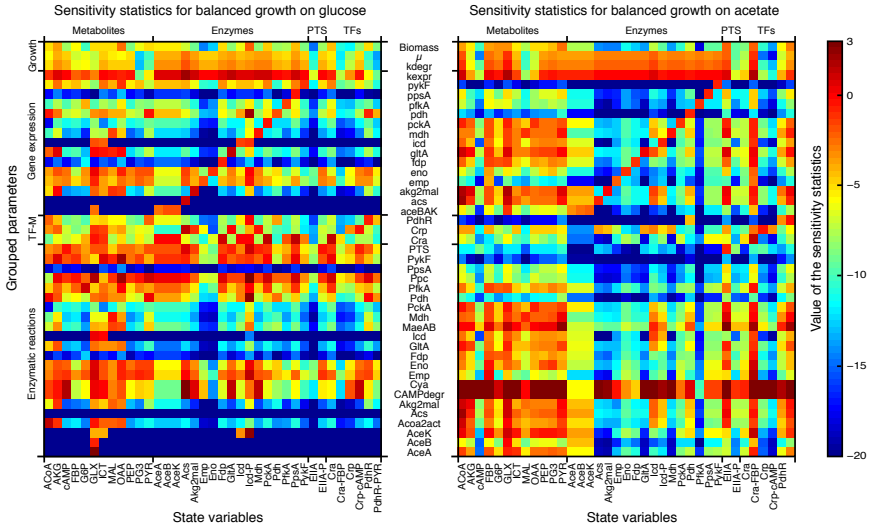
Figure 3.8: Sensitivity statistics of the steady states with respect to 1% parameter perturbations. The used statistics condenses the sensitivities of all parameters appearing in the same rate equation to a single value.

$$S_{i,k} = \frac{1}{n_k} \sum_{j \,|\, p_j \in f_k} S_{i,j}^2$$

as sensitivity statistics, with n_k the number of parameters appearing in f_k. The results of these local sensitivity analyses, one for each of the two steady states, are plotted in Figure 3.8.

This figure reveals that in general, protein and metabolite concentrations respond differently to parameter perturbations. Protein concentrations are most significantly affected by perturbations of the parameters describing their gene's expression but are rather marginally affected by other parameter perturbations, causing the red diagonal line of high protein sensitivities in Figure 3.8. In contrast, the sensitivities of the metabolite concentrations do not form such a line. Therefore, contrary to what might have been expected, the sensitivities of metabolite concentrations to parameter variations of their topological enzyme neighbors and to more distant perturbation are about the same. We suspect that this property of the metabolites' sensitivities is due to the model's densely interconnected enzymatic regulation.

A further observation is that sensitivities that are high on glucose are often low on acetate and vice versa, probably because many reactions are very active in one condition but much less so in the other.

3.6.4 Simulation results

In this section, we present the simulated time profiles of the intracellular metabolites (Figure 3.9), enzymes (Figure 3.10), rates (Figures 3.11 and 3.12), the phosphorylation state of the proteins EIIA and Icd, and the TF activities (Figure 3.13). Figure 3.14 contains a subset of these time profiles to illustrate the connection between the available extracellular carbon sources, the concentrations of the intracellular signal metabolites, and the activities of their target TFs. These time profiles are based on the parameter vector listed in Table 3.5. Note that due to the uncertainty in the parameter values, the ambiguity in the selection of rate laws, and possible effects of not modeled cellular regulations, i.e. those ensuring the here omitted energy, cofactor etc. balances, onto the simulated trajectories, these profiles cannot be considered a quantitatively accurate reproduction of the *in vivo* time profiles. Instead, these profiles are meant to demonstrate that with the chosen parameters, (i) all compound concentrations and rates approach the proper steady state levels and (ii) remain within physiologically reasonable bounds throughout the transitions.

An aspect worth noting is the spiking of some metabolite concentrations following environmental changes. These spikes are in the order of seconds and thus elude experimental detection with current measurement technology. In some cases such as OAA, these spikes might be an artifact of the modeling, caused by the fast change of the used first order rate constants upon a change in environmental conditions (see Section 3.6.1.9). In other cases, however, the spikes arise from the interplay of the mechanistically modeled interactions. For instance, PEP and PYR spike because they are coupled to glucose uptake via the PTS system. These spikes propagate through the system: G6P's small upward spike following the change from glucose to acetate is a consequence of the large upward PEP spike quickly inhibiting the G6P–consuming reaction (in the model, PfkA) while quickly activating the G6P–producing (back-)reaction from FBP (in the model, Fdp); G6P's steady state convergence after that small, initial spike is slow because G6P is immediately formed by the uptake of glucose but needs many reaction steps to be synthesized from acetate.

Another aspect worth mentioning is that the *duration* of the transition period between balanced growth on either substrate is primarily determined by the growth rates and is therefore robust to the inherent uncertainty in the model. The reason for this is that the speed of steady state convergence is given by the slowest significant rates of the model, which are the enzyme dilution rates due to growth. This observation suggests that with μ the growth rate on the condition adapting to, a lower bound (that neglects active protein degradation) for the transcriptional adaptation time of bacteria could be quantified, e.g. in the form of a 'minimal half-adaptation time' $T_{1/2} = \mu^{-1} ln(2)$.
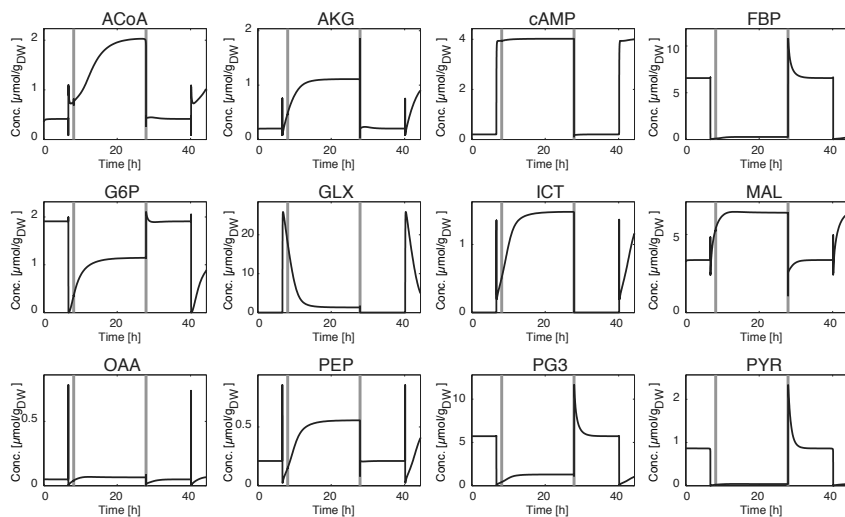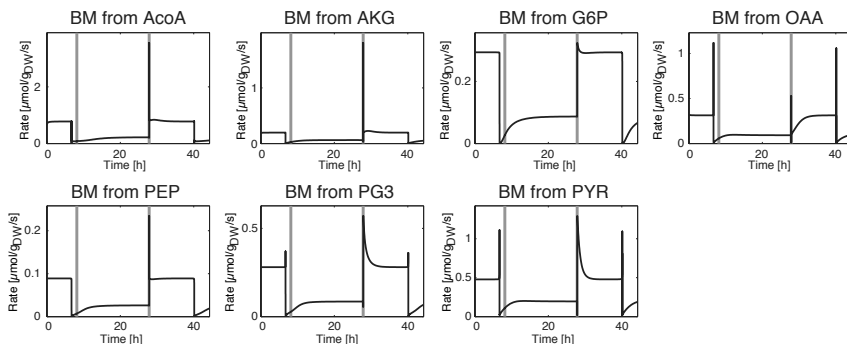
Figure 3.9: Simulated intracellular metabolite concentrations. The vertical gray lines indi-
cate the time instants of external changes to the available carbon sources (see Figure 3.14
or Section 3.4.1). The concentrations approach the measured steady state values listed in
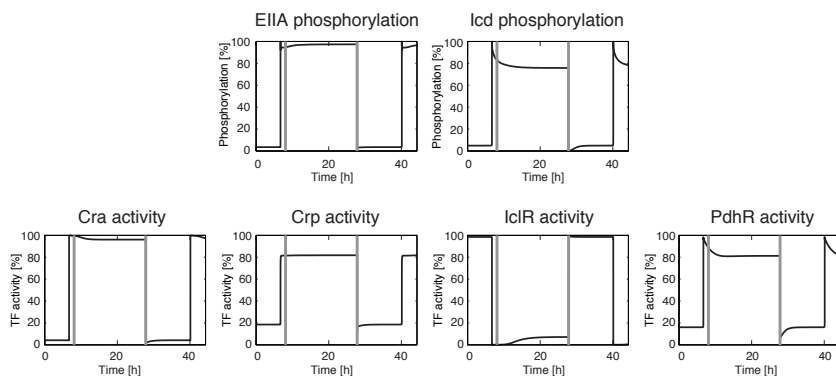Table 3.3.
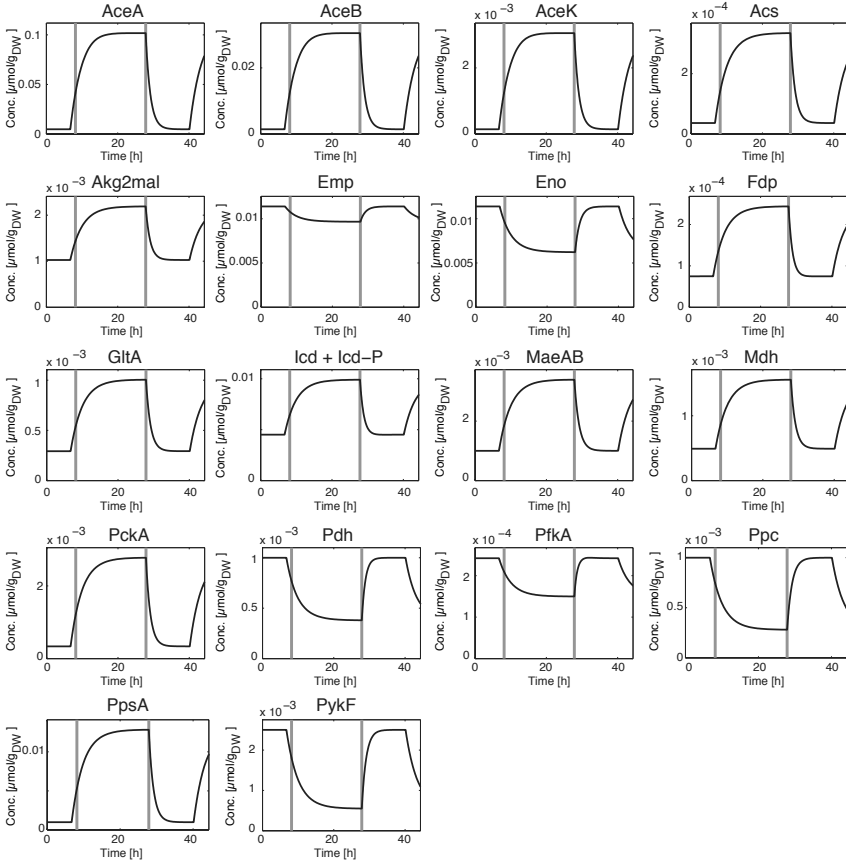
Figure 3.10: Simulated enzyme concentrations. The vertical gray lines indicate the time instants of external changes to the available carbon sources (see Figure 3.14 or Section 3.4.1). The concentrations approach the measured steady state values listed in Table 3.3.
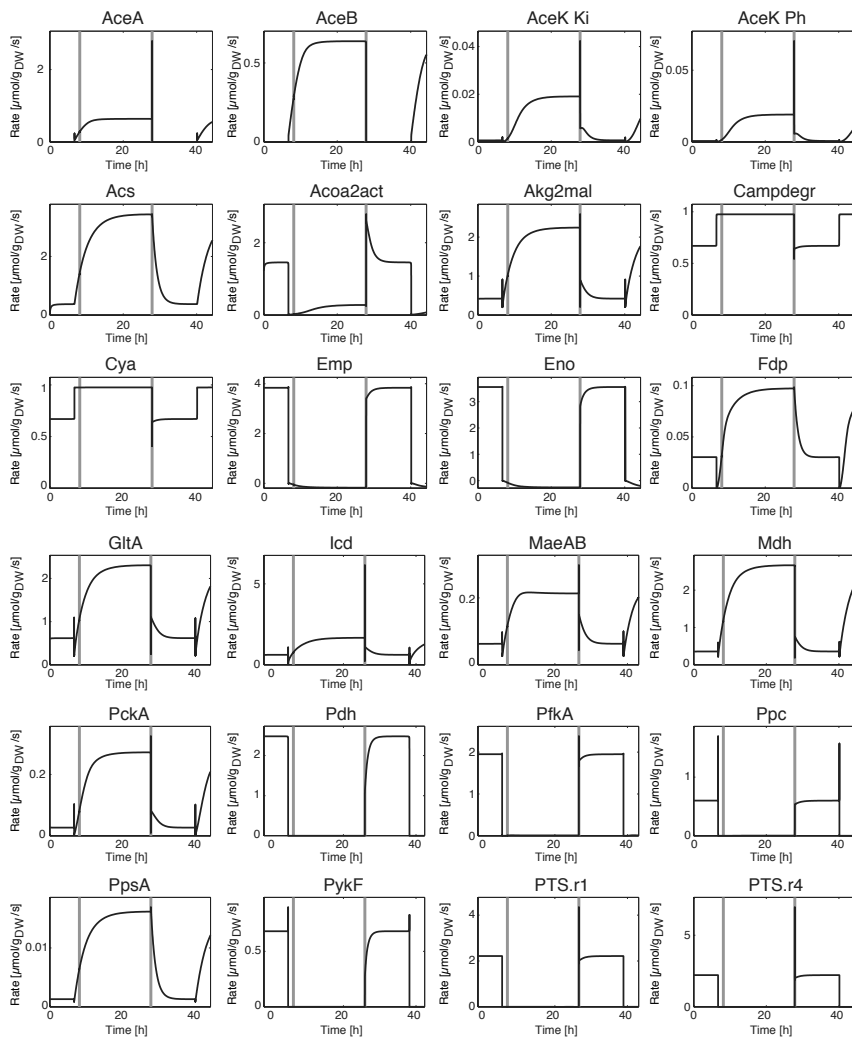
Figure 3.11: Simulated enzymatic reaction rates. The vertical gray lines indicate the time instants of external changes to the available carbon sources (see Figure 3.14 or Section 3.4.1). The rates approach the measured steady state values listed in Table 3.4.

Figure 3.12: Simulated biomass reaction rates. The vertical gray lines indicate the time instants of external changes to the available carbon sources (see Figure 3.14 or Section 3.4.1). The rates approach the measured steady state values listed in Table 3.4.

Figure 3.13: Simulated phosphorylation states of the proteins EIIA and Icd, and simulated TF activities. The vertical gray lines indicate the time instants of external changes to the available carbon sources (see Figure 3.14 or Section 3.4.1). The phosphorylation states and TF activities approach the measured steady state values listed in Table 3.3.
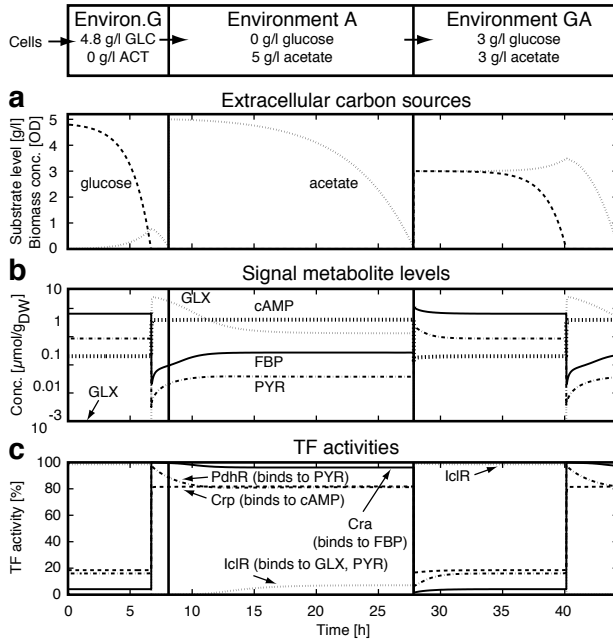
Figure 3.14: The concentrations of the four signal metabolites and the activities of their four target TFs respond to the presence of extracellular carbon sources. **a**, Levels of extracellular carbon sources. **b**, Levels of intracellular signal metabolites. **c**, Activities of the signal metabolites' target TFs.

3.7 Acknowledgments

Chapter 4

Phenotypic bistability in Escherichia coli's central carbon metabolism

Oliver Kotte, Benjamin Volkmer, Sven Panke and Matthias Heinemann

4.1 Summary

Stochastic gene expression causes fluctuations in the abundances of intracellular molecules [28, 35, 50, 143, 146, 151, 167, 203]. Certain regulation architectures can exploit this cell–to–cell variation to create multiple distinct and co-existent phenotypes even among isogenic cells exposed to the same environment [12, 37, 76, 87, 137, 152, 197]. Experimental [2] and complementary theoretical studies [107, 205] suggest that stochasticity–driven systems can prevail

over deterministic designs to confer high population adaptability. Although metabolism continuously adapts to unpredictable environmental changes, and certain substrate uptake pathways were found to exhibit phenotypic bistability [1, 144], central carbon metabolism as a whole is thought to operate deterministically. Here we report that a regulation architecture at the core of central carbon metabolism in *Escherichia coli* splits an isogenic cell population into two phenotypic subpopulations. We found that after a shift from glucose to a gluconeogenic substrate, only a relatively small fraction of cells grows, whereas most cells unexpectedly refrains from growing on the available substrate. We identified the underlying subpopulation–generating mechanism, which comprises two intertwined feedback and feedforward loops on the levels of metabolic and genetic regulation. This circuit 'chooses' between two alternative phenotypes based on measured carbon uptake performance and as such realizes a *performance-based phenotype selection*, assigning only the cells most capable of gluconeogenic growth to the growing phenotype. These results suggest that phenotypic bistability generalizes to central metabolism, and can thus be viewed as an inherent design feature. Metabolic regulation does not ensure the growth of individual cells on a gluconeogenic substrate, but instead uses the strategy of *responsive diversification* to diversify the population once glucose is depleted.

4.2 Introduction

Since the earliest bacterial physiology studies, it has been known that inoculation of a bacterial population into a new medium results in a period of no apparent growth prior to growth on the new carbon source [133]. This 'lag time' is classically attributed to the duration of necessary biochemical adaptation processes within individual cells, which are thought to switch homogeneously and responsively to the new substrate (*responsive switching*, see Figure 4.1).

However, biochemical processes are inherently stochastic and cause molecule abundances to fluctuate. These fluctuations are usually suppressed but can also be amplified and used to generate distinct phenotypes [12, 37, 76, 87, 137, 152, 197]. The possible emergence of multiple coexisting phenotypes within an *isogenic* cell population creates the alternative and hitherto untested hypothesis that the apparent 'lag time' is caused by the exclusive growth of an initially small phenotypic subpopulation (see Figure 4.1).

A growing phenotypic subpopulation could be generated in two ways. First, the cells could 'anticipate' the environmental change by stochastically switching their phenotype at any time, in which case the adaptation to the new carbon source is a passive process accomplished by the presence of pre–adapted cells prior to the environmental change [2, 107, 205] (see Figure 4.1). This mechanism of adaptation, termed *stochastic switching* [107], resembles that of type II persister cells [12]. Second, an initially homogeneous population could actively respond to the sensed environmental change with a diversification of its phenotype, meaning that only

a stochastically generated subset of the cells quickly adapts to growth on the new conditions. This mechanism of adaptation, here termed *responsive diversification*, resembles that of type I persister cells, which acquire antibiotic resistance as a response to an environmental trigger [12].

Here, we investigated which of these three adaptation strategies — responsive switching, responsive diversification, or stochastic switching — is used by *Escherichia coli* to adapt from glycolytic to gluconeogenic growth. On the molecular level, this adaptation requires comprehensive redirections of central carbon flow, which *E.coli* realizes by repressing glycolytic enzymes and inducing gluconeogenic enzymes, particularly those of the acetate uptake pathway (*acs*), the glyoxylate shunt (*aceBAK*), the anaplerotic reactions (*pckA,maeB,sfcA*), and the Embden-Meyerhoff pathway [92].
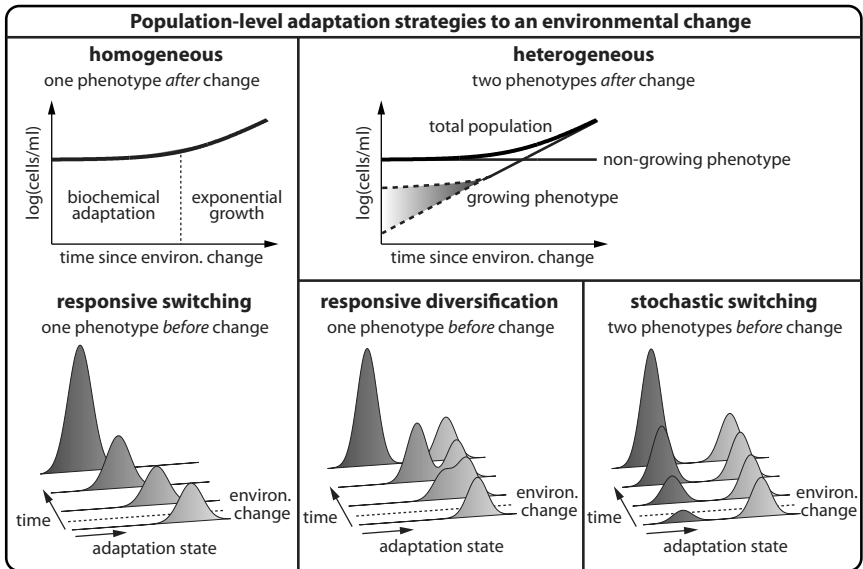
Figure 4.1: Three alternative hypotheses to explain the lag time in the population's growth curve after an environmental change. According to the classical hypothesis, the population responds with a homogeneous adaptation, which takes considerable time before growth is resumed. According to the here presented heterogeneous subpopulation hypothesis, only an initially small subpopulation resumes growth. This phenotypic diversification is either generated from a homogeneous population as a response to the environmental change, or exists already before the environmental change due to stochastically switching cells.

4.3 Results

4.3.1 Phenotypic bistability in central metabolism

To discriminate whether the adaptation of *E. coli*'s central metabolism from glycolytic to gluconeogenic growth involves one or two phenotypes, we conducted substrate shift experiments where we stained the cellular membranes with a red fluorescent, membrane-intercalating dye before changing the carbon source from glucose to different concentrations of acetate (see Figure 4.2a). Hence, a cell's fluorescent intensity is halved with every cell division in the new environment. At multiple time points after the shift to acetate, we determined the populations' fluorescence intensity distributions (see Figure 4.2b). For all acetate concentrations, we do not observe unimodal fluorescence intensity distributions with exponentially decreasing mean values following an initial period of adaptation, which would validate the classic hypothesis (Figure 4.1). Instead, upon the substrate shift to acetate, two subpopulations become visible; cells of the first subpopulation retain a high fluorescence level and therefore do not grow, cells of the second subpopulation lose their fluorescence with a rate approaching the total population's growth rate and therefore do grow (dye toxicity as a possible cause for the non–growing phenotype has been excluded, see dissertation of Benjamin Volkmer). The cause for the two phenotypic subpopulations is not genetic; if the experiment is repeated with cells derived from the growing subpopulation, again two distinct growth phenotypes emerge (see dissertation of Benjamin Volkmer). Therefore, the adaptation from glucose to acetate involves two distinct growth phenotypes, and thus the adaptation strategy cannot be responsive switching.

Next, we discriminated between the remaining two adaptation strategies to determine whether the two phenotypes arise from a passive process (stochastic switching) or an active response (responsive diversification). We thus grew *E. coli* on both glucose and ^{13}C–labelled acetate and measured the intracellular amino acids' ^{13}C enrichment patterns. These patterns (see Section 4.5.1) show that if an acetate–adapted subpopulation with dysfunctional catabolite repression exists at all in the presence of glucose, its size is much too small to generate the growing subpopulation observed in our substrate shift experiment; thus, stochastic switching on glucose can at best generate a small fraction of the growing subpopulation on acetate. Further, we found no experimental evidence for the occurrence of stochastic switching on acetate (see dissertation of Benjamin Volkmer). Hence, the *dominant* mechanism of adaptation must be responsive diversification: When faced with a shift from glucose to acetate, central metabolism responds with diversifying a homogeneous cell population such that only a subpopulation adapts to growth on the new substrate.

The stability of the two phenotypes allowed us to quantify the subpopulation proportions generated by responsive diversification following the substrate shift. For multiple acetate concentrations, estimates for the fractions of acetate–adapting cells at the time instant of the substrate shift, α, and their growth rates μ_g were
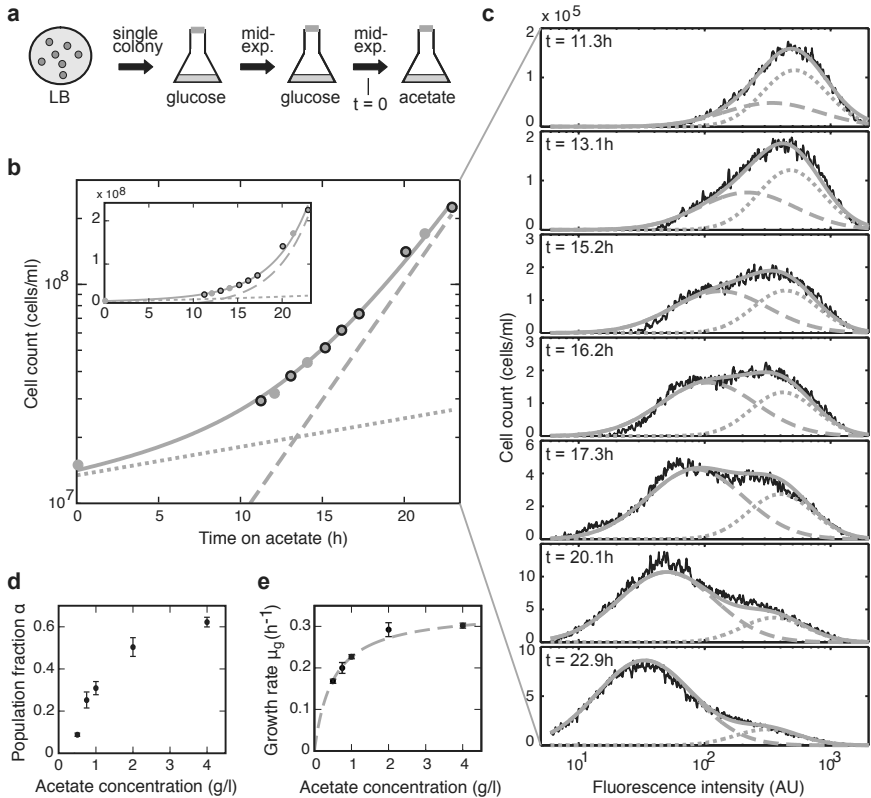
Figure 4.2: Substrate shift experiments from glucose to acetate. **a**, Outline of the experimental procedure. **b**, Measured cell counts of the total population (gray dots) after the substrate shift from glucose to acetate, and deduced growth curves of the total population (regular line) and constituent 'growing' (dashed line) and 'non–growing' (dotted line) subpopulations. Dots with a black circumference indicate the time points for flow cytometry analysis. **c**, Fluorescence intensity distributions at multiple time points after the substrate shift. A Bigaussian fit (dashed gray line, 'growing' subpopulation; dotted gray line, 'non–growing' subpopulation; regular gray line, total population) excellently reproduces the experimental data (black curve). **d**, The fraction of acetate–adapting cells, α, increases with the acetate concentration. **e**, The dependency of the 'growing' subpopulation's growth rate μ_g on the acetate concentration is hyperbolic, with the maximal growth rate $v_{max} = 0.34h^{-1}$ and the Monod constant $K_{Acetate} = 0.5gl^{-1}$.

determined through deconvolution of the two phenotypes' Gaussian contributions to the time progression of the total populations' fluorescence intensity distributions (see Section 4.6.2). We found that the value of α increases with increasing acetate concentration (excluding acidic shock survival as a possible cause for the subpopulations) and levels off at $\alpha \approx 0.5$ (see Figure 4.2d), meaning that for small acetate concentrations, most cells refrain from growing on the present acetate. The value of μ_g follows a hyperbolic dependency on the acetate concentration (see Figure 4.2e), with the maximal growth rate $\mu_{max} = 0.34h^{-1}$ and the Monod constant $K_{Acetate} = 0.5gl^{-1}$, agreeing with an earlier published value [230]. Therefore, rather then being attributable to biochemical adaptation processes within individual cells, the duration of the apparent 'lag time' of total population growth is a combined effect of the values of α and μ_g.

4.3.2 Molecular mechanism generating bistable phenotypes

4.3.2.1 Phenotypic differences on the molecular level

Towards identifying the molecular mechanism generating the responsive diversification, we asked how the two phenotypes differ on the molecular level. One of the necessary regulatory adjustments to adapt from glucose to acetate is performed by the transcription factor Cra, which regulates the reversal of carbon flow through the Embden-Meyerhoff-pathway [178].

To determine if Cra is involved in the generating the responsive diversification, we modulated the Cra abundance through expression from an inducible plasmid in a cra^- background (without this plasmid, cra^- is not viable on acetate, see Figure 4.3e). We found that following the substrate shift to acetate, the total population's growth curve exhibits a markedly prolonged 'lag time' compared to the wild–type when the plasmid is uninduced, whereas growth with the fully induced plasmid reproduces wild–type physiology. When we analyzed this behavior on the single cell level, we found that the variation in 'lag time' arises from a variation of α, not μ_g (see Figure 4.3e). These results show that the abundance of Cra modulates the subpopulation proportions, and therefore strongly suggest that Cra is involved in the molecular mechanism generating the responsive diversification.

4.3.2.2 Identification of a candidate mechanism

Next, bistability theory states that if Cra is involved in generating bistable phenotypes, then it must be an element of a feedback architecture involving cooperative binding events [76, 87, 152, 193]. We therefore screened the known molecular interactions for such a feedback architecture. Since Cra is not known to be transcriptionally regulated, a transcriptional feedback on Cra expression is unlikely. However, Cra activity is next to the Cra abundance also affected by the concentration of Cra's strong inhibitor, the metabolite fructose-1,6-bisphosphate (FBP) [162]. With this interaction, we identified a candidate mechanism for the generation of responsive diversification.
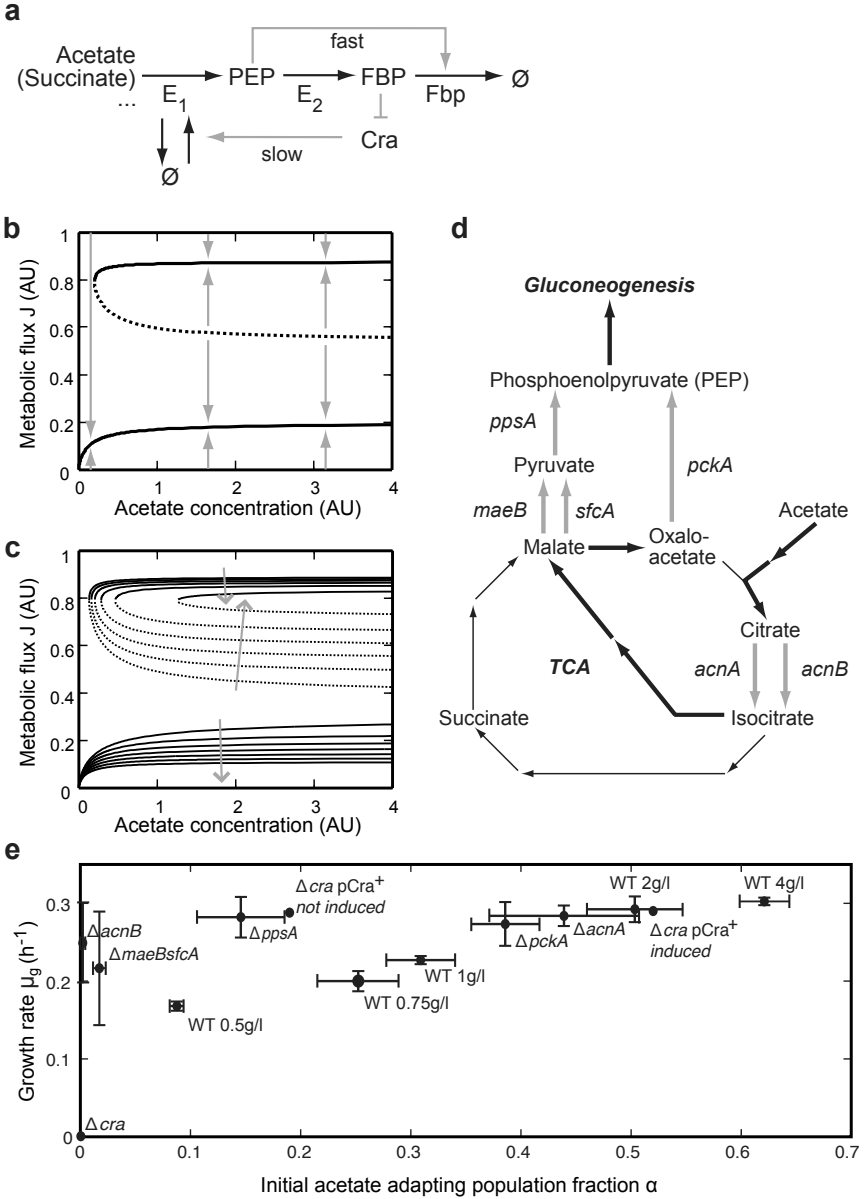
This candidate mechanism, illustrated in Figure 4.3a, comprises a feedback intertwined with a feedforward loop and overarches the metabolic and genetic layers. The isolated feedback of FBP via Cra on the expression of enzymes subsumed into a super–enzyme E_1 is negative; the sign of the closed loop is turned positive by the strong feedforward activation of the enzyme fructose-1,6-bisphosphatase I (Fbp) by phosphoenolpyruvate (PEP) [75], which causes the FBP-producing and consuming reactions to compete for PEP such that FBP *decreases* with increasing PEP and gluconeogenic flux (see Section 4.7.2 and Figure 4.8). The super–enzyme E_1 catalyzes the interconversion of acetate to PEP via the TCA cycle, the glyoxylate shunt, and the anaplerotic reactions.

A bifurcation analysis on a differential equation model of this system confirmed that this candidate mechanism is capable of generating responsive diversification (see 4.7.2). The bifurcation diagram shows two stable acetate uptake fluxes (Figure 4.3b), which can be interpreted as the 'growing' and 'non–growing' phenotypes, and which are separated by a watershed. If, upon the substrate shift, a particular cell immediately realizes an acetate uptake flux above the watershed, it approaches the high steady state and adopts the 'growing' phenotype; otherwise, it approaches the low steady state and adopts the 'non–growing' phenotype. If intracellular randomness causes the relative positions of acetate uptake flux and watershed to vary throughout the population, bistable phenotypes result, and the responsive diversification is generated.

Notably, this mechanism establishes a dependency of FBP on the gluconeogenic flux and propagates the flux information to the genetic layer through binding of FBP to Cra. This process has been analyzed in Chapter 3 and constitutes a *flux measurement* realized by the Cra–FBP *flux sensor*. Remarkably, this flux measurement realizes a *performance–based selection* of the growing phenotype, which is adopted only by cells with a sufficiently high (above-watershed) acetate uptake flux.

4.3.2.3 Validation of the candidate mechanism

The differential equation model's system properties are consistent with the above presented experimental observations. First, higher acetate concentrations in the range of its Monod constant increase the cells' acetate uptake fluxes and potentially lift these above the watershed, increasing α as observed in Figure 4.2d. Second, overexpression of E_1's transcriptional activator Cra potentially raises the production rate of E_1, which increases α by lowering the watershed (the level of the watershed is given by the production rate of E_1, see Section 4.7.2). Third, a knockout of Cra reduces α by raising the watershed. Fourth, the required capacity for *immediate* acetate uptake upon glucose removal is strongly suggested by the results of our ^{13}C experiment (see Section 4.5.1), which show that acetate is taken up already prior to glucose removal — following glucose removal, cessation of glycolytic flux allows the carbon flux derived from acetate, so far cycling through the TCA, to enter the Embden–Meyerhoff pathway as gluconeogenic flux.

a

Acetate
(Succinate)
... E_1 → PEP E_2 → FBP → Fbp → Ø

fast

slow

Cra

Ø

b

c

d

Gluconeogenesis

Phosphoenolpyruvate (PEP)

ppsA

pckA

Pyruvate

maeB *sfcA*

Malate → Oxalo-acetate

Acetate

Citrate

acnA *acnB*

TCA

Succinate

Isocitrate

e

Further observations yield additional confidence in the candidate mechanism. First, we determined the population–averaged Cra abundance to be in the range of 50–100 molecules/cell. This abundance, which is very low compared to most other TF abundances [175], encourages noticeable cell–to–cell variation [14]. Second, the gradual convergence to the two stable acetate uptake fluxes in the bifurcation diagram agrees with an experimentally observed initial, gradual increase or decrease of the two phenotypes' growth rates upon the substrate shift (see dissertation of Benjamin Volkmer).

Next, we challenged the candidate mechanism through experimental validation of predicted perturbation effects. First, a knockout of isoenzymes (*acnA, acnB*) or parallel pathways (*maeBsfcA, ppsA, pckA*) located within E_1 (see Figure 4.3d) *may* (i) reduce the overall production rate of E_1 through the introduction of 'gene expression bottlenecks', which reduce α by raising the level of the watershed (see Figure 4.3 and Section 4.7.2), and/or (ii) reduce the acetate uptake flux through the introduction of 'metabolic flux bottlenecks', which certainly decrease α and possibly also μ_g. Hence, these mutations have either no effect or reduce α more significantly than μ_g. This prediction is confirmed by the results of substrate shift experiments with these mutants (see Figure 4.3e).

Second, if the growth behavior of these mutants differs from that of the wild-

---

Figure 4.3 (preceding page): The bistability–generating mechanism. **a**, Bistability–generating circuit. E_i denote super–enzymes catalyzing lumped reactions. PEP (phosphoenolpyruvate) allosterically activates the enzyme Fbp (fructose-1,6-bisphosphatase). Fructose-1,6-biphosphate (FBP) represses E_1 production by modulating the activity of the transcription factor Cra. **b**, Bifurcation diagram of metabolic steady state fluxes J as a function of extracellular acetate concentration. The system is capable of expressing two stable steady state fluxes (bold lines) and one unstable steady state flux (dashed line) that acts as a watershed separating the convergence regions of the high ('growing' phenotype) and low ('non-growing' phenotype) stable steady states. Arrows show the direction of system dynamics. **c**, A set of bifurcation diagrams for different E_1 production rates. Arrows indicate the direction of decreasing E_1 production rates with which the convergence region of the 'growing' phenotype gradually decreases and that of the 'non–growing' phenotype increases (i.e. α decreases), and the steady state flux J (i.e. the growth rate) of the 'growing' phenotype stays nearly unchanged. **d**, The reactions catalyzed by the super–enzyme E_1. Bold arrows indicate the route of major carbon flux from acetate to PEP, thin arrows complete the TCA cycle. Gray arrows highlight isoenzymes and parallel pathways that are knockout targets for experimentally decreasing the E_1 production rate. **e**, Experimentally derived values for the 'growing' subpopulation fraction α and its growth rate μ_g for different strains. Unless otherwise stated, the initial acetate concentration is 2g/l. In the wild–type, both α and μ_g increase with increasing acetate concentration. Also shown are the behaviors of the *cra* knockout mutant without plasmid (not viable on acetate), with plasmid pCra$^+$ not induced (marked decrease of α but not μ_g), and with plasmid pCra$^+$ induced using an IPTG concentration of 0.1mM (α and μ_g are reverted to wild–type behavior). Further shown are the behaviors of knockout mutants that potentially have a lower E_1 production rate than the wild–type. In the *acnB*, *ppsA*, and *maeBsfcA* mutant strains, the dominant effect is a reduction of α. The *acnA* and *pckA* mutant strains exhibit near–wild–type behavior.

type, the wild-type behavior should be restored through overexpression of the alternative isoenzyme or parallel pathway. Figure 4.3e shows that the $ppsA^-$, $maeBsfcA^-$ and $acnB^-$ strains exhibit markedly reduced values of α. Experiments confirm that $ppsA^-$ is fully reverted to wild-type behavior through overexpression of $pckA$ [88], and that the extensive apparent 'lag times' of $maeBsfcA^-$ and $acnB^-$ are markedly reduced through overexpression of $ppsA$ [88] and $acnA$ (see dissertation of Benjamin Volkmer), respectively.

Third, because merging of the reactions from other gluconeogenic substrates such as succinate, malate, and fumarate to PEP into super–enzymes E_1^* leads to the same topology, the model predicts that responsive diversification also occurs upon shifts to these substrates. The results of the corresponding substrate shift experiments, plotted in Figure 4.7 of Section 4.5.2, confirm this prediction. These results also show that responsive diversification is not an acetate–specific phenomenon and is therefore of a more general nature.

We conclude that upon shifts from glucose to multiple gluconeogenic substrates, responsive diversification is generated at the core of central metabolism. Therefore, phenotypic bistability generalizes to central metabolism and can thus be viewed as an inherent feature of its design.

4.4 Discussion

In this chapter, we have shown that an isogenic *E. coli* population responsively diversifies into two phenotypic subpopulations following shifts from glucose to multiple gluconeogenic substrates. As 'lag time' has here been shown to be caused by the exclusive growth of an initially small phenotypic subpopulation, other occurrences of 'lag time' may arise from multiple phenotypes as well.

The molecular mechanism generating the responsive diversification resides at the core of central metabolism and comprises a feedback architecture overarching the metabolic and genetic layers. This mechanism realizes a performance–based selection of phenotypes through flux measurements. Hence, metabolic flux is a control factor that regulates gene expression, a function that has so far only been demonstrated synthetically [58].

The main result of this chapter is that phenotypic bistability generalizes to central metabolism and can thus be viewed as an inherent feature of its design.

4.5 Supporting results

This section presents experimental results supporting the conclusions of this chapter. Specifically, the exclusion of stochastic switching as the dominant adaptation strategy and the behavior of the wild–type upon shifts to gluconeogenic substrates other than acetate are presented. Refer to the dissertation of Benjamin Volkmer for further supporting results, including

- the effect of culture evaporation on cell count concentration,

- the effect of dye bleaching on fluorescence intensity,

- the validation that membrane staining has no effect on growth behavior and hence the 'non–growing' subpopulation is not a result of dye toxicity,

- the validation that the two phenotypic subpopulations are indeed isogenic and hence the subpopulations do not arise from genetic mutations occurring during the experiment,

- the validation that stochastic switching between the 'growing' and 'non–growing' phenotypes does not occur to a noticeable degree on acetate, and

- the validation that the fraction of cells adapting to growth on acetate, α, is independent of the amount of acetate present in the second consecutive 'glucose' culture; hence, the presence of gluconeogenic substrates (most prominently acetate) in the 'glucose' culture prior to the substrate shift does not prime the cells for the later consumption of these carbon sources.

4.5.1 Exclusion of stochastic switching as the dominant adaptation strategy

This section investigates whether or not an acetate–adapted phenotypic subpopulation exists in the presence of glucose, prior to the substrate switch to acetate, with a size large enough to account for the α values observed in Figure 4.2. The answer to this question allows for a discrimination between the two possible population–level adaptation strategies of passive stochastic switching and active responsive diversification (see Section 4.3.1). From a textbook point of view, the acetate–adapted phenotype is not expected because in *E. coli* and several other bacteria, the usage of alternative carbon sources in the presence of glucose is blocked through a phenomenon known as carbon catabolite repression [200]. Thus, if a phenotype existed that grows on acetate but not on glucose, then in those cells carbon catabolite repression would not be active. Further note that such pre–adaptation to acetate *during* glucose growth is not expected because the value of the initial acetate–adapted population fraction, α, depends on the acetate concentration *after* the glucose growth phase (see Figure 4.2).

If a phenotypic subpopulation adapted to gluconeogenic growth nonetheless existed during the glucose growth phase in our standard transition experiments, we would expect it to be of a detectable size (and not in the range of one cell in few thousands): First, the presence mostly of acetate but also of other gluconeogenic substrates in the glucose culture (see Figure 4.4) would allow this phenotype to grow. Second, if the population–level adaptation strategy was sole passive stochastic switching, then this strategy alone would have to account for the significant population fraction α (see Figure 4.2) that 'decides' to grow on acetate when glucose is removed.
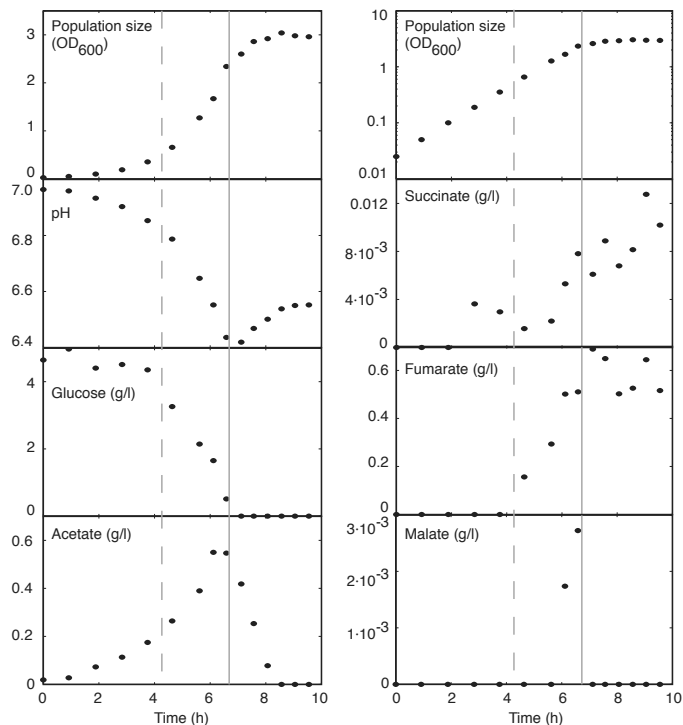
Figure 4.4: Physiological data for diauxic growth on glucose. The population size, measured as the culture's optical density, the pH, and the concentrations of extracellular carbon sources are shown. The dashed gray lines indicate the time instant at which cells are in standard substrate shift experiments transferred to a new medium with a gluconeogenic carbon source. At this time instant, several gluconeogenic carbon sources are already present in the glucose culture, allowing a hypothetical phenotypic subpopulation adapted to these substrates to grow. The solid vertical gray lines indicate the time instant of glucose depletion and the onset of net consumption of at least acetate.

To identify whether or not an acetate–adapted phenotypic subpopulation exists in the glucose growth phase, cells were grown in a M9 glucose medium preculture, transferred to an M9 glucose (5 g/L) medium main culture supplemented with 1 g/L either unlabeled (naturally) or fully ^{13}C–labeled acetate. If a phenotype existed that grows on acetate, then that phenotype would take up ^{13}C labelled acetate and we would find ^{13}C label in, for example, the intracellular amino acids. We took samples (triplicates) from the two main cultures at different optical densities:

- the 'actual sample' at an OD of 1.2 (glucose growth phase) from the culture with the added 1 g/L labeled acetate. This OD roughly represents the OD, at which we usually harvest the cells to inoculate the acetate medium in our standard transition experiments (OD 0.5).

- the 'negative control sample' at an OD of 1.2 (glucose growth phase) from the culture grown with the added 1 g/L unlabeled acetate.

- the 'positive control sample' at an OD of 3.3 from the culture with the added 1 g/L labeled acetate. At this OD the cells have consumed both available substrates, the glucose and the available acetate, consisting of the produced acetate and the added ^{13}C–labeled acetate, and have therefore entered stationary phase.

The measured mass distribution vectors of the amino acids, shown in Figure 4.5, reveal a natural labeling pattern for seven amino acids, and a ^{13}C–enriched labeling pattern for eight amino acids. Therefore, the cells do take up the acetate, but incorporate the carbon derived from the acetate only in a subset of the amino acids. As we know from Figure 4.4 that the cells excrete acetate at the same time, the acetate uptake and excretion must occur simultaneously. This is thermodynamically possible due to the existence of two separate pathways, with the uptake most likely performed by the enzyme Acs, and the excretion by the linear pathway formed by the enzymes Pta and AckAB.
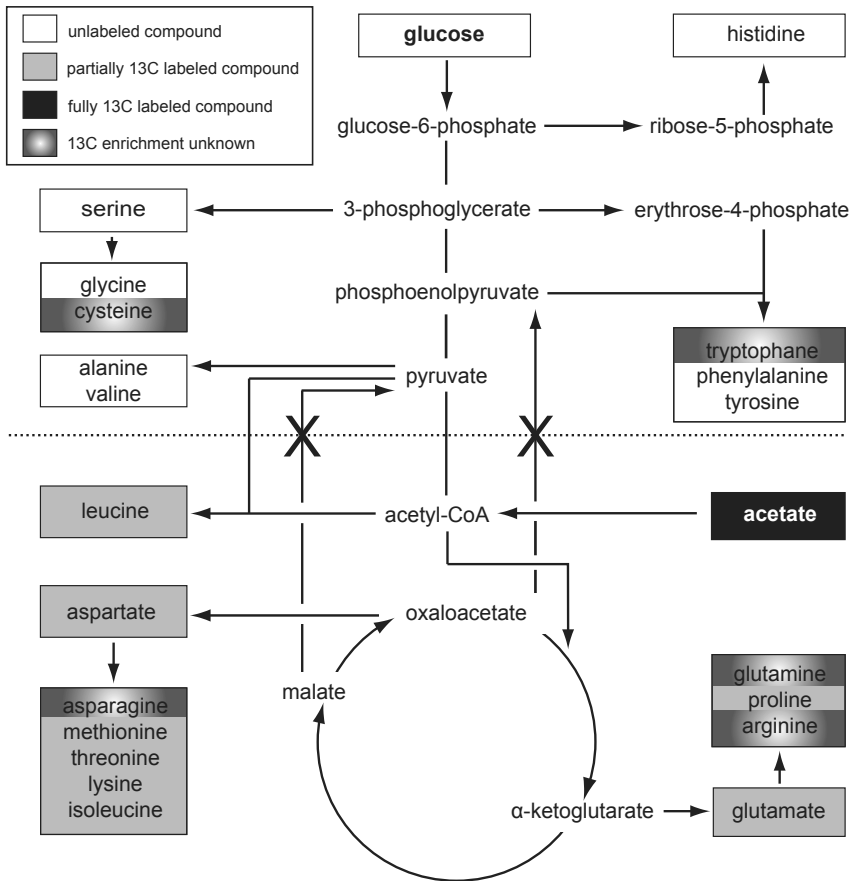
As amino acids are derived from precursor metabolites in central carbon metabolism, we mapped the amino acids to their respective precursor metabolites in central carbon metabolism (see Figure 4.6). The ^{13}C–enrichment occurs only in those amino acids that are derived from metabolites below PEP, but not in those amino acids that are derived from glycolytic intermediates above PEP. Therefore, the carbon derived from acetate is merely cycled through the TCA cycle but does not enter the Embden–Meyerhoff pathway in the gluconeogenic direction. As a phenotypic subpopulation growing on acetate but not on glucose would require gluconeogenesis to synthesize essential biomass precursors (such as the amino acids derived from glycolytic intermediates above PEP), we can conclude that a noticeable phenotypic subpopulation growing exclusively on acetate does not exist in the presence of glucose (as is expected due to the phenomenon of carbon catabolite repression [200]).

Note that due to the 10% uncertainty of the mass spectrometry measurements, we cannot detect an acetate–adapted phenotype comprising less than 10% of the
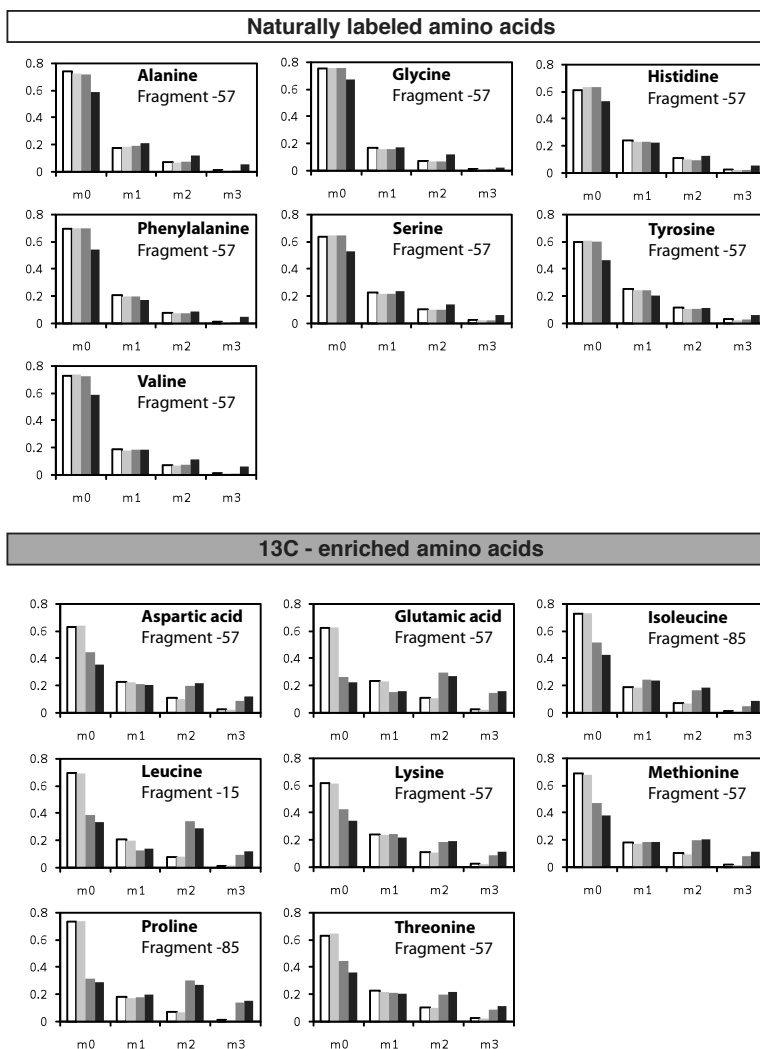
Figure 4.5: Mass distribution vectors (MDV's) of the amino acids. The index $i = 0, 1, 2, 3$ of a MDV indicates the number of the isotope's ^{13}C atoms. White columns, theoretical values of the natural labeling pattern. Light gray columns, negative controls. Dark gray columns, actual samples. Black columns, positive controls. Seven amino acids exhibit a natural labeling pattern — their MDV's are close to the theoretical values and negative controls. Eight amino acids exhibit a ^{13}C–enriched labeling pattern — their MDV's are close to the positive control.

Figure 4.6: Mapping of the amino acids to their respective precursor metabolites in central carbon metabolism. White boxes, naturally labeled compounds. Gray boxes, ^{13}C–enriched compounds. Black box, fully ^{13}C–labeled compound. Boxes with a radial grayscale gradient, compounds with unknown isotopomer distribution. A ^{13}C–enrichment occurs only in those amino acids (gray boxes) that are derived from metabolites below PEP. A natural labeling pattern is exhibited by those amino acids (white boxes) that are derived from metabolites above PEP.

total population. Hence, we cannot exclude that few cells switch their pheno-type stochastically. However, as the acetate–adapted population fraction after the substrate switch to acetate is markedly above 10% for acetate concentrations of at least 0.75 g/l (see Figure 4.2d), the larger fraction of these acetate–adapted cells must be generated by active response to the substrate switch; this active response, among other adjustments, reverses the direction of carbon flow through the Embden–Meyerhoff pathway. Therefore, the *dominant* adaptation strategy must be *responsive diversification* following the substrate switch from glucose to acetate.

4.5.2 General nature of responsive diversification

The central location of the bistability–generating circuit depicted in Figure 4.3a implies that phenotypic subpopulations should be generated not only upon substrate shifts from glucose to acetate but also to other gluconeogenic substrates as well. Therefore, substrate shift experiments were conducted from glucose to 0.75 g/l of either succinate, fumarate, or malate. From these experiments, the fraction of cells adapting to gluconeogenic growth, α, and their growth rates μ_g were determined. The results, plotted in Figure 4.7, reveal that responsive diversification into two phenotypic subpopulations occurs after substrate shifts to succinate, fumarate, and malate. Because responsive diversification is not acetate–specific, responsive diversification is of a general nature, occurring upon substrate shifts from glucose to gluconeogenic substrates.
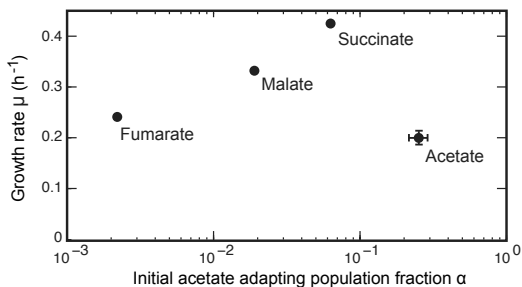
Figure 4.7: Experimentally derived values for the 'growing' subpopulation fraction α and its growth rate μ_g following substrate shifts from glucose to 0.75 g/l of either of the gluconeogenic substrates acetate, succinate, fumarate, or malate. Responsive diversification occurs in all cases and is thus of a general nature.

4.6 Methods Summary

4.6.1 Experimental procedures

This section contains a summary of the experimental methods used in this chapter. Most of these experiments were performed by Benjamin Volkmer. For more details on these methods, refer to the dissertation of Benjamin Volkmer.

Bacterial strains and plasmids   The *Escherichia coli* K–12 BW25113 strain is designated as wild–type throughout this chapter. Single knockout mutant strains were taken from the Keio collection [10]. Cra was overexpressed from the plasmid pCra$^+$, which carries the *cra* gene behind an IPTG–inducible promoter and is derived from pNT3, which was introduced in [179].

Growth media and cultivation protocol   Standard M9 medium was used, supplemented with either 5 g/l glucose, 0.5 g/l, 0.75 g/l, 1 g/l, 2 g/l or 4 g/l acetate, 0.75 g/l succinate, 0.75 g/l malate, or 0.75 g/l fumarate as the sole carbon source. The media were adjusted to pH 7. A single colony was picked from an LB plate, inoculated into a shake flask with 50 ml glucose–M9 medium, grown at 37 °C until an OD of 0.5, and subcultured once. Cells were harvested during mid–exponential growth at an OD of 0.5, washed twice with M9 medium to remove excess glucose, and used to inoculate another carbon source–M9 medium shake flask. All experiments were performed in triplicate, except experiments for which no error bars are shown, which were performed once.

Membrane staining and flow cytometry   In some experiments, cells were stained with a red–fluorescent, membrane-intercalating dye (PKH26, Sigma-Aldrich) before transfer into a new carbon source. 3 ml of culture was taken at an OD of 0.5 and centrifuged for 4 min at 4 °C at 4000 rpm in an Eppendorf 5804 R centrifuge. The supernatant was removed and the cells resuspended in 500 μl dilution buffer C. A freshly prepared mixture of 8 μl PKH26 staining and 500 μl dilution buffer C was added. After 3 min at room temperature, 4 ml of cold, carbon source–free M9 minimal media containing 1% BSA was added. The cells were washed twice through centrifugation and resuspension first in 5 ml cold, carbon source–free media, then in 1 ml cold, carbon–source free media. For flow cytometry, 1 ml of culture was taken at different time points and analyzed using a BD FACS Calibur flow cytometer.

^{13}C cultivation and mass spectrometry   Cells were grown in glucose–M9 medium and subcultured once, then grown in M9 medium supplemented with 5 g/l glucose and 1 g/l either naturally or fully ^{13}C–labeled acetate. Taken samples were fast–centrifuged and shock–frozen in liquid nitrogen. Proteinogenic amino acids were extracted in 60:40 EtOH:H$_2$O solution with 10 mM NH$_4$$^+CH_3COO^-$ at 78 °C, dried in a Kühner RVC 2-33 SpeedVac at 0.12mbar, and derivatized. The mass

isotopomers were analyzed using an Agilent 6890N gas chromatograph coupled to an Agilent 5973 mass spectrometer.

Measurements of protein abundance   The population–averaged Cra abundances were measured with selected–reaction monitoring (SRM) as described in [108, 154].

Physiological measurements   Cells were taken from the second consecutive glucose–M9 medium culture. Extracellular metabolites were quantified with a high–performance liquid chromatography (HPLC) system (Agilent HP1100), equipped with a polymer column (Aminex HPX-87H from Bio-Rad), as described in [70].

4.6.2 Bigaussian model fit to experimental data

Throughout this chapter, the presented values for the population fractions adapting to gluconeogenic growth, α, and their growth rates μ_g are determined by fitting a Bigaussian model (with α and μ_g as parameters) to the measured time progressions of the total populations' fluorescence intensity distributions. A representative such fit is shown in Figure 4.2c. The employed fitting strategy does *not* involve independent fits of two Gaussians to the fluorescence intensity distributions at individual time points but *jointly* considers the fluorescence intensity distributions (and the cell count concentrations) measured at multiple time points. Specifically, the determination of α and μ_g requires (i) the cell count concentration and the fluorescence intensity distribution of the total population measured immediately after the substrate shift at $t = t_0$, designated $c(t_0)$ and $\mathcal{D}(t_0)$, respectively, and (ii) a sequence of k cell count concentrations and fluorescence intensity distributions measured at time instants $t = t_i$, $i = 1 \ldots k$ within the observation window $[t_1, t_k]$, designated $c(t_i)$ and $\mathcal{D}(t_i)$ with $i = 1 \ldots k$.

To obtain the $k + 1$ cell count concentrations c and fluorescence intensity distributions $\mathcal{D}$, the raw data provided by flow cytometry measurements was processed by

1. smoothing the fluorescence data with a 3–point moving average filter,

2. discarding the lowest 200 of in total 1024 fluorescence intensity bins due to unacceptably low signal–to–noise ratios,

3. correcting the fluorescence intensities for background noise by subtracting a noise distribution–approximating, decaying exponential from the total signal.

4. correcting the measured cell count concentrations with the culture's evaporation rate of 0.142 ml/h, and

5. scaling the smoothed and noise–corrected fluorescence intensity distributions with the evaporation–corrected cell count concentrations.

The measured, time–dependent, total population's fluorescence intensity distribution data $\mathcal{D}(t)$ is approximated with a model, $\mathcal{M}(t)$. In the model, the time–dependent, total population's fluorescence intensity distributions are given as the sum of the two time–dependent distributions $\mathcal{M}_g(t)$ and $\mathcal{M}_n(t)$ of the 'growing' and 'non–growing' subpopulations, respectively. Hence,

$$\mathcal{M}(t) = \mathcal{M}_g(t) + \mathcal{M}_n(t) \ . \tag{4.1}$$

Both $\mathcal{M}_g(t)$ and $\mathcal{M}_n(t)$ are modeled as Gaussian distributions $\mathcal{N}(m, \sigma)$ with time–dependent mean intensities $m_g(t)$ and $m_n(t)$, respectively, and time–invariant standard deviations σ_g and σ_n, respectively. The areas of the Gaussians grow with the subpopulations' growth rates μ_g and μ_n, respectively, and are for the $\mathcal{M}_g(t_1)$ and $\mathcal{M}_n(t_1)$–distributions given by $c_{g,1}$ and $c_{n,1}$, respectively. Hence, the time progressions of $\mathcal{M}_g(t)$ and $\mathcal{M}_n(t)$ within the observation window $[t_1, t_k]$ are given by

$$\begin{aligned} \mathcal{M}_g(t) &= c_{g,1} \ \mathrm{e}^{\mu_g(t-t_1)} \ \mathcal{N}(m_g(t), \sigma_g)[256\log_{10}(\cdot)] \\ \mathcal{M}_n(t) &= c_{n,1} \ \mathrm{e}^{\mu_n(t-t_1)} \ \mathcal{N}(m_n(t), \sigma_n)[256\log_{10}(\cdot)] \ , \end{aligned} \tag{4.2}$$

with the argument $[256\log_{10}(\cdot)]$ accounting for the exponentially increasing width of the 1024 fluorescence intensity–bins into which the flow cytometry data is discretized. Therefore, $\mathcal{M}_g(t)$ and $\mathcal{M}_n(t)$ are symmetric on a log–scale and skewed on a linear scale.

At $t = t_1$, the mean fluorescence intensities $m_g(t)$ and $m_n(t)$ are given by the parameters $m_{g,1}$ and $m_{n,1}$, respectively. At all times, the mean fluorescence intensities $m_g(t)$ and $m_n(t)$ have two contributions, (i) the natural fluorescence of the cells, F_N, and (ii) the fluorescence conferred by the dye. The natural fluorescence remains constant, whereas the fluorescence conferred by the dye decreases over time with the cells' constant growth rates μ_g and μ_n, respectively, and with the degradation rate of the dye, $\delta = 0.015\mathrm{h}^{-1}$. Hence, the time progressions of the mean fluorescence intensities $m_g(t)$ and $m_n(t)$ within the observation window $[t_1, t_k]$ are given by

$$\begin{aligned} m_g(t) &= F_N + (m_{g,1} - F_N) \, \mathrm{e}^{-(\mu_g+\delta)(t-t_1)} \\ m_n(t) &= F_N + (m_{n,1} - F_N) \, \mathrm{e}^{-(\mu_n+\delta)(t-t_1)} \ . \end{aligned} \tag{4.3}$$

With the growth rate of the non–growing subpopulation experimentally determined as $\mu_n = 0.03\mathrm{h}^{-1}$, the optimal values $\mathbf{p_{opt}}$ of the remaining eight parameters

$$\mathbf{p} = \begin{bmatrix} \mu_g & \sigma_g & \sigma_n & c_{g,1} & c_{n,1} & m_{g,1} & m_{n,1} & F_N \end{bmatrix} \tag{4.4}$$

are determined by cell count–weighted minimization of the sum–of–squares distance between the model (Equations 4.1, 4.2, and 4.3) and the data within the observation window; hence, the optimal values $\mathbf{p_{opt}}$ for the eight parameters $\mathbf{p}$ are determined by

$$\mathbf{p_{opt}} = \min_{\mathbf{p}} \sum_{i=1}^{k} c(t_i)^{-1} \left(\mathcal{M}(t_i) - \mathcal{D}(t_i)\right)^{\mathrm{T}} \left(\mathcal{M}(t_i) - \mathcal{D}(t_i)\right) \ . \tag{4.5}$$

The model fit in the observation window $[t_1, t_k]$ is used to calculate α, the fraction of cells adapting to growth on gluconeogenic substrates upon the substrate shift at t_0. This calculation is based on the definition of α,

$$\alpha = \frac{A_g(t_0)}{A(t_0)} \, , \tag{4.6}$$

with $A_g(t_0)$ the areal contribution of those cells that are about to assume the growing phenotype to $A(t_0)$, the total area of the total population's fluorescence intensity distribution at t_0. Because the area of any $\mathcal{N}(m, \sigma)$–distribution is unity, $A(t_0)$ is given by the measured cell count concentration at t_0; hence,

$$A(t_0) = c(t_0) \, . \tag{4.7}$$

Given $d(t)$, the number of cell divisions of the growing phenotype between t_0 and $t \in [t_1, t_k]$, the area $A_g(t_0)$ can be back–calculated from any area $A_g(t)$ within the observation window $[t_1, t_k]$ via

$$A_g(t_0) = \frac{A_g(t)}{2^{d(t)}} \, . \tag{4.8}$$

Within the observation window $[t_1, t_k]$, $A_g(t)$, the areal contribution of the growing subpopulation to the total population's fluorescence intensity distribution at time t, is given by

$$A_g(t) = c_{g,1} \, e^{\mu_g (t - t_1)} \, . \tag{4.9}$$

The number of cell divisions of the growing phenotype between t_0 and $t \in [t_1, t_k]$, $d(t)$, is inferred from the loss of fluorescence intensity in this time period via

$$d(t) = \log \frac{1}{2} \, \log \frac{m_g(t_0) - F_N}{m_g(t) - F_N} \, , \tag{4.10}$$

with $m_g(t_0)$ the mean fluorescence intensity of the growing phenotype at t_0. Because t_0 occurs immediately after the cells are stained, at t_0, the mean fluorescence intensity of the growing subpopulation, $m_g(t_0)$, equals m_0, the mean fluorescence intensity of the total population; hence,

$$m_g(t_0) = m_0 \, . \tag{4.11}$$

The mean fluorescence intensity of the total population at t_0, m_0, is determined through the minimization of the sum–of–squares distance between $\mathcal{D}(t_0)$ and the single Gaussian model

$$\mathcal{M}(t_0) = c(t_0) \cdot \mathcal{N}(m_0, \sigma_0)[256 \log_{10}(\cdot)] \, , \tag{4.12}$$

with σ_0 the standard deviation of the total population's fluorescence intensity distribution at t_0.

The beauty of this approach is that the determined values for α and μ_g are *independent* of the population's growth behavior between the shift to the gluconeogenic substrate at t_0 and the start of the observation window at t_1. This property is important because the quantification of the growth behavior within the early time period $[t_0, t_1]$ is highly uncertain due to, for instance, possible reductive cell divisions of the 'non–growing' phenotype and the early, gradual growth rate–acceleration of the 'growing' phenotype (see dissertation of Benjamin Volkmer).

4.6.3 Model of the bistability–generating circuit

To investigate if the identified regulatory circuit (see Figure 4.3a) can generate the experimentally observed subpopulations, we described the molecular interactions with a mathematical model (see Section 4.7 for details), resulting in:

$$\frac{dE_1}{dt} = f(E_1, FBP_{SS}) = v_{e_1,max} \left(1 - \frac{FBP_{SS}^{n_{e_1}}}{FBP_{SS}^{n_{e_1}} + (K_{e1,FBP})^{n_{e_1}}} \right) - d \cdot E_1 \quad (4.13)$$

$$FBP_{SS}(J): \quad 0 = J - \frac{v_{Fbp,max} \frac{FBP_{SS}}{K_{Fbp,FBP}} \left(1 + \frac{FBP_{SS}}{K_{Fbp,FBP}} \right)^3}{\left(1 + \frac{FBP_{SS}}{K_{Fbp,FBP}} \right)^4 + L_{Fbp} \left(1 + \frac{J \cdot K_{E_2,PEP}}{K_{Fbp,PEP} \cdot (v_{E_2,max} - J)} \right)^{-4}} \quad (4.14)$$

$$J(E_1, acetate) = \frac{k_{E_1,cat} \cdot E_1 \cdot acetate}{acetate + K_{E_1,acetate}} \quad (4.15)$$

Equation 4.13 combines the regulation of E_1 production through Cra activity and the regulation of Cra activity through FBP. $v_{e_1,max}$, $K_{e1,FBP}$ and n_{e_1} parameterise the Hill–type kinetics of E_1 production, and d represents the combined E_1 degradation and dilution rate due to cell growth. Equation 4.14 yields the steady state concentration of FBP as an implicit function of the metabolic flux J (see Figure 4.8a). $v_{E_2,max}$ and $K_{E_2,PEP}$ parameterise the Michaelis–Menten kinetics of E_2, and $v_{Fbp,max}$, $K_{Fbp,FBP}$, $K_{Fbp,PEP}$ and L_{Fbp} parameterise the Monod–Wyman–Changeux kinetics of the tetrameric enzyme Fbp. Equation 4.15 gives J established through the Michaelis–Menten kinetics of E_1, parameterised by $k_{E_1,cat}$ and $K_{E_1,acetate}$.

We found that the model structure tightly constrains the possible responses. This enabled us to focus on qualitative predictions without knowing the exact values of the poorly identifiable parameters. Of the parameter sets that lead to biologically meaningful behaviour (stable E_1 dynamics, $E_1 \geq 0$, $PEP_{SS} \geq 0$, $FBP_{SS} \geq 0$, $J \geq 0$), some create a monostable and others a bistable E_1 concentration. The parameters leading to bistability are realistic enough for the biological

system to potentially operate within this regime, and the model behavior reasonably robust to variations of the model structure (see Section 4.7.2 for a detailed analysis).

4.7 Model details

In this section, we present the development and analysis of the mathematical model of the bistability–generating circuit shown in Figure 4.3a. First, we describe in detail the derivation of the model's differential and algebraic equations. After that, we present the bifurcation analysis of the model underlying Figures 4.3b and 4.3c. Finally, we show that the inclusion of additional known molecular interactions only modulates, but does not fundamentally change, the circuit's bifurcation properties.

4.7.1 Model development

We here derive the differential and algebraic equations that model the regulatory architecture shown in Figure 4.3a.

A 'super–enzyme' E_1 subsumes the entire pathway from acetate uptake to phosphoenolpyruvate (PEP) and is modeled with a Michaelis-Menten kinetic,

$$v_{E_1}(acetate, E_1) = \frac{k_{E_1,cat} \cdot E_1 \cdot acetate}{acetate + K_{E_1,acetate}}, \tag{4.16}$$

where $acetate$ is the extracellular acetate concentration, $k_{E_1,cat}$ is E_1's maximal turnover rate, and $K_{E_1,acetate}$ is the Michaelis–Menten constant.

In *E. coli*, the production of many enzymes in this pathway (i.e., AceA, AceB, AcnA, PckA, PpsA) is activated by the transcription factor Cra [177]. Fructose-1,6-biphosphate (FBP) inactivates Cra by binding to it, forming Cra-FBP [177]. The inhibition that FBP thus exerts on E_1 production involves binding events — FBP to Cra, and Cra to the promoter region of the e_1 gene — both of which may or may not involve cooperativity. The effect of FBP on e_1 expression is lumped into the Hill-type function

$$f_1(FBP) = v_{e_1,max} \left(1 - \frac{FBP^{n_{e_1}}}{FBP^{n_{e_1}} + K_{e1,FBP}^{n_{e_1}}} \right), \tag{4.17}$$

where $v_{e_1,max}$ is the maximal e_1 expression rate, n_{e_1} is the degree of cooperativity, and $K_{e1,FBP}$ is the FBP concentration required for 50% expression. For $n_{e_1} = 1$, the lumped binding processes do not contain cooperative binding events so that FBP exerts a parabolic influence on E_1 production. For $n_{e_1} > 1$, these processes contain positively cooperative binding events, such that FBP exerts a sigmoidal influence on e_1 expression.

With E_1 degradation and dilution due to cell growth modelled as a linear function of E_1 itself with slope d, the differential equation for E_1 becomes

$$\frac{dE_1}{dt} = f_1(FBP) - d \cdot E_1. \tag{4.18}$$

The dilution of E_1 due to cell growth depends on the growth rate μ, therefore d increases with μ. Also, the concentrations of ribosome and DNA polymerase complexes, which affect the rate of E_1 synthesis, f_1, increase with μ. We assumed that these contributions cancel each other out and did not model the dependencies of d and f_1 on μ. Although this simplification may introduce an error, it is still more desirable than the alternative, which would be the introduction of a high degree of error–prone structural uncertainty and complexity in an attempt to reduce the original error. Note that the dependency of f_1 on ribosomes and DNA polymerases levels, and the dependencies of these levels on μ, are unclear. Further, μ itself would have to be modelled as a dependent variable. The structure of this function remains unclear, and previous attempts to identify it have failed even in much more detailed kinetic models of metabolism [24].

The enzyme E_2 subsumes the entire gluconeogenic pathway from PEP to FBP and is also modeled with a Michaelis-Menten kinetic,

$$v_{E_2}(PEP) = \frac{v_{E_2,max} \cdot PEP}{PEP + K_{E_2,PEP}}, \tag{4.19}$$

where $v_{E_2,max}$ is the maximal turnover rate, and $K_{E_2,PEP}$ is the Michaelis–Menten constant.

Fbp represents the tetrameric enzyme fructose-1,6-bisphosphatase (Fbp), which converts FBP to fructose-6-phosphate. As PEP is a strong allosteric activator of Fbp [75], Fbp is modelled with a Monod-Wyman-Changeux (MWC) kinetic with $n = 4$:

$$v_{Fbp}(PEP, FBP) = \frac{v_{Fbp,max} \frac{FBP}{K_{Fbp,FBP}} \left(1 + \frac{FBP}{K_{Fbp,FBP}}\right)^3}{\left(1 + \frac{FBP}{K_{Fbp,FBP}}\right)^4 + L_{Fbp}\left(1 + \frac{PEP}{K_{Fbp,PEP}}\right)^{-4}}, \tag{4.20}$$

where $v_{Fbp,max}$ is the maximal turnover rate, and L_{Fbp}, $K_{Fbp,PEP}$ and $K_{Fbp,FBP}$ are shape parameters of the MWC kinetic.

4.7.2 Bifurcation analysis

In this section, we perform a bifurcation analysis on the model equations derived in the previous section. This analysis results in Figures 4.3b and 4.3c.

The steady state assumption on the fluxes

$$v_{E_1} = v_{E_2} = v_{Fbp} = J \tag{4.21}$$

allows the calculation of the steady state metabolite concentrations as function of J. With Equation 4.21, Equation 4.19 is rearranged to give PEP_{SS}, the steady state concentration of PEP:

$$PEP_{SS}(J) = \frac{J \cdot K_{E_2,PEP}}{v_{E_2,max} - J}. \tag{4.22}$$

Then, Equations 4.20, 4.21 and 4.22 are combined to yield an implicit equation for FBP_{SS}, the steady state concentration of FBP:

$$FBP_{SS}(J): \quad 0 = J - \frac{v_{Fbp,max}\frac{FBP_{SS}}{K_{Fbp,FBP}}\left(1+\frac{FBP_{SS}}{K_{Fbp,FBP}}\right)^3}{\left(1+\frac{FBP_{SS}}{K_{Fbp,FBP}}\right)^4 + L_{Fbp}\left(1+\frac{J\cdot K_{E_2,PEP}}{K_{Fbp,PEP}\cdot(v_{E_2,max}-J)}\right)^{-4}}.$$
(4.23)

Through empirically probing the parameter space, we found that the model structure tightly constrains the possible responses. This enabled us to focus on qualitative predictions without knowing the exact values of the poorly identifiable parameters. Of the possible system responses, many are biologically not meaningful. For instance, if E_1 realizes the flux J and $v_{E_2,max} < J$, then E_2 cannot realize J, and Equation 4.22 has no positive steady state solution for $PEP_{SS}(J)$ because the metabolite accumulates indefinitely. By 'educated trial and error', we found parameter sets that lead to a stable E_1 dynamics with $E_1 > 0$, positive steady state metabolite concentrations ($FBP_{SS}(J), PEP_{SS}(J) > 0$), and an uptake of acetate ($J > 0$). These constraints resemble the system's biological function.

Of the parameter sets that lead to such biologically meaningful behaviour, we were specifically interested in those that enable two stable steady states, because a bistable acetate uptake rate can be interpreted as the experimentally observed 'growing' and 'non–growing' phenotypes. We found that such parameter sets exist, although it was harder to find these sets than it was to find sets that produce only one stable steady state. This suggests that the bistable region in parameter space is smaller compared to the monostable region.

However, the bistable region is not unrealistically small for the system to operate within this regime: First, we were able to find the bistable region within the 12–dimensional parameter space using only empiric means. Second, parameter values of a parameter set within the bistable region can be varied to a comfortable extent without leaving the bistable region. Third, parameter values that lead to bistable behaviour seem to be realistic. For instance, the parameters determining a substrate or effector concentration required for 50% activation are in the range of the respective compound concentrations. Further, we found that a high value of L_{Fbp} is essential to move the system to the bistable region. L_{Fbp} is a shape parameter of the MWC kinetics describing the tetrameric enzyme Fbp, which is substantially activated by PEP. Although a value for L_{Fbp} is not available in literature, another study [24] experimentally fitted the shape parameter L of the MWC kinetics describing the tetrameric enzyme $PykF$, which is substantially activated by FBP, and obtained a similarly high value. This provides confidence that high values of L are realistic for tetrameric enzymes that are substantially activated by a metabolite.

In the following, we report on our investigation of the system's operation in the bistable region. With parameters according to Table 4.1, Figure 4.8a shows the steady state metabolite concentrations (Equations 4.22 and 4.23) with varying steady state flux J. Note that PEP_{SS} monotonically increases with increasing
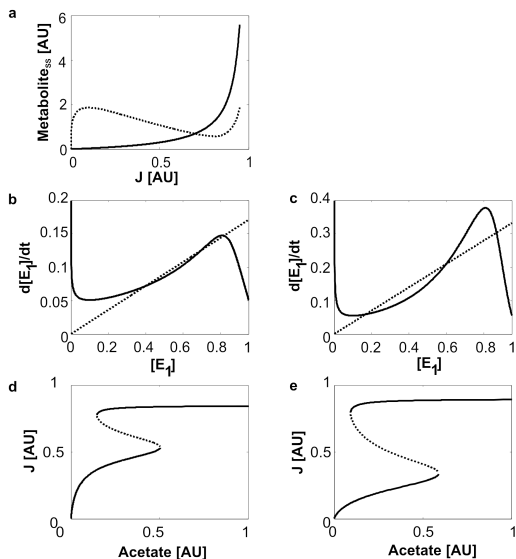
J, whereas FBP_{SS} initially sharply increases, then gradually decreases, and finally increases again. The shape of the FBP_{SS} curve can be explained as follows. For small metabolic fluxes J, the PEP_{SS} concentration is too low to noticeably activate the enzyme Fbp. Therefore, Fbp can only provide the flux J if its substrate concentration FBP_{SS} is relatively high. When J is increased further, then the enzyme E_2 can only provide a higher flux if it increases its substrate concentration PEP_{SS}. This increase in PEP_{SS} significantly activates Fbp so that this enzyme can provide the increased flux J with a relatively low substrate concentration FBP_{SS}. The activation effect of PEP_{SS} on Fbp however saturates with increasing PEP_{SS}. Therefore, if J and thereby PEP_{SS} are raised even further, then Fbp can only provide the flux J if the concentration of its substrate FBP_{SS} is increased again.

It is important to note that the thus established inverse dependence of FBP_{SS} on J within a wide range of J changes the sign of the closed feedback loop on E_1 production from negative to positive. As a single positive feedback is capable of generating two stable steady states, but a single negative feedback loop is not, this change of sign is essential for the generation of two subpopulations.

Table 4.1: Parameters used in the simulations. *This notation means that the parameter was increased from 0.7 to 1.3, in steps of 0.1.

Parameter	Value in Fig. 4.8a	Value in Fig. 4.8b(c)	Value in Fig. 4.8d(e)	Value in Fig. 4.3b	Value in Fig. 4.3c
$k_{E_1,cat}$	-	1	1	1	1
$K_{E_1,acetate}$	-	-	0.02 (0.1)	0.1	0.1
$v_{E_2,max}$	1	1	1	1	1
$K_{E_2,PEP}$	0.3	0.3	0.3	0.3	0.3
$v_{Fbp,max}$	1	1	1	1	1
$K_{Fbp,PEP}$	0.1	0.1	0.1	0.1	0.1
$K_{Fbp,FBP}$	0.1	0.1	0.1	0.1	0.1
L_{Fbp}	$4 \cdot 10^6$	$4 \cdot 10^6$	$4 \cdot 10^6$	$4 \cdot 10^6$	$4 \cdot 10^6$
d	-	0.18 (0.25)	0.18 (0.25)	0.35	0.35
$v_{e_1,max}$	-	1.1	1.1	1	0.7:0.1:1.3*
n_{e_1}	-	1 (2)	1 (2)	2	2
$K_{e_1,FBP}$	-	0.1 (0.45)	0.1 (0.45)	0.45	0.45

In a next step, we investigated how such a flux–dependent, S–shaped concentration profile of the metabolite FBP controls the expression of the enzyme E_1. With the parameters listed in Table 4.1, Figure 4.8b reveals that the E_1 production and degradation/dilution curves (the two terms in Equation 4.18) intersect at three points for both parabolic and sigmoidal regulation of E_1 production. Of these three steady states, the middle one is unstable whereas the outer two are stable. Therefore, the system is capable of expressing bistable E_1 concentrations.

Figure 4.8: a. Steady state PEP (solid line) and FBP (dotted line) metabolite concentrations over steady state flux, which are identical for both parabolic ($n_{e_1} = 1$) and sigmoidal ($n_{e_1} = 2$) influence of FBP on E_1 production. **b,c.** E_1 production (solid line) and degradation/dilution (dotted line) curves for parabolic and sigmoidal regulation of E_1 production by FBP, respectively. These two figures were generated assuming saturation of the enzyme E_1 with its substrate acetate ($acetate \gg K_{E_1,acetate}$), such that $v_{E_1} = k_{E_1,cat} \cdot E_1$. **d,e.** Bifurcation diagrams showing stable (solid lines) and unstable (dotted lines) steady states for parabolic and sigmoidal regulation of E_1 production by FBP, respectively.

The S–shaped form of the E_1 production curve stems mainly from the S–shape of the FBP concentration curve (Figure 4.8a), which in turn is the result of cooperative allosteric activation of Fbp through PEP, as described previously. Hence, allosteric regulation of enzyme activity can be sufficient to generate bistability in gene expression if the allosterically introduced effect of cooperativity on the metabolic level is propagated to the gene expression level via transcription factor–metabolite binding. Introduction of additional cooperativity in the process of propagating the FBP concentration to promoter activity pronounces the characteristic shape of the E_1 production curve (Figure 4.8c), thereby making it 'easier' for the production and degradation/dilution curves to have three intersections. Thus, additional cooperativity widens the region of bistability.

Figures 4.8d and e show the steady state fluxes J through the pathway as a function of the *acetate* concentration in the medium. Qualitatively, the same curve shapes are observed for both the parabolic and sigmoidal case, with a wider region of bistability for the latter. For low *acetate* concentrations, only one stable steady state flux exists at a low level. When the *acetate* concentration is increased, then, at a critical concentration, a second stable steady state flux at a high level is created along with an unstable steady state flux through a saddle-node bifurcation. Further increase of the *acetate* concentration through the region of bistability can lead to a second saddle-node bifurcation where the unstable and the lower stable steady state fluxes collide and annihilate each other. For *acetate* concentrations above this bifurcation point, only one stable steady state flux on a high level would remain. However, although extracellular *acetate* concentrations can reach arbitrarily high levels, the capacity to take up *acetate* is limited due to substrate saturation of the enzyme E_1. This saturation can cause the second saddle–node bifurcation to be unreachable, such that two stable steady state fluxes persist for arbitrarily high *acetate* concentrations. Such behaviour resembles our experimentally observed bistable phenotypes, which we observed for arbitrarily high *acetate* concentrations. These conditions lead to the bifurcation diagrams shown in Figures 4.3b and 4.3c.

4.7.3 Inclusion of further interactions into the model

As we have seen, the presented and analyzed core model, depicted in Figure 4.3a, is capable of reproducing the bistable phenotypes even at high acetate concentrations. However, the core model includes only a subset of all known regulatory interactions between the modelled components. On the metabolic level, all known enzymatic regulations *between the modelled components* are accounted for with the activation of Fbp by PEP. On the genetic level, two known regulatory actions of Cra on modelled components were not included in the core model: The activation of Fbp production, and the repression of E_2 production.

In the following sections, we show that inclusion of these additional molecular interactions only modulates, but not fundamentally changes, the core model's characteristic behaviour. Thereby we show that our mathematical results are not

due to the system boundary of the core model, but due to the dynamics of the known molecular interactions.

4.7.3.1 Activation of Fbp production by Cra

In *E. coli*, the transcription factor Cra, whose activity is modulated through FBP, also activates the production of fructose-1,6-bisphosphatase I (Fbp). In our model, this regulation corresponds to a repression of Fbp production by FBP. To examine the influence of this additional regulation on the system behaviour, we included this additional interaction and then analyzed the extended model.

The extended model equations include the algebraic reaction rate equations for the enzymes E_1 and E_2 (Equations 4.16 and 4.19) as well as the differential equation for the concentration of E_1 (Equation 4.18). A second differential equation for the concentration of the enzyme Fbp needs to be introduced in analogy to Equation 4.18:

$$\frac{dFbp}{dt} = f_{Fbp}(FBP) - d \cdot Fbp, \tag{4.24}$$

with

$$f_{fbp}(FBP) = v_{fbp,max} \left(1 - \frac{FBP^{n_{fbp}}}{FBP^{n_{fbp}} + (K_{fbp,FBP})^{n_{fbp}}} \right), \tag{4.25}$$

where $v_{fbp,max}$ is the maximal production rate of the gene fbp, n_{fbp} is the degree of cooperativity involved in the activation of the gene fbp by FBP, and $K_{fbp,FBP}$ is a shape parameter. Further, the earlier formulated algebraic equation for the reaction rate of the enzyme Fbp (Equation 4.20) needs to be adapted to account for a dynamic Fbp concentration:

$$v_{Fbp}(PEP, FBP) = \frac{k_{Fbp,cat} \cdot Fbp \cdot \frac{FBP}{K_{Fbp,FBP}} \left(1 + \frac{FBP}{K_{Fbp,FBP}} \right)^3}{\left(1 + \frac{FBP}{K_{Fbp,FBP}} \right)^4 + L_{Fbp} \left(1 + \frac{PEP}{K_{Fbp,PEP}} \right)^{-4}}, \tag{4.26}$$

where $k_{Fbp,cat}$ is the enzyme's maximal turnover rate.

To analyze the extended model, we again make the steady state assumption according to Equation 4.21. Then, PEP_{SS}, the steady state concentration of PEP, is given by Equation 4.22. To obtain an implicit equation for the steady state concentration of the metabolite FBP (FBP_{SS}), the steady state concentration of the enzyme Fbp,

$$Fbp_{SS} = \frac{f_{Fbp}(FBP)}{d}, \tag{4.27}$$

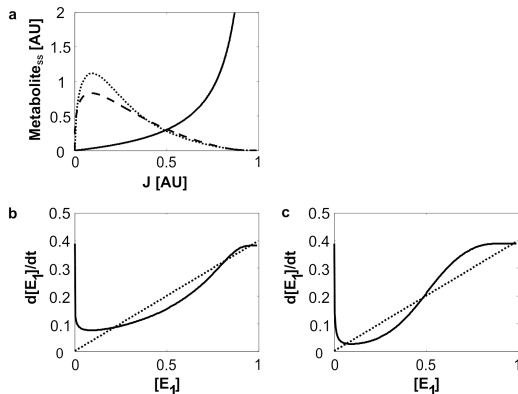is plugged into Equation 4.26, considering Equations 4.21 and 4.22:

$$FBP_{SS}(J): \quad 0 = J - \frac{k_{Fbp,cat} \cdot \frac{f_{Fbp}(FBP)}{d} \cdot \frac{FBP_{SS}}{K_{Fbp,FBP}} \left(1 + \frac{FBP_{SS}}{K_{Fbp,FBP}} \right)^3}{\left(1 + \frac{FBP_{SS}}{K_{Fbp,FBP}} \right)^4 + L_{Fbp} \left(1 + \frac{J \cdot K_{E_2,PEP}}{K_{Fbp,PEP} \cdot (v_{E_2,max} - J)} \right)^{-4}}. \tag{4.28}$$

Table 4.2: Parameter values used in the simulations.

Parameter	Value in Fig. 4.9a parabolic (sigmoidal) case	Value in Fig. 4.9b (c)
$k_{E_1,cat}$	-	1
$K_{E_1,acetate}$	-	-
$v_{E_2,max}$	1	1
$K_{E_2,PEP}$	0.3	0.3
$k_{Fbp,cat}$	20	20
$K_{Fbp,PEP}$	0.1	0.1
$K_{Fbp,FBP}$	0.1	0.1
L_{Fbp}	$1 \cdot 10^6$	$1 \cdot 10^6$
d	0.4	0.4
$v_{e_1,max}$	-	0.39
n_{e_1}	-	1 (2)
$K_{e_1,FBP}$	-	0.2 (0.3)
$v_{fbp,max}$	0.5	0.5
n_{fbp}	1 (2)	1 (2)
$K_{fbp,FBP}$	0.2 (0.3)	0.2 (0.3)

With model parameters according to Table 4.2, Figure 4.9a shows the steady state metabolite concentrations when the steady state flux J is varied. Compare Figure 4.8a and Figure 4.9a. The PEP_{SS} curve remains unchanged. The FBP_{SS} curve, however, does change: At high fluxes J, FBP_{SS} rises in the core model, but not in the here considered extended model. This missing rise of FBP_{SS} in the extended model can be easily explained: When FBP_{SS} decreases with increasing J, then the production of the enzyme Fbp gets activated. Because of the resulting increase in the concentration of Fbp, there is no need to also increase the concentration of its substrate FBP in order to realize a high flux J. This is in contrast to the core model, where the concentration of the enzyme Fbp is constant, so that the only way for Fbp to realize a flux J close to the enzyme's maximal turnover rate is to increase the concentration of its substrate FBP.

Does the observed change in the concentration profile of FBP_{SS} only modulate, or does it fundamentally change, the model's bifurcation properties? As can be seen from a comparison of Figure 4.8b with Figure 4.9b and Figure 4.8c with 4.9c, the profiles of the E_1 production curves differ between the core model and the here considered extended model. These differences arise because the different profiles of the FBP_{SS} curves propagate into the E_1 production curves. In particular, in the extended model, the flattening of the FBP_{SS} curve at high fluxes J propagates into a leveling off of the E_1 production curve at high E_1 concentrations, whereas in the core model, the rise of the FBP_{SS} curve at high fluxes J propagates into a drop of the E_1 production curve at high E_1 concentrations.

Figure 4.9: a. Steady state PEP (solid line) and FBP (dashed line: parabolic, dotted line: sigmoidal regulation of E_1 production) metabolite concentrations over steady state flux. **b,c.** E_1 production (solid line) and degradation/dilution (dotted line) curves for parabolic and sigmoidal regulation of E_1 production, respectively. These two figures were generated assuming saturation of the enzyme E_1 with its substrate acetate ($acetate \gg K_{E_1,acetate}$), such that $v_{E_1} = k_{E_1,cat} \cdot E_1$.

The important point is that regardless of whether the E_1 production curve drops or levels off at high E_1 concentrations, the resulting production curve profile can still intersect with the linear degradation/dilution curve at three points, and thereby establish two stable phenotypes. Thus, inclusion of the activation of *Fbp* production by Cra does not fundamentally change the core model's bifurcation properties. However, inclusion of this additional regulation does modulate the characteristic bifurcation properties of the core model. Since the E_1 production curve levels off (as opposed to drop) at high E_1 concentrations while the degradation/dilution curve remains unchanged, the third intersection point, and thereby the high steady state flux, is inevitably shifted towards higher E_1 concentrations. Therefore, the effect of the activation of *Fbp* production by Cra is a further separation of the stable steady states from each other. Thus, this interaction only modulates (i.e. strengthens) the expression of the bistable phenotypes and is not part of the generating mechanism.

4.7.3.2 Repression of E_2 production by Cra

In *E. coli*, the gluconeogenic pathway from PEP to FBP comprises five reactions. Of the five catalyzing enzymes, the first (which follows PEP consumption) and the last three (which precede FBP formation) are repressed by Cra. These additional regulations are not included in the core model.

To include these additional regulations in the model, the model must be ex-

tended through inclusion of Cra repression on the production rate of E_2, which would be modeled as activation of E_2 production by FBP. Instead of extending the model and performing a subsequent model analysis, we argue that the result of such a model analysis can be foreseen by reasoning and is somewhat analogous to the result obtained by analyzing the additional activation of *Fbp* production by Cra (see Section 4.7.3.1).

It is important to note here that *provided a given flux J*, the concentration of a metabolite is governed solely by the kinetic equation of its consuming enzyme (including possible enzyme and effector concentrations that appear therein). Hence, provided an influx J to PEP, initially only the PEP concentration is affected through introduction of E_2 repression by Cra. Next, the PEP concentration affects the rates of both the FBP–producing (where PEP acts as substrate) and FBP– consuming (where PEP acts as activator) reactions; hence these two reactions compete for PEP. As E_2 production is *repressed* by Cra, this additional interaction modulates the competition of the FBP–producing and FBP–consuming reactions for PEP *in favor for the FBP–consuming reaction*: The PEP concentration rises because with its activity repressed, E_2 can only match the given flux J through increasing the concentration of its substrate. As PEP activates the FBP–consuming reaction, a higher PEP concentration leads to a lower FBP concentration (the FBP–consuming enzyme can match the given flux J with a lower concentration of its substrate). A lower FBP concentration in turn reduces the FBP–mediated repression of E_2 and thus counters the here described effect; hence the process approaches a steady state. However, compared to the core model, the *added* repression of E_2 *must* lead to higher PEP and lower FBP steady state concentrations. Therefore, the inverse dependency of FBP on PEP, which has been shown in Section 4.7.2 to be essential for the generation of multiple steady states, is *strengthened* through inclusion of this additional regulation.

Therefore, the effect of the repression of E_2 production by Cra is a further separation of the stable steady states from each other. Thus, this interaction only modulates (i.e. strengthens) the expression of the bistable phenotypes and is not part of the generating mechanism.

4.8 Acknowledgements

Chapter 5

On the optimality of responsive phenotypic diversification

Oliver Kotte, Martin Ackermann and Matthias Heinemann

5.1 Summary

Isogenic populations of bacteria are known to diversify into phenotypic subpopulations in order to anticipate future environmental changes, an adaptation strategy known as bet–hedging. The generation of phenotypic diversity through the mechanism of stochastic switching has been shown to maximize population size in certain stochastic environments. However, the optimality of another adaptation mechanism, responsive diversification, has not yet been investigated. Therefore, in this study, we investigate the optimality of responsive phenotypic diversification in stochastic environments where the nutrient concentrations are subject not only to external fluctuations but also to the growth behavior of the cell population. By means of a mathematical model, we show that in certain such environments,

population growth is *not* maximized by maximization of each cell's growth rate but by responsive diversification of phenotypes. We show that in such environments, bet–hedging 'resolves' an *apparent* tragedy of the commons dilemma and has therefore a social component attached. We connect the bet–hedging strategy to two molecular trade–offs imposing limitations on the behavior of individual cells. The trade–off between self–preservation and nutritional capability has been introduced earlier, the trade–off between fast growth and nutritional flexibility is conceptually novel and introduced in this study. We argue that the genotype deals with these trade–offs on the population level through the expression of phenotypic subpopulations that specialize in either of the trade–offs' extremes. Remarkably, responsive diversification of phenotypes upon carbon source fluctuations can *simultaneously* deal with the two molecular trade–offs on the population level, maximize population size in the absence of stresses, *and* increase the robustness of the total population to stresses.

5.2 Introduction

To anticipate environmental changes, isogenic bacterial populations are known to diversify into multiple distinct and coexisting phenotypic subpopulations. Examples of such phenotypic multistability, which is generated through propagation of stochastic events during gene expression (for instance, see [50, 100, 146, 168]), include the swarming motility [89], the sporulation [212] and the competence for DNA uptake from the environment [202] of *Bacillus subtilis*, the acquisition of antibiotic resistance [12, 106] and the reversal of central carbon flow (see Chapter 4) in *Escherichia coli*, and flagellin phase variation in *Salmonella enterica* [232]. The employed adaptation strategy is known as bet–hedging [187, 213]: Multiple phenotypes increase the chances that at least some cells will be well–adapted in an uncertain future environment [41]. In recent years, this microbial adaptation strategy has received much interest [7, 201, 229] and has been shown to be an evolvable trait [2].

Intriguingly, the benefits of bet–hedging strategies are surprisingly multifaceted. For instance, the benefits for a sporulating *B. subtilis* subpopulation [214] are resistances to various stresses, whereas the benefits for cells that delay or avoid sporulation are (i) slow proliferation on nutrients released from lysis of other cells, a behavior termed cannibalism or fratricide [39], and (ii) the capability for rapid resumption of growth in the event of returning nutrient conditions (cells that have sporulated are committed to a long-term process of spore formation and subsequent germination) [213]. Additionally, non–growing cells can indirectly benefit actively metabolizing cells by not competing for limited resources; hence, in certain circumstances, selfish bet–hedging has been proposed to have social components attached [61]. Further, if bet–hedging through spontaneous stochastic phenotype switching makes sensing and quick response mechanisms otherwise required to adapt to extracellular conditions obsolete, and the saved costs of these mechanisms

exceed the incurred costs of not having all individuals well–adapted, bet–hedging maximizes the size of the total isogenic cell population [107, 205].

In Chapter 4 of this thesis, it has been shown that following a carbon source shift from glucose to gluconeogenic substrates such as acetate, an isogenic *E. coli* population diversifies into two phenotypic subpopulations. One of these subpopulations grows on the gluconeogenic substrate, whereas the other, surprisingly large subpopulation refrains from growth and, as strongly suggested by the morphology of its cells, enters the protective state of stationary phase rest. It is still unclear *why E. coli* employs the bet–hedging strategy of responsive phenotypic diversification to adapt to this type of environmental change — whereas stochastic phenotype switching has been shown to maximize population size in certain stochastic environments [2, 107, 205], the optimality of responsive phenotypic diversification has not yet been investigated.

Hence, this study investigates the multifaceted benefits of bet–hedging through responsive phenotypic diversification following shifts from rich (glycolytic) to limited concentrations of poor (gluconeogenic) carbon sources. By means of a mathematical model, we show that for certain such substrate shifts, (i) population growth is counter–intuitively *not* maximized by maximization of each cell's growth rate but by responsive phenotypic diversification, and that in such environments, (ii) bet–hedging has an attached social component by 'resolving' an *apparent* tragedy of the commons dilemma [67, 123, 163]. We connect the population–level study to the molecular level and argue that the two distinct phenotypes, which are a form of specialization [213], are a strategy of the genotype to cope with limitations of individual cells imposed by two molecular trade–offs. One of these trade–offs has been proposed earlier [94], the other rests on the results of Chapter 4 and is conceptually novel.

This chapter is structured as follows. First, we present a mathematical model that describes the growth of an isogenic cell population in and its impact on a stochastic environment. Then, we analyze the model and discuss the identified social component of the bet–hedging strategy. Next, we review one and introduce another molecular trade–off and argue that the genotype deals with these trade–offs on the population level through phenotypic specialization in either of the trade–offs' extremes. Finally, we combine these results into a coherent explanation for the optimality of responsive diversification upon carbon source fluctuations.

5.3 Model

To analyze if following a substrate shift from a rich to a poor carbon source, the bet–hedging strategy of responsive phenotypic diversification has attached social components, we developed a differential equation model.

First, we give an overview of the model. Then, we describe the model in detail and present the model equations.
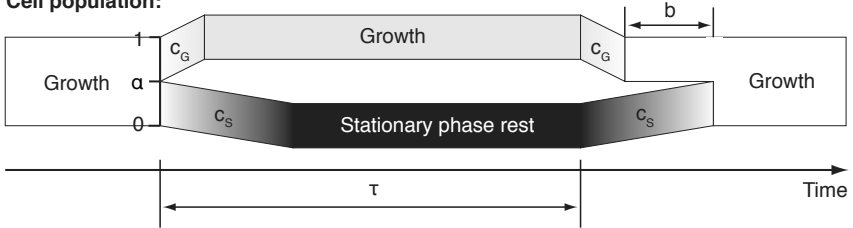
5.3.1 Overview

The model describes the growth of an isogenic cell population in an environment with two carbon sources. One of these carbon sources is richer than the other, so the growth rate on the rich substrate, μ_R, is greater than the growth rate on the poor substrate, μ_P. It is not possible for the cells to grow on both substrates at the same time (this occurs, for instance, when the rich carbon source is glycolytic and the poor carbon source gluconeogenic [178]). The cells preferentially adapt to the rich substrate, meaning that whenever the rich carbon source is present, the poor substrate is ignored (this is realized, for instance, through the mechanism of catabolite repression). Presence of the rich carbon source resembles *feast conditions*. When the rich substrate is absent, the cells can either grow on the poor substrate if it is available, or they can enter the protective state of stationary phase rest [73]. Such an environment resembles *famine conditions*.

Intuitively, one would expect that upon depletion of the rich substrate, all cells adapt to growth on the poor substrate until (i) the rich substrate returns to the environment, causing immediate re–adaptation to the rich substrate, or (ii) the poor substrate becomes depleted, in which case no substrates are available and the cells enter stationary phase rest. However, in the case of an isogenic *E. coli* population, glucose as the rich substrate ($\mu_R = 0.64\text{h}^{-1}$), and acetate as the poor substrate ($\mu_A = 0.20\text{h}^{-1}$), a recent experimental observation (see Chapter 4) contradicts this intuitive reasoning: When glucose becomes depleted but acetate is still available, only a phenotypic subpopulation adapts to growth on acetate. The other subpopulation refrains from growth and, as strongly suggested by the morphology of its cells, enters the state of stationary phase rest.

We set up a differential equation model to investigate the effect of such counterintuitive behavior on long–term growth of an isogenic cell population in an environment with fluctuating carbon sources. The model subjects an isogenic cell population to a period of famine conditions between two periods of feast conditions to compare the adaptation strategies of responsive switching (all cells respond in the same way to changes in carbon source availability) and responsive diversification (different cells respond in different ways to changes in carbon source availability) with each other. The comparison is made by means of a cost–benefit analysis yielding the optimal adaptation strategy that maximizes the size of the total population. The model includes subpopulation dynamics during famine conditions, consumption of the poor substrate by competitors as well as transition times and costs between the different phenotypes.

5.3.2 Model equations

The differential equation model describes how an initially homogeneous population, which is still adapted to the rich carbon source, splits into two phenotypic subpopulations upon the transition from feast to famine conditions (see Figure 5.1). Certain cells directly enter the state of stationary phase rest, other cells adapt to

Environmental conditions:

Feast	Famine	Feast

Cell population:

Figure 5.1: Illustration of the cost–benefit model analysis. The model confronts a cell population with a single period of famine conditions, lasting for the duration τ, between two periods of feast conditions. By varying α between 0 (all cells enter stationary phase rest) and 1 (all cells grow on the poor carbon source), the model allows for a comparison between the adaptation strategies of responsive switching ($\alpha = 1$) and responsive diversification ($0 < \alpha < 1$). b, growth benefit of metabolically active cells due to shorter adaptation time. c_S, transition costs incurred when adapting between growth on different substrates. c_S, transition costs incurred when entering or leaving stationary phase rest.

the poor substrate and grow until (i) the rich substrate returns, or (ii) the poor substrate becomes depleted, whereupon these cells join the stationary subpopulation. To determine the adaptation strategy that maximizes the size of the total population after completed re–adaptation to feast conditions, this final population size is maximized through variation of the initial subpopulation proportions. Of these processes,

1. the adaptation from feast to famine conditions is described by a function of the phenotypic split ratio, yielding the initial conditions of the state variables,

2. the population behavior during famine conditions is described by differential equations,

3. the adaptation from famine back to feast conditions and the calculation of final population size is described by a function of the differential equations' terminal state values.

Refer to Table 5.1 for an overview of the model's nomenclature.

The first process described by the model is the transition from feast to famine conditions. Upon depletion of the rich substrate at time $t = 0$, a homogeneous population of size P_0, which is still adapted to the rich carbon source, splits into two phenotypic subpopulations, P_S and P_P. The subpopulation P_S directly enters stationary phase rest, incurring transition costs c_S. The subpopulation P_P adapts to growth on the poor substrate, incurring transition costs c_G. The ratio of this

phenotypic split is quantified by the parameter α, which describes the fraction of the P_0–population that 'decides' to adapt to the poor substrate; thus, $0 \leq \alpha \leq 1$. The concentration of the poor substrate, S_P, is at $t = 0$ given by $S_{P,0}$. The transition from feast to famine conditions is modeled through the calculation of the initial conditions of the model's three state variables P_S, P_P and S_P, according to

$$
\begin{aligned}
P_S(0) &= P_0 \, c_S \, (1 - \alpha) \\
P_P(0) &= P_0 \, c_G \, \alpha \\
S_P(0) &= S_{P,0} \, .
\end{aligned}
\tag{5.1}
$$

The second process described by the model covers the subpopulation dynamics during the period of famine conditions, which lasts for τ hours. During this period, the subpopulation P_S, in stationary phase rest, grows with a nonpositive rate of μ_S. The other subpopulation, P_P, grows on the poor substrate with the rate μ_A, a function of the substrate concentration S_P. When the poor substrate becomes depleted ($S_P = 0$), cells of the subpopulation P_P enter stationary phase rest and thus join the subpopulation P_S; the transition from the P_P to the P_S subpopulation is modeled with the S_P-dependent rate ρ and incurs transition costs c_S. Growth of the subpopulation P_P reduces the concentration of the poor substrate S_P by converting S_P to biomass with yield Y. The poor substrate S_P is further drained from the environment with rate δ, reflecting the consumption of S_P by competitors. The subpopulation dynamics during the period of famine conditions ($t \in [0, \tau]$) are given by

$$
\begin{aligned}
\dot{P}_S &= \mu_S \, P_S + \rho \, P_P \, c_S \\
\dot{P}_P &= (\mu_P - \rho) \, P_P \\
\dot{S}_P &= -Y^{-1} \, \mu_P \, P_P \, .
\end{aligned}
\tag{5.2}
$$

The growth rate μ_P is given by

$$
\mu_P = \frac{\mu_{P,max} \, S_P}{S_P + K_P},
\tag{5.3}
$$

with $\mu_{P,max}$ the maximal growth rate and K_P the Monod constant for growth of P_P on the poor substrate S_P.

The S_P-dependent switching rate ρ of cells from the P_P to the P_S subpopulation is given by

$$
\rho = \rho_{max} \, (\mu_{P,max} - \mu_P),
\tag{5.4}
$$

with ρ_{max} the maximal switching rate.

The third and final process covered by the model is the transition from famine back to feast conditions. This transition is triggered by the return of the rich substrate to the environment at time $t = \tau$. Both subpopulations immediately adapt to growth on the rich substrate, incurring transition costs of c_S for P_S cells, and of c_G for P_P cells. This transition is described by a function of the model's terminal states, i.e. the values of the state variables at time $t = \tau$. This function,

$$
P_F = P_P(\tau) \, c_G \, b + P_S(\tau) \, c_S \, ,
\tag{5.5}
$$

Table 5.1: Nomenclature of the model.

Symbol	Classification	Meaning
b	parameter	relative growth benefit
c_G	parameter	transition cost of switching between growth on either substrate
c_S	parameter	transition cost of entering or leaving stationary phase rest
K_P	parameter	Monod constant for the poor substrate
P_P	variable	size of the (sub)population adapted to the poor substrate
P_0	parameter	initial size of the (total) population, adapted to the rich substrate
P_F	function	final size of the (total) population, adapted to the rich substrate
P_S	variable	size of the (sub)population in stationary phase rest
S_P	variable	concentration of the poor substrate
$S_{P,0}$	parameter	initial concentration of the poor substrate
Y	parameter	yield coefficient
α	parameter	population fraction adapting to growth on the poor carbon source
δ	parameter	fierceness of competition
μ_P	function	growth rate on the poor substrate
$\mu_{P,max}$	parameter	maximal growth rate on the poor substrate
μ_S	parameter	growth rate in stationary phase rest
ρ	function	switching rate to stationary phase rest
ρ_{max}	parameter	maximal switching rate to stationary phase rest
τ	parameter	duration of famine conditions

calculates P_F, the size of the total, homogeneous population after completed re-adaptation of cells from both subpopulations to growth on the rich substrate. In addition to the terminal states and transition costs, P_F also depends on the parameter b. This parameter quantifies the relative growth benefit of cells adapting from the metabolically active (former) subpopulation, P_P, compared to cells adapting from the subpopulation in stationary phase rest, P_S.

Cells adapting from the P_P–subpopulation harvest this relative growth benefit because they adapt to the rich substrate faster than cells adapting from the P_S–subpopulation (see Figure 5.1). The faster adaptation of P_P–cells arises from the fact that these cells are already metabolically active whereas the P_S–cells need to slowly undo the complex physiological and morphological changes they built up as they entered stationary phase rest [73]. Therefore, when the former P_S-cells have completed their adaptation to the rich substrate, the former P_P cells were already growing on the rich substrate with the fast rate μ_G for a certain period of time. As Equation 5.5 calculates the total population size after *completed* re-adaptation of both subpopulations to the rich substrate, which occurs a certain period of time *after* the return of the rich substrate to the environment at $t = \tau$, the relative growth benefit b due to faster adaptation of formerly P_P–cells must be accounted for. Note that in the context of the sporulation of *B. subtilis*, the capability for rapid resumption of growth in the event of returning nutrient conditions has already been proposed as a benefit of actively metabolizing cells over spores [213].

Three assumptions underlie the presented model. First, entering and leaving stationary phase rest is assumed to be energetically more expensive than adapting from one carbon source to another; thus, $c_S > c_G$. Second, cells growing on the poor carbon source are assumed to adapt to growth on the rich carbon source faster than cells adapting from stationary phase rest; thus, $b \geq 0$. Third, the error made by not considering the adaptation time from feast to famine conditions is negligible. This error is probably indeed negligible because (i) the result of our ^{13}C–experiment presented in Chapter 4 strongly suggests that the adaptation from the rich substrate glucose to the poor substrate acetate is very quick, and (ii) the non–positive 'growth rate' of cells in stationary phase rest, μ_S, is probably only slightly below zero and therefore, the error incurred from too early 'growth' of stationary phase–cells on the poor carbon source is probably very small (if μ_S is zero, no error is made at all).

5.4 Results

5.4.1 Maximization of population size in the presence of stresses

Previous studies [2, 107, 205] have shown that in certain stochastic environments, multiple specialized phenotypic subpopulations can maximize the long–term growth rate of an isogenic cell population. For these results to hold for the here studied case, each of the here investigated two phenotypes must have a higher

growth rate than the other in one of two available environments. As this study analyses the optimal strategy to bridge periods of famine conditions characterized by the absence of the rich carbon source, in the context of this section, the two available environments are A: the poor carbon source is available, and B: the poor carbon source is not available.

In environment A, the growth rate of the growing phenotype is obviously higher than the growth rate of the non–growing phenotype. But what happens in environment B? The morphology of *E. coli* cells expressing the non–growing phenotype during famine conditions strongly suggests that these cells have entered the state of stationary phase rest. Although not considered differentiated, *E. coli* cells in stationary phase rest have many properties in common with the spores into which some gram–positive bacteria such as *B. subtilis* differentiate [212]. For instance, both *B. subtilis* spores and *E. coli* cells in stationary phase rest have a strong multiple–stress resistance and can survive prolonged periods of starvation [73].

Therefore, *if environment B poses environmental threats to the cells*, then the stationary, 'non–growing' phenotype has a higher (less negative) growth rate than the not stress–resistant phenotype adapted to the (not present) poor carbon source. For this case, the above mentioned studies [2, 107, 205] have shown that multiple phenotypes can maximize population size by saving the costs for the sensing and quick response mechanisms otherwise required to counter environmental threats.

5.4.2 Maximization of population size in the absence of stresses

Can responsive phenotypic diversification maximize population size in environments where the non–growing, stationary phenotype does not harvest any benefits from growth restraint? This question is addressed by the analysis of the model presented in Section 5.3 — environmental stresses during famine conditions do not occur in this model.

5.4.2.1 Optimal adaptation to specific carbon source fluctuations

With the presented model, we determined the optimal strategy (in terms of the maximization of total population size) for a cell population to bridge single periods of famine conditions between two periods of feast conditions. Mathematically, the optimal adaptation strategy α_{opt} is determined by maximizing P_F, the final population size after completed re-adaptation to feast conditions, over α, the ratio of the initial phenotypic split upon the transition from feast to famine conditions. Therefore, a function of the terminal states (Equation 5.5) is maximized by variation of the state's initial conditions. Therefore, the task

$$\alpha_{opt} \ = \ \max_{\alpha} P_F \tag{5.6}$$

is a small–scale dynamic optimization problem, which we solved with a brute force approach.

To investigate how the optimal adaptation strategy depends on the characteristics of the famine condition to be bridged, we solved the optimization problem for many different periods of famine conditions, which we generated by varying

- the duration of the famine conditions, τ,
- the initial concentration of the poor substrate, $S_{P,0}$, and
- the fierceness of competition, δ.

Table 5.2 lists the parameter values used in the optimization. As this study investigates a general effect and specific units only introduce a scaling factor that can be countered by inverse scaling of the parameter values, all parameters units were kept arbitrary.

Table 5.2: Used parameter values. All units are arbitrary. $\mathcal{N}(m,\sigma)$ denotes the normal distribution with mean m and standard deviation σ.

Parameter	Used value or distribution
b	4
c_G	0.9999
c_S	0.66
K_P	0.02
P_0	1
$S_{P,0}$	In Section 5.4.2.1 (Figure 5.2): varies from 0 to 15
	In Section 5.4.2.2 (Figure 5.3): $\mathcal{N}(7,3)$–distribution
Y	0.4
α	subject to optimization
δ	In Section 5.4.2.1 (Figure 5.2): 0.1
	In Section 5.4.2.2 (Figure 5.3): varies from 0 to 2
$\mu_{P,max}$	0.24
μ_S	0
ρ_{max}	4
τ	In Section 5.4.2.1 (Figure 5.2): varies from 0 to 20
	In Section 5.4.2.2 (Figure 5.3): $\mathcal{N}(9,3)$–distribution

The dependency of the optimal phenotypic split ratio, α_{opt}, on the duration of famine conditions, τ, and the initial concentration of the poor carbon source, $S_{P,0}$, is shown in Figure 5.2. This dependency establishes three distinct regions in the τ-$S_{P,0}$ plane.

- In Region 1, $S_{P,0}$ is so high that there is more poor substrate available than the cell population and its competitors can consume before the rich substrate returns. Therefore, the population size is maximized by a uniform adaptation to growth on the poor substrate (responsive switching, $\alpha_{opt} = 1$).

- In Region 2, $S_{P,0}$ is low enough to become a limited resource. By restricting growth to only a subpopulation, the total population stretches the available substrate over time. It thereby maintains an actively metabolizing subpopulation until the rich substrate returns. The actively metabolizing subpopulation saves costly transitions into and out of stationary phase rest, and benefits from a faster and cheaper adaptation to feast conditions. Therefore, the optimal adaptation strategy is responsive diversification, $0 < \alpha < 1$.

- In Region 3, τ is so high that maintaining an actively metabolizing subpopulation becomes impossible in the presence of defecting competitors growing on the common, poor substrate as fast as possible. Therefore, the optimal strategy is an aggressive competition for the common substrate through a uniform adaptation to growth in famine conditions (responsive switching, $\alpha_{opt} = 1$). The size of Region 3 depends on the fierceness of competitors for the common substrate, δ. Region 3 is nonexistent for $\delta = 0$; with increasing δ, Region 3 gradually takes over Region 2 until Region 2 becomes extinct and responsive switching is globally optimal.
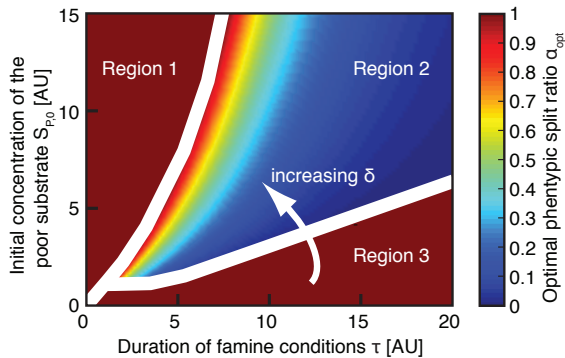
Figure 5.2: The optimal population response α_{opt} to a single carbon source fluctuation depends on the duration of the famine condition, τ, the initial concentration of the poor carbon source $S_{P,0}$, and the fierceness of competition, δ. The dependency on these three parameters subdivides the τ–$S_{P,0}$ plane into three regions. A uniform adaptation to growth on the poor substrate (responsive switching; red) is optimal in Regions 1 and 3, whereas a diversification response (responsive diversification; other colors) is optimal in Region 2. Region 3 is nonexistent for $\delta = 0$; with increasing δ, Region 3 gradually takes over Region 2 until Region 2 becomes extinct and responsive switching is globally optimal.

To conclude, the optimal strategy to bridge a single period of famine conditions between two periods of feast conditions depends on the duration of the famine condition, the available amount of the poor carbon source, and the fierceness of competitors for the common substrate. Depending on these parameters, the

optimal strategy is either responsive switching or responsive diversification into two subpopulations.

5.4.2.2 Optimal adaptation in a stochastic environment

Obviously, the cell population cannot anticipate the optimal adaptation strategy α_{opt} to bridge a particular, upcoming period of famine conditions. However, cells live in an environment that repeatedly alternates between feast and famine conditions. Therefore, the cell population can respond to each period of famine conditions with the constant, evolutionary optimal phenotypic split ratio $\alpha_{opt}^\star$ that maximizes the *long–term* population size in an environment with alternating feast and famine conditions. An alternating sequence of feast and famine conditions can be described by *distributions* over τ, $S_{P,0}$ and δ.

To determine the optimal adaptation strategy in such stochastic environments, we imposed normal distributions over τ and S_{P_0} (see Table 5.2) and set δ to a constant value. We then generated many such environments that share the same distributions over τ and S_{P_0} but differ in the constant value of δ. Therefore, cells in each such environment are faced with an infinite sequence of feast and famine conditions, with different, single periods of famine conditions having different durations and substrate concentrations but the same fierceness of competition. Figure 5.3 shows that the optimal adaptation strategy in such stochastic environments depends strongly on the constant value for the fierceness of competition, δ. The weaker the competition is, the smaller is the optimal size of the growing subpopulation on the poor carbon source, $\alpha_{opt}^\star$, and the larger is the accomplished gain in population size of the optimal response over the uniform response. With increasing competition, the optimal response $\alpha_{opt}^\star$ first approaches and then becomes the uniform response.

The exact value of the constant, evolutionary optimal adaptation strategy $\alpha_{opt}^\star$ depends strongly on the choice of the distributions over τ, $S_{P,0}$ and δ. In essence, responsive diversification ($0 < \alpha_{opt}^\star < 1$) maximizes the long–term population size if the environment comprises sufficient diversification–favoring periods of famine conditions (which lie in Region 2 of Figure 5.2). A diversification strategy with a constant, evolutionary optimal $\alpha_{opt}^\star$ can only be optimal if the competition for the poor carbon source is sufficiently weak. The reason for this is that the prudent resource usage of the diversification strategy only pays off if the population's growth behavior influences the environmental substrate concentration to a sufficient degree. With increasing fierceness of competitors, the scope of the population's control over the environmental substrate pool decreases continuously, and thus responsive diversification becomes ever less beneficial to the total population in terms of maximizing its size.

To conclude, responsive diversification can increase the size of the total population not only by saving the costs of sensing and quick response mechanisms (see Section 5.4.1), but *additionally* by utilizing the poor carbon source prudently. It is important to understand that these two contributions do not exclude each other
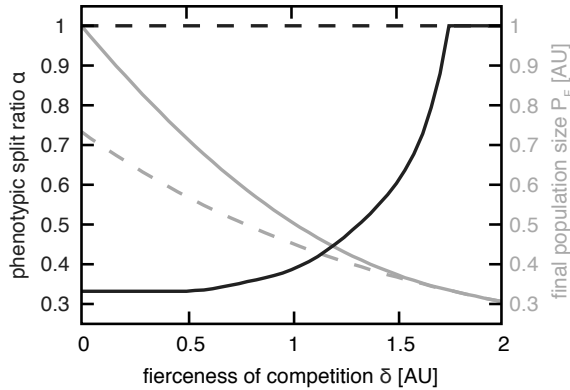
Figure 5.3: In stochastic environments characterized by normal distributions over τ and $S_{P,0}$, the constant, evolutionary optimal phenotypic split ratio $\alpha^{\star}_{opt}$ strongly depends on the constant fierceness of competitors, δ. The weaker the competition is, the smaller is the optimal phenotypic split ratio $\alpha^{\star}_{opt}$ (solid black line). The optimal response $\alpha^{\star}_{opt}$ becomes the homogeneous response ($\alpha = 1$, dashed black line) for sufficiently fierce competition δ. When the fierceness of competition is weak enough, the optimal response yields a higher population size (solid gray line) than the uniform response (dashed gray line).

— sensing and response cost savings increase population size for a certain range of environmental switching rates [107], whereas prudent resource usage increases population size if the famine conditions fall in Region 2 of Figure 5.2.

5.4.2.3 Social component of responsive diversification

Having understood in detail how responsive phenotypic diversification can maximize the size of the total population in certain stochastic environments, in this section, we zoom out and assume a game–theory perspective. This perspective is suited to understand the optimality of responsive diversification in generic terms and allows for a classification of the seemingly cooperative behavior of the phenotypic subpopulations suggested by the model analysis. A game in the context of game theory means any situation in which an individual's success in making choices depends on the choices of others. In the here studied scenario of responsive diversification, individual cell of the population play a game because maximization of each cell's individual growth does not maximize the growth of the total population. More specifically, responsive diversification in the here studied context is a strategy of a so–called social goods game, or N–person prisoners dilemma game (for example, see [68]).

Such a game is played whenever individuals have access to a common resource that can be exploited *either* rapidly *or* efficiently [112, 123]. This dilemma was

first noted by [115], is since [67] known as the tragedy of the commons, and has already been identified in biological systems (for instance, see [90]). The here present tragedy of the commons dilemma can be classified as a Type 1 tragedy, which arises from the competition of individuals over an extrinsic resource [163], in our case over the poor carbon source. The dilemma can be further classified as a collapsing tragedy, which arises when selfish behavior of individuals, in our case a uniform population response in a diversification–favoring environment, results in the entire resource vanishing [163].

It has been convincingly argued [67] that unless the access to the common resource is regulated, selfish behavior of individuals depletes the resource rapidly and inefficiently, with potentially disastrous consequences for the population as a whole. For instance, when human society is confronted with a collapsing Type 1 tragedy of the commons dilemma, which occurs e.g. when overfishing threatens to drive the fish population extinct, the tragedy can only be averted through the imposition of fishing laws. As the tragedy can only be avoided through *higher–level* regulation [67], and natural selection acts at the level of the gene [46], the here present tragedy can only be resolved if such higher–level regulation is provided by the genotype of the isogenic cell population in the form of *cooperative behavior*. In the model, cells of the non–growing subpopulation P_S even act *altruistically* in order to benefit the cells of the growing subpopulation P_P. The cooperative behavior is also reflected in the fact that cooperation can be easily exploited by cheating defectors (see for instance [215]), and the optimality of the here investigated diversification response is lost in the presence of sufficiently fierce competitors (see Figure 5.3).

To conclude, *in the absence of environmental stresses*, the expression of two phenotypic subpopulations during famine conditions maximizes the genotype's representation in future generations by resolving a collapsing Type 1 tragedy of the commons dilemma *cooperatively*. This cooperative behavior contrasts the by definition selfish nature of bet–hedging (where each cell places its own individual bet on the uncertain future) and implies that in friendly environments, responsive diversification has an attached social component. However, *in the presence of environmental stresses*, the increase in population size achieved by prudent resource usage is not altruistic because the non–growing, stationary phenotype harvests the benefit of stress resistance from growth restraint and as such acts primarily selfish (it has merely placed a different individual bet on future conditions). In these cases, the tragedy of the commons dilemma is only an apparent dilemma (which are quite common in biology [163]).

5.4.3 Population–level solution to molecular–level trade–offs

In this section, we review one and introduce another molecular trade–off that impose limitations on the behavior of individual cells. We argue the genotype deal with these limitations of individual cells on the population level by expressing phenotypic subpopulations that specialize in the trade–offs' opposing extremes.

5.4.3.1 Self–preservation or nutritional capability

We have argued in Section 5.4.1 that the increased stress resistance of the non–growing subpopulation confers a benefit to growth restraint. In this context, a conceptual trade–off between self–preservation and nutritional capability has been observed [94]. This trade–off is connected to the concentration of the sigma factor σ^S (or RpoS) [196], which is responsible for the general stress response of *E. coli* [72, 117, 223]. The trade–off is manifested in the fact that strains with high σ^S concentrations are more resistant to stress but grow more slowly and on fewer nutrients under a variety of conditions [54].

The existence of this trade–off means that an individual cell cannot specialize in both self–preservation and nutritional capability. However, the genotype of an isogenic cell population may deal with this limitation on the population level by expressing phenotypic subpopulations that specialize in the trade–off's opposing extremes.

5.4.3.2 Fast growth or nutritional flexibility

In this section, we propose the conceptually novel trade–off between fast growth on rich glycolytic carbon sources and nutritional flexibility, meaning the capability to adapt to poor gluconeogenic carbon sources after depletion of glycolytic carbon sources.

The here proposed novel trade–off between nutritional flexibility and fast growth is predicted by the results presented in Chapter 4 of this thesis. In that chapter, we identified the molecular mechanism responsible for the generation of two phenotypic subpopulations upon glucose–to–acetate substrate shifts in *E. coli* (see Figure 4.3a). We showed that the magnitude of the gluconeogenic flux realized immediately after glucose depletion must be sufficiently high in order to adapt from the glycolytic substrate glucose to the gluconeogenic substrate acetate. Because a strong expression of glycolytic genes implies a strong repression of gluconeogenic genes, glycolytic and gluconeogenic growth are mutually exclusive [178]. Therefore, the more a cell expresses its glycolytic genes, the more it represses its gluconeogenic genes, and the more difficult it should be for the cell, upon depletion of glycolytic substrates, to achieve an immediate gluconeogenic flux above the watershed required for gluconeogenic growth (see Figure 4.3b). Therefore, the results of Chapter 4 predict that the more a cell dedicates itself to growth on glycolytic carbon sources, the more difficult its adaptation to gluconeogenic carbon sources should become.

This model–derived prediction is strongly supported by the results of an evolution experiment in a seasonal environment containing glucose and acetate [110]. The authors of that study observed a genetic diversification of an initially isogenic *E. coli* population into two coexisting strains. One of these strains exhibited a strong expression of genes in the acetate excretion pathway and grew fast on glucose but had severe difficulties to adapt to acetate after glucose depletion; the other

strain exhibited a strong expression of the acetate uptake pathway and grew slower on glucose but was capable to quickly adapt to acetate. Although the authors of that study focus on the genetic diversification and the coexistence of these two strains in the seasonal environment, the growth behavior and gene expression patterns of the two evolved strains strongly suggest the existence of the here proposed trade–off between fast growth and nutritional flexibility.

The existence of this trade–off means that an individual cell cannot specialize in both fast growth and nutritional flexibility. However, the genotype of an isogenic cell population may deal with this limitation on the population level by expressing phenotypic subpopulations that specialize in the trade–off's opposing extremes.

5.5 Discussion

In this chapter, a mathematical model was presented to investigate population growth in stochastic environments with nutrient concentrations subject not only to external fluctuations but also to the growth behavior of an isogenic cell population. From the model analysis, it was concluded that in certain stochastic environments, the expression of phenotypic subpopulations is the evolutionary optimal adaptation strategy maximizing the size of the total population. The counter–intuitive, population size–increasing effect was classified as a solution to a collapsing Type 1 tragedy of the commons dilemma, which reduces to an apparent dilemma in the presence of environmental stresses. It was argued that this effect can be cumulative to the increase in population size achieved by saving the costs for not required sensing and quick response mechanisms.

In a next step, the population–level study was connected to two molecular trade–offs, which impose limitations in individual cells. The trade–off between between nutritional capability and self–preservation has been studied thoroughly and is connected to the RpoS concentration [54, 94]; the trade–off between nutritional flexibility and fast growth is conceptually novel, rests on a recent experimental observation [110] and on the results presented in Chapter 4 of this thesis, and is connected to the competition between glycolytic and gluconeogenic gene expression. We have argued that the genotype deals with these trade–offs on the population level through phenotypic specialization in either of these trade–offs' extremes.

The combined results of this chapter provide a coherent picture for the optimality of responsive phenotypic diversification upon carbon source fluctuations. When taken together, the results of this chapter suggest that although the two phenotypic subpopulations become manifest *after* depletion of the rich carbon source, they may be determined already *well before* the transition from feast to famine conditions through hedging of bets:

1. A cell can 'decide' to refrain from growing on the rich substrate as fast as possible to be able to adapt to growth on the poor substrate later. Hence, the cell bets that the duration of feast conditions will be short, that the amount

of available poor substrate will be large, and that the protective measures provided by the state of stationary phase rest will not be required as long as the poor substrate will be available. With regard to its two molecular trade–offs, the cell bets on nutritional flexibility instead of fast growth on the rich substrate, and on nutritional capability instead of self–preservation on the poor substrate.

2. A cell can also 'decide' to grow on the rich carbon source with such a fast rate that it won't be able to adapt to the poor carbon source later; when the rich substrate will be depleted, the cell enters the state of stationary phase rest instead. Hence, the cell bets that the duration of feast conditions will be long, that the duration of famine conditions will be short, and that the protective measures provided by the state of stationary phase will be required while the poor substrate is available. With regard to its two molecular trade–offs, the cell bets on fast growth instead of nutritional flexibility on the rich substrate, and on self–preservation instead of nutritional capability on the poor substrate.

Therefore, the expression of phenotypic subpopulations that specialize in the extremes — either nutritional capability *and* nutritional flexibility or fast growth on the rich substrate *and* self–preservation on the poor substrate — might be a population–level strategy to cope with *both* molecular trade–offs *jointly*. Note that we do not explicitly show that phenotypic specialization is more optimal to the stress resistance of the genotype than a homogeneous population balanced in between the two extremes. However, because multiple phenotypes generally increase the robustness of the population to stresses [29, 98, 199], phenotypic specialization is very likely the optimal solution to the trade–off between self–preservation and nutritional capability. Next, if this trade–off is best solved through responsive diversification of phenotypes, then the optimality of phenotypic specialization to solve the molecular trade–off between fast growth and nutritional flexibility inevitably follows because there is no reason for the later non–growing subpopulation to *not* grow as fast as possible on the rich carbon source as long as it is available.

To conclude, in this chapter, we have shown that responsive phenotypic diversification upon carbon source fluctuations is a bet–hedging strategy that can *simultaneously*

1. increase the robustness of the total population to stresses because of the mere fact of having multiple phenotypes expressed [29],

2. deal with the molecular trade–off between nutritional capability and self–preservation on the population level through phenotypic specialization in the opposing extremes,

3. deal with the molecular trade–off between nutritional flexibility and fast

> growth on the population level through phenotypic specialization in the op-
> posing extremes, and

4. maximize the representation of the genotype in future generations through
   prudent usage of the limited, poor carbon source during famine conditions.

If stresses are absent from the environment, the employed bet–hedging strategy has
social traits attached. As recent studies have already proposed a cooperative role of
persister cells [61] and spores [213], and this study has identified a cooperative role
of a stationary phase rest—subpopulation, bet–hedging strategies that involve the
expression of a protective but growth–impaired phenotype might thus in general
carry a social component.

5.6 Outlook

In its present form, this chapter is a purely computational study. The scientific con-
tribution of this chapter can be improved by combining the presented theoretical
work with experiments. The following specific experiments are suggested.

As experimental model system linked to the computational study, the glucose
to acetate shift of an isogenic *E. coli* population is suggested. Note that usage of
this model system requires adjusted parameter values and thus a repetition of the
model analysis leading to Figures 5.2 and 5.3. In particular, the Monod constant
for the poor carbon source, K_P, should be set to the literature value for acetate,
which is in the 7 to 10 mM range [230]. A literature study might further reveal
an *E. coli*–specific estimate for the transition costs c_G and c_S. The value b for the
relative growth benefit of metabolically active cells over those in stationary phase
rest could be determined experimentally. To determine this parameter, shake flask
cell cultures growing exponentially on acetate as well as cultures that have been in
stationary phase rest for different amounts of time could be inoculated into fresh
glucose medium. Once all of these cultures have reached exponential growth on
glucose, the biomass concentration of the culture that was metabolically active on
acetate is expected to be greater than the biomass concentrations of the cultures
that were in stationary phase rest. The value of b is given by the difference of
these biomass concentrations. If cultures that have been in stationary phase rest
for different amounts of time exhibit different biomass concentrations on glucose,
then the value of b is not constant but depends on the time the cells have spent
in stationary phase rest; if this is the case, the mathematical model should be
extended to include this dependency.

Using shake flasks, substrate shift experiments from e.g. 5 gl^{-1} glucose to 2
gl^{-1} acetate and back to 5 gl^{-1} glucose can be performed to experimentally vali-
date that growth restraint of a phenotypic subpopulation can maximize the size of
the total population upon completed re–adaptation to glucose. To obtain different
ratios of the phenotypic split upon the glucose to acetate substrate change, two
parallel experiments with two strains exhibiting different phenotypic split ratios α

are suggested. For instance, *E. coli* wild–type could be used to obtain a split–ratio of ≈ 0.5, and the *E. coli maeBsfcA* mutant strain could be used to obtain a split ratio of ≈ 0.017 (see Chapter 4). Then, the duration of growth on acetate is chosen such that the biomass concentration of the still exponentially growing *maeBsfcA* culture, which exhibits a longer apparent lag phase than the wild–type, exceeds the biomass concentration of the wild–type culture, which has already entered stationary phase rest because it has consumed the acetate in its medium. At such a time point, an equal volume of both cultures is transferred to new shake flasks that contain glucose as the carbon source. After both cultures have reached exponential growth, the biomass concentration of the *maeBsfcA* culture (which grows on glucose almost with wild–type rate) should exceed the biomass concentration of the wild–type culture. The important expected result of this experiment is that the expression of *less* acetate–consuming cells in the second shake flask culture increases the size of the total population in the third shake flask culture. This result validates the computational results derived in Section 5.4.2.1.

Further, a variant of the suggested glucose–acetate–glucose substrate shift experiment could be additionally performed. This additional experiment would be identical to the above suggested experiment except that both the wild–type and the *maeBsfcA* culture are inoculated into the *same* shake flask. Therefore, these strains compete for the available acetate. It is expected that the wild–type outgrows the *maeBsfcA* mutant strain because it has more acetate–adapted cells that grow on the common substrate. Ideally, but perhaps not necessarily, the composition of the mixed culture is determined at multiple time points e.g. through insertion of green fluorescent markers in one strain but not in the other and a microscopic analysis of samples. Whereas the above suggested substrate shift experiment is expected to show that the *maeBsfcA* strain performs better than the wild–type in separate cultures, the expected result of the substrate shift experiment with a mixed culture is that the wild–type performs better than the *maeBsfcA* mutant. Because the benefits derived from cooperativity are known to diminish in the presence of competitors, this expected result would corroborate the notion that the two expressed phenotypic subpopulation are *apparently cooperative*.

5.7 Acknowledgements

Chapter 6

Conclusions

To advance our understanding of the general principles surrounding and generating phenotypic diversification of isogenic cell populations, this thesis investigated the emergence of multiple phenotypes in the central carbon metabolism of the bacterium *Escherichia coli*.

The take–home message of this thesis is that phenotypic multistability, which has already been identified in diverse cellular processes, generalizes to central metabolism and can thus be viewed as an inherent feature of its design. Because metabolism plays a vital role in diseases such as cancer, and diversification of such diseases severely hinders their successful treatment, the discovery of phenotypic diversification within central metabolism may have significant impact beyond fundamental research. Most notably, sick cells have been interpreted as being 'trapped' in a disease–conferring stable steady state (or phenotype) of cellular operation that is in a healthy cell suppressed [98]. For this reason, systematic research on how to control cellular states without triggering mutations has been proposed as a crucial component of future cancer–therapy research [97, 99]. The general principles proposed in this thesis further our fundamental understanding of how multiple simultaneously available stable steady states are generated and how individual cells 'choose' between the available phenotypes. This understanding may guide systematic disease research in finding ways to manipulate the expression of multiple phenotypes such that cells trapped in a sick state of cellular operation are forced back to a healthy state.

In detail, the novel general principles proposed in this thesis are

- two motifs that establish *flux–signalling metabolites*, whose levels indicate the presence of extracellular carbon sources (see Chapter 3),

- *molecular sensors for intracellular metabolic flux* (see Chapter 3),

- the sensing of extracellular carbon sources through *distributed sensing of metabolic fluxes* (see Chapter 3),

- the population–level adaptation strategy of *responsive diversification* (see Chapter 4),

- *performance–based selection* of one phenotype over another (see Chapter 4),

- *flux sensor–centered* generation of multiple phenotypes (see Chapter 4), and

- the molecular *trade–off between fast growth and nutritional flexibility* (see Chapter 5).

This thesis has further shown that

- the interplay of known interactions in *E. coli*'s central metabolism can explain in molecular-level detail the system-wide adaptations of metabolic operation between glycolytic and gluconeogenic carbon sources (such as glucose and acetate, see Chapter 3),

- metabolic flux is a control factor that regulates gene expression in vivo (see Chapter 4), a function that has so far only been demonstrated synthetically [58].

- lag time upon environmental perturbations can indicate the existence of phenotypic subpopulations (see Chapter 4), and

- the population–level adaptation strategy underlying responsive diversification is bet–hedging with possible social components that 'resolve' an apparent tragedy of the commons dilemma (see Chapter 5).

The principles proposed in Chapter 4 are experimentally validated whereas those proposed in Chapter 3 stem from a model–based analysis and therefore still requires direct experimental validation. Nevertheless, a flux–sensor that was proposed by the model–based analysis in Chapter 3 has been in Chapter 4 experimentally validated to generate bistable phenotypes based on the measured metabolic flux. Therefore, the experimental results presented in Chapter 4 do provide indirect experimental evidence for the principles proposed in Chapter 3. Similarly, the molecular trade–off proposed in Chapter 5 rests on the experimentally validated function of performance–based selection (see Chapter 4). However, such cross–chapter evidence does not cover the model–derived mechanism of *distributed flux sensing* (the higher–level working together of multiple flux sensors), whose experimental verification is difficult precisely because its nature is distributed. Hence, although distributed flux sensing provides a consistent explanation of how a multitude of known molecular interactions fits into a coherent systems picture, it remains to be seen whether this explanation withstands future observations or whether a refined explanation will be required.

The explanation of distributed flux sensing has been derived through a modeling approach that puts system behavior before parameter values. This approach, when carefully applied as in Chapter 2, can predict new functions that emerge from the modeled complexity. The approach can but in general does not need to be assisted by the method developed in Chapter 2. It will be interesting to see if the here chosen focus on system behavior instead of on parameter values will influence the future modeling of biochemical systems.

References

1. ACAR, M., BECSKEI, A., AND VAN OUDENAARDEN, A. Enhancement of cellular memory by reducing stochastic transitions. *Nature 435*, 7039 (2005), 228–32.

2. ACAR, M., METTETAL, J. T., AND VAN OUDENAARDEN, A. Stochastic switching as a survival strategy in fluctuating environments. *Nat Genet 40*, 4 (2008), 471–5.

3. ANDERSON, W. B., SCHNEIDER, A. B., EMMER, M., PERLMAN, R. L., AND PASTAN, I. Purification of and properties of the cyclic adenosine 3',5'-monophosphate receptor protein which mediates cyclic adenosine 3',5'-monophosphate-dependent gene transcription in *Escherichia coli*. *J Biol Chem 246*, 19 (1971), 5929–5937.

4. ARITA, M. The metabolic world of *Escherichia coli* is not small. *Proc Natl Acad Sci U S A 101*, 6 (2004), 1543–7.

5. ARKIN, A., ROSS, J., AND MCADAMS, H. H. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics 149*, 4 (1998), 1633–48.

6. AUSTIN, D. W., ALLEN, M. S., MCCOLLUM, J. M., DAR, R. D., WILGUS, J. R., SAYLER, G. S., SAMATOVA, N. F., COX, C. D., AND SIMPSON, M. L. Gene network shaping of inherent noise spectra. *Nature 439*, 7076 (2006), 608–11.

7. AVERY, S. V. Microbial cell individuality and the underlying sources of heterogeneity. *Nat Rev Microbiol 4*, 8 (2006), 577–87.

8. AXELROD, R., AXELROD, D. E., AND PIENTA, K. J. Evolution of cooperation among tumor cells. *Proc Natl Acad Sci U S A 103*, 36 (2006), 13474–9.

9. AXELROD, R., AND HAMILTON, W. D. The evolution of cooperation. *Science 211*, 4489 (1981), 1390–6.

10. BABA, T., ARA, T., HASEGAWA, M., TAKAI, Y., OKUMURA, Y., BABA, M., DATSENKO, K. A., TOMITA, M., WANNER, B. L., AND MORI, H. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol 2* (2006), 2006.0008.

References

11. BAHAR, R., HARTMANN, C. H., RODRIGUEZ, K. A., DENNY, A. D., BUSUT-TIL, R. A., DOLLÉ, M. E. T., CALDER, R. B., CHISHOLM, G. B., POLLOCK, B. H., KLEIN, C. A., AND VIJG, J. Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature 441*, 7096 (2006), 1011–4.

12. BALABAN, N. Q., MERRIN, J., CHAIT, R., KOWALIK, L., AND LEIBLER, S. Bacterial persistence as a phenotypic switch. *Science 305*, 5690 (2004), 1622–5.

13. BALÁZSI, G., BARABÁSI, A.-L., AND OLTVAI, Z. N. Topological units of environmental signal processing in the transcriptional regulatory network of *Escherichia coli. Proc Natl Acad Sci U S A 102*, 22 (2005), 7841–6.

14. BAR-EVEN, A., PAULSSON, J., MAHESHRI, N., CARMI, M., O'SHEA, E., PILPEL, Y., AND BARKAI, N. Noise in protein expression scales with natural protein abundance. *Nat Genet 38*, 6 (2006), 636–43.

15. BARABÁSI, A.-L., AND OLTVAI, Z. N. Network biology: understanding the cell's functional organization. *Nat Rev Genet 5*, 2 (2004), 101–13.

16. BECSKEI, A., KAUFMANN, B. B., AND VAN OUDENAARDEN, A. Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nat Genet 37*, 9 (2005), 937–44.

17. BECSKEI, A., SÉRAPHIN, B., AND SERRANO, L. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J 20*, 10 (2001), 2528–35.

18. BECSKEI, A., AND SERRANO, L. Engineering stability in gene networks by autoregulation. *Nature 405*, 6786 (2000), 590–3.

19. BENNETT, B. D., KIMBALL, E. H., GAO, M., OSTERHOUT, R., VAN DIEN, S. J., AND RABINOWITZ, J. D. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli. Nat Chem Biol 5*, 8 (2009), 593–9.

20. BENZER, S. Induced synthesis of enzymes in bacteria analyzed at the cellular level. *Biochimica Et Biophysica Acta 11*, 3 (1953), 383–395.

21. BERG, O. G. A model for the statistical fluctuations of protein numbers in a microbial population. *J Theor Biol 71*, 4 (1978), 587–603.

22. BERG, O. G., PAULSSON, J., AND EHRENBERG, M. Fluctuations and quality of control in biological cells: zero-order ultrasensitivity reinvestigated. *Biophys J 79*, 3 (2000), 1228–36.

23. BERGMANN, S., IHMELS, J., AND BARKAI, N. Similarities and differences in genome-wide expression data of six organisms. *PLoS Biol 2*, 1 (2004), E9.

24. BETTENBROCK, K., FISCHER, S., KREMLING, A., JAHREIS, K., SAUTER, T., AND GILLES, E.-D. A quantitative approach to catabolite repression in *Escherichia coli*. *J Biol Chem 281*, 5 (2006), 2578–84.

25. BETTENBROCK, K., SAUTER, T., JAHREIS, K., KREMLING, A., LENGELER, J. W., AND GILLES, E.-D. Correlation between growth rates, EIIACrr phosphorylation, and intracellular cyclic AMP levels in *Escherichia coli* K-12. *J Bacteriol 189*, 19 (2007), 6891–900.

26. BIJLSMA, J. J. E., AND GROISMAN, E. A. Making informed decisions: regulatory interactions between two-component systems. *Trends Microbiol 11*, 8 (2003), 359–66.

27. BLAKE, W. J., AND COLLINS, J. J. And the noise played on: stochastic gene expression and HIV-1 infection. *Cell 122*, 2 (2005), 147–9.

28. BLAKE, W. J., KAERN, M., CANTOR, C. R., AND COLLINS, J. J. Noise in eukaryotic gene expression. *Nature 422*, 6932 (2003), 633–7.

29. BOOTH, I. R. Stress and the single cell: intrapopulation diversity is a mechanism to ensure survival upon exposure to stress. *Int J Food Microbiol 78*, 1-2 (2002), 19–30.

30. BOTSFORD, J. L., AND HARMAN, J. G. Cyclic AMP in prokaryotes. *Microbiol Rev 56*, 1 (1992), 100–22.

31. BRANDMAN, O., FERRELL, JR, J. E., LI, R., AND MEYER, T. Interlinked fast and slow positive feedback loops drive reliable cell decisions. *Science 310*, 5747 (2005), 496–8.

32. BREHM-STECHER, B. F., AND JOHNSON, E. A. Single-cell microbiology: tools, technologies, and applications. *Microbiol Mol Biol Rev 68*, 3 (2004), 538–59.

33. BREMER, H., AND DENNIS, P. P. Modulation of chemical composition and other parameters of the cell at different exponential growth rates. In *EcoSal — Escherichia coli and Salmonella: Cellular and Molecular Biology*, A. Böck, R. Curtiss III, J. Kaper, P. D. Karp, F. C. Neidhardt, T. Nyström, J. M. Slauch, C. L. Squires, and D. Ussery, Eds. ASM Press, Washington, DC, 2008, ch. 97.

34. BROWN, K. S., HILL, C. C., CALERO, G. A., MYERS, C. R., LEE, K. H., SETHNA, J. P., AND CERIONE, R. A. The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Phys Biol 1*, 3-4 (2004), 184–95.

References

35. CAI, L., FRIEDMAN, N., AND XIE, X. S. Stochastic protein expression in individual cells at the single molecule level. *Nature 440*, 7082 (2006), 358–62.

36. CHASSAGNOLE, C., NOISOMMIT-RIZZI, N., SCHMID, J. W., MAUCH, K., AND REUSS, M. Dynamic modeling of the central carbon metabolism of *Escherichia coli. Biotechnol Bioeng 79*, 1 (2002), 53–73.

37. CHOI, P. J., CAI, L., FRIEDA, K., AND XIE, X. S. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science 322*, 5900 (2008), 442–6.

38. CHUNG, T., RESNIK, E., STUELAND, C., AND LAPORTE, D. C. Relative expression of the products of glyoxylate bypass operon: contributions of transcription and translation. *J Bacteriol 175*, 14 (1993), 4572–5.

39. CLAVERYS, J.-P., AND HÅVARSTEIN, L. S. Cannibalism and fratricide: mechanisms and raisons d'être. *Nat Rev Microbiol 5*, 3 (2007), 219–29.

40. CLUTTON-BROCK, T. H., O'RIAIN, M. J., BROTHERTON, P. N., GAYNOR, D., KANSKY, R., GRIFFIN, A. S., AND MANSER, M. Selfish sentinels in cooperative mammals. *Science 284*, 5420 (1999), 1640–4.

41. COHEN, D. Optimizing reproduction in a randomly varying environment. *J Theor Biol 12*, 1 (1966), 119–29.

42. COHEN, G. N., AND MONOD, J. Bacterial permeases. *Bacteriol Rev 21*, 3 (1957), 169–94.

43. COVERT, M. W., KNIGHT, E. M., REED, J. L., HERRGARD, M. J., AND PALSSON, B. O. Integrating high-throughput and computational data elucidates bacterial networks. *Nature 429*, 6987 (2004), 92–6.

44. CRASNIER-MEDNANSKY, M., PARK, M. C., STUDLEY, W. K., AND SAIER, JR, M. H. Cra-mediated regulation of *Escherichia coli* adenylate cyclase. *Microbiology 143 (Pt 3)* (1997), 785–92.

45. CSETE, M. E., AND DOYLE, J. C. Reverse engineering of biological complexity. *Science 295*, 5560 (2002), 1664–9.

46. DAWKINS, R. *The Selfish Gene*. Oxford University Press, 1976.

47. DIONISIO, F., AND GORDO, I. The tragedy of the commons, the public goods dilemma, and the meaning of rivalry and excludability in evolutionary biology. *Evol. Ecol. Res. 8*, 2 (2006), 321–332.

48. EL MANSI, E. M., NIMMO, H. G., AND HOLMS, W. H. Pyruvate metabolism and the phosphorylation state of isocitrate dehydrogenase in *Escherichia coli. J Gen Microbiol 132*, 3 (1986), 797–806.

49. ELF, J., PAULSSON, J., BERG, O. G., AND EHRENBERG, M. Near-critical phenomena in intracellular metabolite pools. *Biophys J 84*, 1 (2003), 154–70.

50. ELOWITZ, M. B., LEVINE, A. J., SIGGIA, E. D., AND SWAIN, P. S. Stochastic gene expression in a single cell. *Science 297*, 5584 (2002), 1183–6.

51. ENDY, D., KONG, D., AND YIN, J. Intracellular kinetics of a growing virus: A genetically structured simulation for bacteriophage T7. *Biotechnol Bioeng 55*, 2 (1997), 375–89.

52. EPSTEIN, W., ROTHMAN-DENES, L. B., AND HESSE, J. Adenosine 3':5'-cyclic monophosphate as mediator of catabolite repression in *Escherichia coli*. *Proc Natl Acad Sci U S A 72*, 6 (1975), 2300–4.

53. FAMILI, I., FORSTER, J., NIELSEN, J., AND PALSSON, B. O. *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 (2003), 13134–9.

54. FERENCI, T. Maintaining a healthy SPANC balance through regulatory and mutational adaptation. *Mol Microbiol 57*, 1 (2005), 1–8.

55. FÖRSTER, J., FAMILI, I., FU, P., PALSSON, B. Ø., AND NIELSEN, J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res 13*, 2 (2003), 244–53.

56. FOSTER, K., FORTUNATO, A., STRASSMANN, J., AND QUELLER, D. The costs and benefits of being a chimera. *Proceedings of the Royal Society of London Series B-Biological Sciences 269*, 1507 (Nov. 2002), 2357–2362.

57. FUHRER, T., FISCHER, E., AND SAUER, U. Experimental identification and quantification of glucose metabolism in seven bacterial species. *J Bacteriol 187*, 5 (2005), 1581–90.

58. FUNG, E., WONG, W. W., SUEN, J. K., BULTER, T., LEE, S.-G., AND LIAO, J. C. A synthetic gene-metabolic oscillator. *Nature 435*, 7038 (2005), 118–22.

59. FUTREAL, P. A., COIN, L., MARSHALL, M., DOWN, T., HUBBARD, T., WOOSTER, R., RAHMAN, N., AND STRATTON, M. R. A census of human cancer genes. *Nat Rev Cancer 4*, 3 (2004), 177–83.

60. GADKAR, K. G., GUNAWAN, R., AND DOYLE, 3RD, F. J. Iterative approach to model identification of biological networks. *BMC Bioinformatics 6* (2005), 155.

61. GARDNER, A., WEST, S. A., AND GRIFFIN, A. S. Is bacterial persistence a social trait? *PLoS One 2*, 1 (2007), e752.

References

62. GARDNER, T. S., CANTOR, C. R., AND COLLINS, J. J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature 403*, 6767 (2000), 339–42.

63. GOLDBETER, A. Computational approaches to cellular rhythms. *Nature 420*, 6912 (2002), 238–45.

64. GORE, J., YOUK, H., AND VAN OUDENAARDEN, A. Snowdrift game dynamics and facultative cheating in yeast. *Nature 459*, 7244 (2009), 253–6.

65. GOTTESMAN, M. Bacteriophage lambda: the untold story. *J Mol Biol 293*, 2 (1999), 177–80.

66. GUTENKUNST, R. N., WATERFALL, J. J., CASEY, F. P., BROWN, K. S., MYERS, C. R., AND SETHNA, J. P. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol 3*, 10 (2007), 1871–78.

67. HARDIN, G. The tragedy of the commons. *Science 162*, 5364 (1968), 1243–8.

68. HARDIN, R. Collective action as an agreeable n-prisoners' dilemma. *Behavioral Science 16*, 5 (1971), 472–481.

69. HARTWELL, L. H., HOPFIELD, J. J., LEIBLER, S., AND MURRAY, A. W. From molecular to modular cell biology. *Nature 402*, 6761 Suppl (1999), C47–52.

70. HEER, D., AND SAUER, U. Identification of furfural as a key toxin in lignocellulosic hydrolysates and evolution of a tolerant yeast strain. *Microbial Biotechnology 1*, 6 (2008), 497–506.

71. HEINEMANN, M., KÜMMEL, A., RUINATSCHA, R., AND PANKE, S. In silico genome-scale reconstruction and validation of the *Staphylococcus aureus* metabolic network. *Biotechnol Bioeng 92*, 7 (2005), 850–64.

72. HENGGE-ARONIS, R. Signal transduction and regulatory mechanisms involved in control of the sigma(S) (RpoS) subunit of RNA polymerase. *Microbiol Mol Biol Rev 66*, 3 (2002), 373–95, table of contents.

73. HENGGE-ARONIS, R. Regulation of gene expression during entry into stationary phase. In *EcoSal — Escherichia coli and Salmonella: Cellular and Molecular Biology*, A. Böck, R. Curtiss III, J. Kaper, P. D. Karp, F. C. Neidhardt, T. Nyström, J. M. Slauch, C. L. Squires, and D. Ussery, Eds. ASM Press, Washington, DC, 2008, ch. 93.

74. HENGL, S., KREUTZ, C., TIMMER, J., AND MAIWALD, T. Data-based identifiability analysis of non-linear dynamical models. *Bioinformatics 23*, 19 (2007), 2612–8.

75. HINES, J. K., FROMM, H. J., AND HONZATKO, R. B. Novel allosteric activation site in *Escherichia coli* fructose-1,6-bisphosphatase. *J Biol Chem 281*, 27 (2006), 18386–93.

76. HOOSHANGI, S., THIBERGE, S., AND WEISS, R. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc Natl Acad Sci U S A 102*, 10 (2005), 3581–6.

77. HÜBSCHER, J., JANSEN, A., KOTTE, O., SCHÄFER, J., MAJCHERCZYK, P. A., HARRIS, L. G., BIERBAUM, G., HEINEMANN, M., AND BERGER-BÄCHI, B. Living with an imperfect cell wall: compensation of femab inactivation in *Staphylococcus aureus. BMC Genomics 8* (2007), 307.

78. IBARRA, R. U., EDWARDS, J. S., AND PALSSON, B. O. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature 420*, 6912 (2002), 186–9.

79. INGRAM, P. J., STUMPF, M. P. H., AND STARK, J. Network motifs: structure does not determine function. *BMC Genomics 7* (2006), 108.

80. ISHIHAMA, Y., SCHMIDT, T., RAPPSILBER, J., MANN, M., HARTL, F. U., KERNER, M. J., AND FRISHMAN, D. Protein abundance profiling of the *Escherichia coli* cytosol. *BMC Genomics 9* (2008), 102.

81. ISHII, N., NAKAHIGASHI, K., BABA, T., ROBERT, M., SOGA, T., KANAI, A., HIRASAWA, T., NABA, M., HIRAI, K., HOQUE, A., HO, P. Y., KAKAZU, Y., SUGAWARA, K., IGARASHI, S., HARADA, S., MASUDA, T., SUGIYAMA, N., TOGASHI, T., HASEGAWA, M., TAKAI, Y., YUGI, K., ARAKAWA, K., IWATA, N., TOYA, Y., NAKAYAMA, Y., NISHIOKA, T., SHIMIZU, K., MORI, H., AND TOMITA, M. Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science 316*, 5824 (2007), 593–7.

82. JAIN, A., PRABHAKAR, S., AND PANKANTI, S. On the similarity of identical twin fingerprints. *Pattern Recognition 35*, 11 (Nov. 2002), 2653–2663.

83. JANAUSCH, I. G., ZIENTZ, E., TRAN, Q. H., KRÖGER, A., AND UNDEN, G. C4-dicarboxylate carriers and sensors in bacteria. *Biochim Biophys Acta 1553*, 1-2 (2002), 39–56.

84. JANGA, S. C., SALGADO, H., MARTÍNEZ-ANTONIO, A., AND COLLADO-VIDES, J. Coordination logic of the sensing machinery in the transcriptional regulatory network of *Escherichia coli*. *Nucleic Acids Res 35*, 20 (2007), 6963–72.

85. JAQAMAN, K., AND DANUSER, G. Linking data to models: data regression. *Nat Rev Mol Cell Biol 7*, 11 (2006), 813–9.

References

86. JEONG, H., TOMBOR, B., ALBERT, R., OLTVAI, Z. N., AND BARABÁSI, A. L. The large-scale organization of metabolic networks. *Nature 407*, 6804 (2000), 651–4.

87. KAERN, M., ELSTON, T. C., BLAKE, W. J., AND COLLINS, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet 6*, 6 (2005), 451–64.

88. KAO, K. C., TRAN, L. M., AND LIAO, J. C. A global regulatory role of gluconeogenic genes in *Escherichia coli* revealed by transcriptome network analysis. *J Biol Chem 280*, 43 (2005), 36079–87.

89. KEARNS, D. B., CHU, F., RUDNER, R., AND LOSICK, R. Genes governing swarming in *Bacillus subtilis* and evidence for a phase variation mechanism controlling surface motility. *Mol Microbiol 52*, 2 (2004), 357–69.

90. KERR, B., NEUHAUSER, C., BOHANNAN, B. J. M., AND DEAN, A. M. Local migration promotes competitive restraint in a host-pathogen 'tragedy of the commons'. *Nature 442*, 7098 (2006), 75–8.

91. KERR, B., RILEY, M. A., FELDMAN, M. W., AND BOHANNAN, B. J. M. Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. *Nature 418*, 6894 (2002), 171–4.

92. KESELER, I. M., BONAVIDES-MARTÍNEZ, C., COLLADO-VIDES, J., GAMA-CASTRO, S., GUNSALUS, R. P., JOHNSON, D. A., KRUMMENACKER, M., NOLAN, L. M., PALEY, S., PAULSEN, I. T., PERALTA-GIL, M., SANTOS-ZAVALETA, A., SHEARER, A. G., AND KARP, P. D. Ecocyc: a comprehensive view of *Escherichia coli* biology. *Nucleic Acids Res 37*, Database issue (2009), D464–70.

93. KIM, J.-W., AND DANG, C. V. Multifaceted roles of glycolytic enzymes. *Trends Biochem Sci 30*, 3 (2005), 142–50.

94. KING, T., ISHIHAMA, A., KORI, A., AND FERENCI, T. A regulatory trade-off as a source of strain variation in the species *Escherichia coli*. *J Bacteriol 186*, 17 (2004), 5614–20.

95. KITANO, H. Computational systems biology. *Nature 420*, 6912 (2002), 206–10.

96. KITANO, H. Systems biology: a brief overview. *Science 295*, 5560 (2002), 1662–4.

97. KITANO, H. Cancer robustness: tumour tactics. *Nature 426*, 6963 (2003), 125.

98. KITANO, H. Biological robustness. *Nat Rev Genet 5*, 11 (2004), 826–37.

99. KITANO, H. Cancer as a robust system: implications for anticancer therapy. *Nat Rev Cancer 4*, 3 (2004), 227–35.

100. KOROBKOVA, E., EMONET, T., VILAR, J. M. G., SHIMIZU, T. S., AND CLUZEL, P. From molecular noise to behavioural variability in a single bacterium. *Nature 428*, 6982 (2004), 574–8.

101. KOTTE, O., AND HEINEMANN, M. A divide-and-conquer approach to analyze underdetermined biochemical models. *Bioinformatics 25*, 4 (2009), 519–25.

102. KREMLING, A., BETTENBROCK, K., AND GILLES, E. D. A feed-forward loop guarantees robust behavior in *Escherichia coli* carbohydrate uptake. *Bioinformatics 24*, 5 (2008), 704–10.

103. KRISHNA, S., SEMSEY, S., AND SNEPPEN, K. Combinatorics of feedback in cellular uptake and metabolism of small molecules. *Proc Natl Acad Sci U S A 104*, 52 (2007), 20815–9.

104. KUEPFER, L., PETER, M., SAUER, U., AND STELLING, J. Ensemble modeling for analysis of cell signaling dynamics. *Nat Biotechnol 25*, 9 (2007), 1001–6.

105. KÜMMEL, A., PANKE, S., AND HEINEMANN, M. Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Mol Syst Biol 2* (2006), 2006.0034.

106. KUSSELL, E., KISHONY, R., BALABAN, N. Q., AND LEIBLER, S. Bacterial persistence: a model of survival in changing environments. *Genetics 169*, 4 (2005), 1807–14.

107. KUSSELL, E., AND LEIBLER, S. Phenotypic diversity, population growth, and information in fluctuating environments. *Science 309*, 5743 (2005), 2075–8.

108. LANGE, V., PICOTTI, P., DOMON, B., AND AEBERSOLD, R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol 4* (2008), 222.

109. LAPORTE, D. C., WALSH, K., AND KOSHLAND, JR, D. E. The branch point effect. Ultrasensitivity and subsensitivity to metabolic control. *J Biol Chem 259*, 22 (1984), 14068–75.

110. LE GAC, M., BRAZAS, M. D., BERTRAND, M., TYERMAN, J. G., SPENCER, C. C., HANCOCK, R. E. W., AND DOEBELI, M. Metabolic changes associated with adaptive diversification in *Escherichia coli*. *Genetics 178*, 2 (2008), 1049–60.

References

111. Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I., Zeitlinger, J., Jennings, E. G., Murray, H. L., Gordon, D. B., Ren, B., Wyrick, J. J., Tagne, J.-B., Volkert, T. L., Fraenkel, E., Gifford, D. K., and Young, R. A. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science 298*, 5594 (2002), 799–804.

112. Leigh, E. G. How does selection reconcile individual advantage with the good of the group? *Proc Natl Acad Sci U S A 74*, 10 (1977), 4542–4546.

113. Lewis, K. Persister cells, dormancy and infectious disease. *Nat Rev Microbiol 5*, 1 (2007), 48–56.

114. Liebermeister, W., and Klipp, E. Biochemical networks with uncertain parameters. *Syst Biol (Stevenage) 152*, 3 (2005), 97–107.

115. Lloyd, W. F. *Two lectures on the checks to population.* Goldsmiths'-Kress library of economic literature ; no. 27925. Oxford : Printed for the author by S. Collingwood, 1833.

116. Løbner-Olesen, A. Distribution of minichromosomes in individual *Escherichia coli* cells: implications for replication control. *EMBO J 18*, 6 (1999), 1712–21.

117. Loewen, P. C., Hu, B., Strutinsky, J., and Sparling, R. Regulation in the rpoS regulon of *Escherichia coli*. *Can J Microbiol 44*, 8 (1998), 707–17.

118. Lorca, G. L., Ezersky, A., Lunin, V. V., Walker, J. R., Altamentova, S., Evdokimova, E., Vedadi, M., Bochkarev, A., and Savchenko, A. Glyoxylate and pyruvate are antagonistic effectors of the *Escherichia coli* IclR transcriptional regulator. *J Biol Chem 282*, 22 (2007), 16476–91.

119. Lowry, O. H., Carter, J., Ward, J. B., and Glaser, L. The effect of carbon and nitrogen sources on the level of metabolic intermediates in *Escherichia coli*. *J Biol Chem 246*, 21 (1971), 6511–21.

120. Lowy, F. D. *Staphylococcus aureus* infections. *N Engl J Med 339*, 8 (1998), 520–32.

121. Ma, H.-W., Buer, J., and Zeng, A.-P. Hierarchical structure and modules in the *Escherichia coli* transcriptional regulatory network revealed by a new top-down approach. *BMC Bioinformatics 5* (2004), 199.

122. Ma, H.-W., Zhao, X.-M., Yuan, Y.-J., and Zeng, A.-P. Decomposition of metabolic network into functional modules based on the global connectivity structure of reaction graph. *Bioinformatics 20*, 12 (2004), 1870–6.

123. MacLean, R. C. The tragedy of the commons in microbial populations: insights from theoretical, comparative and experimental studies. *Heredity 100*, 5 (2008), 471–7.

124. MacLean, R. C., and Gudelj, I. Resource competition and social conflict in experimental populations of yeast. *Nature 441*, 7092 (2006), 498–501.

125. Maloney, P. C., and Rotman, B. Distribution of suboptimally induced b-D-galactosidase in *Escherichia coli*. The enzyme content of individual cells. *J Mol Biol 73*, 1 (1973), 77–91.

126. Manczak, M., Park, B. S., Jung, Y., and Reddy, P. H. Differential expression of oxidative phosphorylation genes in patients with Alzheimer's disease: implications for early mitochondrial dysfunction and oxidative damage. *Neuromolecular Med 5*, 2 (2004), 147–62.

127. Martin, S. J., Beekman, M., Wossler, T. C., and Ratnieks, F. L. W. Parasitic cape honeybee workers, *Apis mellifera capensis*, evade policing. *Nature 415*, 6868 (2002), 163–5.

128. Martínez-Antonio, A., Janga, S. C., Salgado, H., and Collado-Vides, J. Internal-sensing machinery directs the activity of the regulatory network in *Escherichia coli. Trends Microbiol 14*, 1 (2006), 22–7.

129. Maynard Smith, J. *Evolution and the Theory of Games.* Cambridge University Press, 1982.

130. McAdams, H. H., and Arkin, A. Stochastic mechanisms in gene expression. *Proc Natl Acad Sci U S A 94*, 3 (1997), 814–9.

131. McAdams, H. H., and Arkin, A. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet 15*, 2 (1999), 65–9.

132. Mettetal, J. T., Muzzey, D., Gómez-Uribe, C., and van Oudenaarden, A. The frequency dependence of osmo-adaptation in *Saccharomyces cerevisiae. Science 319*, 5862 (2008), 482–4.

133. Monod, J. The growth of bacterial cultures. *Annu. Rev. Microbiol. 3* (1949), 371–94.

134. Monod, J., and Jacob, F. Teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harb Symp Quant Biol 26* (1961), 389–401.

135. Mootha, V. K., Lindgren, C. M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., Houstis, N., Daly, M. J., Patterson, N., Mesirov, J. P.,

References

Golub, T. R., Tamayo, P., Spiegelman, B., Lander, E. S., Hirschhorn, J. N., Altshuler, D., and Groop, L. C. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet 34*, 3 (2003), 267–73.

136. Newlands, S., Levitt, L. K., Robinson, C. S., Karpf, A. B., Hodgson, V. R., Wade, R. P., and Hardeman, E. C. Transcription occurs in pulses in muscle fibers. *Genes Dev 12*, 17 (1998), 2748–58.

137. Novick, A., and Weiner, M. Enzyme induction as an all-or-none phenomenon. *Proc Natl Acad Sci U S A 43*, 7 (1957), 553–66.

138. Nowak, M. A., Bonhoeffer, S., and May, R. M. Spatial games and the maintenance of cooperation. *Proc Natl Acad Sci U S A 91*, 11 (1994), 4877–81.

139. Nowak, M. A., and Sigmund, K. Bacterial game dynamics. *Nature 418*, 6894 (2002), 138–9.

140. Nurse, P. Systems biology: understanding cells. *Nature 424*, 6951 (2003), 883.

141. Oh, M. K., and Liao, J. C. Gene expression profiling by DNA microarrays and metabolic fluxes in *Escherichia coli*. *Biotechnol Prog 16*, 2 (2000), 278–86.

142. Oh, M.-K., Rohlin, L., Kao, K. C., and Liao, J. C. Global expression profiling of acetate-grown *Escherichia coli*. *J Biol Chem 277*, 15 (2002), 13175–83.

143. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D., and van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nat Genet 31*, 1 (2002), 69–73.

144. Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I., and van Oudenaarden, A. Multistability in the lactose utilization network of *Escherichia coli*. *Nature 427*, 6976 (2004), 737–40.

145. Patil, K. R., and Nielsen, J. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc Natl Acad Sci U S A 102*, 8 (2005), 2685–9.

146. Paulsson, J. Summing up the noise in gene networks. *Nature 427*, 6973 (2004), 415–8.

147. Paulsson, J. Models of stochastic gene expression. *Physics of Life Reviews 2*, 2 (June 2005), 157–175.

148. PAULSSON, J. Prime movers of noisy gene expression. *Nat Genet 37*, 9 (2005), 925–6.

149. PAULSSON, J., BERG, O. G., AND EHRENBERG, M. Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation. *Proc Natl Acad Sci U S A 97*, 13 (2000), 7148–53.

150. PAULSSON, J., AND EHRENBERG, M. Noise in a minimal regulatory network: plasmid copy number control. *Q Rev Biophys 34*, 1 (2001), 1–59.

151. PEDRAZA, J. M., AND PAULSSON, J. Effects of molecular memory and bursting on fluctuations in gene expression. *Science 319*, 5861 (2008), 339–43.

152. PEDRAZA, J. M., AND VAN OUDENAARDEN, A. Noise propagation in gene networks. *Science 307*, 5717 (2005), 1965–9.

153. PENG, L., ARAUZO-BRAVO, M. J., AND SHIMIZU, K. Metabolic flux analysis for a ppc mutant *Escherichia coli* based on 13C-labelling experiments together with enzyme activity assays and intracellular metabolite measurements. *FEMS Microbiol Lett 235*, 1 (2004), 17–23.

154. PICOTTI, P., BODENMILLER, B., MUELLER, L. N., DOMON, B., AND AEBERSOLD, R. Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell 138*, 4 (2009), 795–806.

155. PLUMBRIDGE, J. Regulation of gene expression in the PTS in *Escherichia coli*: the role and interactions of Mlc. *Curr Opin Microbiol 5*, 2 (2002), 187–93.

156. PTASHNE, M. Regulation of transcription: from lambda to eukaryotes. *Trends Biochem Sci 30*, 6 (2005), 275–9.

157. QUAIL, M. A., AND GUEST, J. R. Purification, characterization and mode of action of PdhR, the transcriptional repressor of the pdhR-aceEF-lpd operon of *Escherichia coli*. *Mol Microbiol 15*, 3 (1995), 519–29.

158. RAHMAN, M., HASAN, M. R., OBA, T., AND SHIMIZU, K. Effect of rpoS gene knockout on the metabolism of *Escherichia coli* during exponential growth phase and early stationary phase based on gene expressions, enzyme activities and intracellular metabolite concentrations. *Biotechnol Bioeng 94*, 3 (2006), 585–95.

159. RAINEY, P. B., AND RAINEY, K. Evolution of cooperation and conflict in experimental bacterial populations. *Nature 425*, 6953 (2003), 72–4.

160. RAJ, A., VAN DEN BOGAARD, P., RIFKIN, S. A., VAN OUDENAARDEN, A., AND TYAGI, S. Imaging individual mrna molecules using multiple singly labeled probes. *Nat Methods 5*, 10 (2008), 877–9.

References

161. RAJ, A., AND VAN OUDENAARDEN, A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell 135*, 2 (2008), 216–26.

162. RAMSEIER, T. M., NÈGRE, D., CORTAY, J. C., SCARABEL, M., COZZONE, A. J., AND SAIER, JR, M. H. In vitro binding of the pleiotropic transcriptional regulatory protein, FruR, to the fru, pps, ace, pts and icd operons of *Escherichia coli* and *Salmonella typhimurium*. *J Mol Biol 234*, 1 (1993), 28–44.

163. RANKIN, D. J., BARGUM, K., AND KOKKO, H. The tragedy of the commons in evolutionary biology. *Trends Ecol Evol 22*, 12 (2007), 643–51.

164. RAO, C. V., AND ARKIN, A. P. Control motifs for intracellular regulatory networks. *Annu Rev Biomed Eng 3* (2001), 391–419.

165. RAO, C. V., KIRBY, J. R., AND ARKIN, A. P. Design and diversity in bacterial chemotaxis: a comparative study in *Escherichia coli* and *Bacillus subtilis*. *PLoS Biol 2*, 2 (2004), E49.

166. RAO, C. V., WOLF, D. M., AND ARKIN, A. P. Control, exploitation and tolerance of intracellular noise. *Nature 420*, 6912 (2002), 231–7.

167. RASER, J. M., AND O'SHEA, E. K. Control of stochasticity in eukaryotic gene expression. *Science 304*, 5678 (2004), 1811–4.

168. RASER, J. M., AND O'SHEA, E. K. Noise in gene expression: origins, consequences, and control. *Science 309*, 5743 (2005), 2010–3.

169. RAVASZ, E., SOMERA, A. L., MONGRU, D. A., OLTVAI, Z. N., AND BARABÁSI, A. L. Hierarchical organization of modularity in metabolic networks. *Science 297*, 5586 (2002), 1551–5.

170. REA, S. L., WU, D., CYPSER, J. R., VAUPEL, J. W., AND JOHNSON, T. E. A stress-sensitive reporter predicts longevity in isogenic populations of *Caenorhabditis elegans*. *Nat Genet 37*, 8 (2005), 894–8.

171. REECE, R. J., BEYNON, L., HOLDEN, S., HUGHES, A. D., RÉBORA, K., AND SELLICK, C. A. Nutrient-regulated gene expression in eukaryotes. *Biochem Soc Symp*, 73 (2006), 85–96.

172. REED, J. L., VO, T. D., SCHILLING, C. H., AND PALSSON, B. O. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol 4*, 9 (2003), R54.

173. ROBERTSON, D. L., HAHN, B. H., AND SHARP, P. M. Recombination in aids viruses. *J Mol Evol 40*, 3 (1995), 249–59.

174. ROBERTSON, D. L., SHARP, P. M., MCCUTCHAN, F. E., AND HAHN, B. H. Recombination in hiv-1. *Nature 374*, 6518 (1995), 124–6.

175. ROBISON, K., MCGUIRE, A. M., AND CHURCH, G. M. A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete *Escherichia coli* K-12 genome. *J Mol Biol 284*, 2 (1998), 241–54.

176. RUSSO-MARIE, F., ROEDERER, M., SAGER, B., HERZENBERG, L. A., AND KAISER, D. Beta-galactosidase activity in single differentiating bacterial cells. *Proc Natl Acad Sci U S A 90*, 17 (1993), 8194–8.

177. SAIER, JR, M. H., AND RAMSEIER, T. M. The catabolite repressor/activator (Cra) protein of enteric bacteria. *J Bacteriol 178*, 12 (1996), 3411–7.

178. SAIER, JR, M. H., AND RAMSEIER TOM M. AN REIZER, J. Regulation of carbon utilization. In *EcoSal — Escherichia coli and Salmonella: Cellular and Molecular Biology*, A. Böck, R. Curtiss III, J. Kaper, P. D. Karp, F. C. Neidhardt, T. Nyström, J. M. Slauch, C. L. Squires, and D. Ussery, Eds. ASM Press, Washington, DC, 2008, ch. 85.

179. SAKA, K., TADENUMA, M., NAKADE, S., TANAKA, N., SUGAWARA, H., NISHIKAWA, K., ICHIYOSHI, N., KITAGAWA, M., MORI, H., OGASAWARA, N., AND NISHIMURA, A. A complete set of *Escherichia coli* open reading frames in mobile plasmids facilitating genetic studies. *DNA Res 12*, 1 (2005), 63–8.

180. SAMOILOV, M., ARKIN, A., AND ROSS, J. Signal processing by simple chemical systems. *Journal of Physical Chemistry a 106*, 43 (Oct. 2002), 10205–10221.

181. SAUER, U., AND EIKMANNS, B. J. The PEP-pyruvate-oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMS Microbiol Rev 29*, 4 (2005), 765–94.

182. SAVAGEAU, M. A. Comparison of classical and autogenous systems of regulation in inducible operons. *Nature 252*, 5484 (1974), 546–9.

183. SAVAGEAU, M. A. *Biochemical systems analysis: a study of function and design in molecular biology*. Addison-Wesley, Reading, MA, 1976.

184. SCHLEGEL, A., DANOT, O., RICHET, E., FERENCI, T., AND BOOS, W. The N terminus of the *Escherichia coli* transcription activator MalT is the domain of interaction with MalY. *J Bacteriol 184*, 11 (2002), 3069–77.

185. SCHOEBERL, B., EICHLER-JONSSON, C., GILLES, E. D., AND MÜLLER, G. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat Biotechnol 20*, 4 (2002), 370–5.

References

186. Schroedinger, E. *What is Life?* Cambridge University Press, 1944.

187. Seger, J., and Brockman, H. What is bet-hedging? In *Oxford Surveys in Evolutionary Biology*, P. Harvey and L. Partridge, Eds. Oxford University Press, 1987, pp. 182–211.

188. Seggewiss, J., Becker, K., Kotte, O., Eisenacher, M., Yazdi, M. R. K., Fischer, A., McNamara, P., Al Laham, N., Proctor, R., Peters, G., Heinemann, M., and von Eiff, C. Reporter metabolite analysis of transcriptional profiles of a *Staphylococcus aureus* strain with normal phenotype and its isogenic hemB mutant displaying the small-colony-variant phenotype. *J Bacteriol 188*, 22 (2006), 7765–77.

189. Seshasayee, A. S. N., Bertone, P., Fraser, G. M., and Luscombe, N. M. Transcriptional regulatory networks in bacteria: from input signals to output responses. *Curr Opin Microbiol 9*, 5 (2006), 511–9.

190. Shen-Orr, S., Milo, R., Mangan, S., and Alon, U. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genetics 31*, 1 (May 2002), 64–68.

191. Shi, Y., and Shi, Y. Metabolic enzymes and coenzymes in transcription–a direct link between metabolism and transcription? *Trends Genet 20*, 9 (2004), 445–52.

192. Shin, T., Kraemer, D., Pryor, J., Liu, L., Rugila, J., Howe, L., Buck, S., Murphy, K., Lyons, L., and Westhusin, M. A cat cloned by nuclear transplantation. *Nature 415*, 6874 (2002), 859.

193. Smits, W. K., Kuipers, O. P., and Veening, J.-W. Phenotypic variation in bacteria: the role of feedback regulation. *Nat Rev Microbiol 4*, 4 (2006), 259–71.

194. Smolen, P., Baxter, D. A., and Byrne, J. H. Frequency selectivity, multistability, and oscillations emerge from models of genetic regulatory systems. *Am J Physiol 274*, 2 Pt 1 (1998), C531–42.

195. Southward, C. M., and Surette, M. G. The dynamic microbe: green fluorescent protein brings bacteria to light. *Mol Microbiol 45*, 5 (2002), 1191–6.

196. Spira, B., Hu, X., and Ferenci, T. Strain variation in ppGpp concentration and RpoS levels in laboratory strains of *Escherichia coli* K-12. *Microbiology 154*, Pt 9 (2008), 2887–95.

197. Spudich, J. L., and Koshland, Jr, D. E. Non-genetic individuality: chance in the single cell. *Nature 262*, 5568 (1976), 467–71.

198. STELLING, J. Mathematical models in microbial systems biology. *Curr Opin Microbiol 7*, 5 (2004), 513–8.

199. STELLING, J., SAUER, U., SZALLASI, Z., DOYLE, 3RD, F. J., AND DOYLE, J. Robustness of cellular functions. *Cell 118*, 6 (2004), 675–85.

200. STÜLKE, J., AND HILLEN, W. Carbon catabolite repression in bacteria. *Curr Opin Microbiol 2*, 2 (1999), 195–201.

201. STUMPF, M. P. H., LAIDLAW, Z., AND JANSEN, V. A. A. Herpes viruses hedge their bets. *Proc Natl Acad Sci U S A 99*, 23 (2002), 15234–7.

202. SÜEL, G. M., GARCIA-OJALVO, J., LIBERMAN, L. M., AND ELOWITZ, M. B. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature 440*, 7083 (2006), 545–50.

203. SÜEL, G. M., KULKARNI, R. P., DWORKIN, J., GARCIA-OJALVO, J., AND ELOWITZ, M. B. Tunability and noise dependence in differentiation dynamics. *Science 315*, 5819 (2007), 1716–9.

204. TAO, H., BAUSCH, C., RICHMOND, C., BLATTNER, F. R., AND CONWAY, T. Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media. *J Bacteriol 181*, 20 (1999), 6425–40.

205. THATTAI, M., AND VAN OUDENAARDEN, A. Stochastic gene expression in fluctuating environments. *Genetics 167*, 1 (2004), 523–30.

206. TONG, A. H. Y., LESAGE, G., BADER, G. D., DING, H., XU, H., XIN, X., YOUNG, J., BERRIZ, G. F., BROST, R. L., CHANG, M., CHEN, Y., CHENG, X., CHUA, G., FRIESEN, H., GOLDBERG, D. S., HAYNES, J., HUMPHRIES, C., HE, G., HUSSEIN, S., KE, L., KROGAN, N., LI, Z., LEVINSON, J. N., LU, H., MÉNARD, P., MUNYANA, C., PARSONS, A. B., RYAN, O., TONIKIAN, R., ROBERTS, T., SDICU, A.-M., SHAPIRO, J., SHEIKH, B., SUTER, B., WONG, S. L., ZHANG, L. V., ZHU, H., BURD, C. G., MUNRO, S., SANDER, C., RINE, J., GREENBLATT, J., PETER, M., BRETSCHER, A., BELL, G., ROTH, F. P., BROWN, G. W., ANDREWS, B., BUSSEY, H., AND BOONE, C. Global mapping of the yeast genetic interaction network. *Science 303*, 5659 (2004), 808–13.

207. TSIEN, R. Y. The green fluorescent protein. *Annu Rev Biochem 67* (1998), 509–44.

208. TURNER, P. E., AND CHAO, L. Prisoner's dilemma in an RNA virus. *Nature 398*, 6726 (1999), 441–3.

209. UEDA, H. R., HAGIWARA, M., AND KITANO, H. Robust oscillations within the interlocked feedback model of *Drosophila* circadian rhythm. *J Theor Biol 210*, 4 (2001), 401–6.

References

210. VAN RIEL, N. A. W. Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments. *Brief Bioinform 7*, 4 (2006), 364–74.

211. VARMA, A., AND PALSSON, B. O. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl Environ Microbiol 60*, 10 (1994), 3724–31.

212. VEENING, J.-W., HAMOEN, L. W., AND KUIPERS, O. P. Phosphatases modulate the bistable sporulation gene expression pattern in *Bacillus subtilis*. *Mol Microbiol 56*, 6 (2005), 1481–94.

213. VEENING, J.-W., SMITS, W. K., AND KUIPERS, O. P. Bistability, epigenetics, and bet-hedging in bacteria. *Annu Rev Microbiol 62* (2008), 193–210.

214. VEENING, J.-W., STEWART, E. J., BERNGRUBER, T. W., TADDEI, F., KUIPERS, O. P., AND HAMOEN, L. W. Bet-hedging and epigenetic inheritance in bacterial cell development. *Proc Natl Acad Sci U S A 105*, 11 (2008), 4393–8.

215. VELICER, G. J. Social strife in the microbial world. *Trends Microbiol 11*, 7 (2003), 330–7.

216. VOLTERRA, V. Variations and fluctuations of the number of individuals in animal species living together. In *Animal Ecology*. McGraw–Hill, 1931.

217. VON MERING, C., KRAUSE, R., SNEL, B., CORNELL, M., OLIVER, S. G., FIELDS, S., AND BORK, P. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature 417*, 6887 (2002), 399–403.

218. VON NEUMANN, J., AND MORGENSTERN, O. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

219. VULIC, M., AND KOLTER, R. Evolutionary cheating in *Escherichia coli* stationary phase cultures. *Genetics 158*, 2 (2001), 519–26.

220. WAGNER, A., AND FELL, D. A. The small world inside large metabolic networks. *Proc Biol Sci 268*, 1478 (2001), 1803–10.

221. WALSH, K., AND KOSHLAND, JR, D. E. Determination of flux through the branch point of two metabolic cycles. The tricarboxylic acid cycle and the glyoxylate shunt. *J Biol Chem 259*, 15 (1984), 9646–54.

222. WARBURG, O. On respiratory impairment in cancer cells. *Science 124*, 3215 (1956), 269–70.

223. WEBER, H., POLEN, T., HEUVELING, J., WENDISCH, V. F., AND HENGGE, R. Genome-wide analysis of the general stress response network in *Escherichia coli*: sigmaS-dependent genes, promoters, and sigma factor selectivity. *J Bacteriol 187*, 5 (2005), 1591–603.

224. WEINBERGER, L. S., BURNETT, J. C., TOETTCHER, J. E., ARKIN, A. P., AND SCHAFFER, D. V. Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. *Cell 122*, 2 (2005), 169–82.

225. WEINIG, C., JOHNSTON, J. A., WILLIS, C. G., AND MALOOF, J. N. Antagonistic multilevel selection on size and architecture in variable density settings. *Evolution 61*, 1 (2007), 58–67.

226. WERNER, M., SEMSEY, S., SNEPPEN, K., AND KRISHNA, S. Dynamics of uptake and metabolism of small molecules in cellular response systems. *PLoS One 4*, 3 (2009), e4923.

227. WIENER, N. *Cybernetics or Control and Communication in the Animal and the Machine.* MIT Press, Cambridge, MA, 1948.

228. WOLF, D. M., AND ARKIN, A. P. Motifs, modules and games in bacteria. *Curr Opin Microbiol 6*, 2 (2003), 125–34.

229. WOLF, D. M., VAZIRANI, V. V., AND ARKIN, A. P. Diversity in times of adversity: probabilistic strategies in microbial survival games. *J Theor Biol 234*, 2 (2005), 227–53.

230. WOLFE, A. J. The acetate switch. *Microbiol Mol Biol Rev 69*, 1 (2005), 12–50.

231. YAGIL, G., AND YAGIL, E. On the relation between effector concentration and the rate of induced enzyme synthesis. *Biophys J 11*, 1 (1971), 11–27.

232. YAMAMOTO, S., AND KUTSUKAKE, K. FljA-mediated posttranscriptional control of phase 1 flagellin expression in flagellar phase variation of *Salmonella enterica serovar Typhimurium*. *J Bacteriol 188*, 3 (2006), 958–67.

233. ZHAO, J., BABA, T., MORI, H., AND SHIMIZU, K. Effect of zwf gene knockout on the metabolism of *Escherichia coli* grown on glucose or acetate. *Metab Eng 6*, 2 (2004), 164–74.

234. ZWANZIG, R. *Nonequilibrium statistical mechanics.* Oxford University Press, 2001.

Acknowledgements

Many people contributed to this thesis in diverse ways. I am grateful to

- ... Matthias Heinemann for accepting me as a PhD student, for being an enthusiastic scientist, for his understanding when I needed it, for his supervision throughout the years, and for being on my PhD committee.
- ... Uwe Sauer and Sven Panke for providing a friendly scientific environment, for many helpful discussions, and for being on my PhD committee.
- ... John Lygeros for scientific help on a side project, and for being on my PhD committee.
- ... the ETH–INIT project 'Single Cell Metabolomics' (SCM) and the Roche Research Foundation for funding.
- ... Martin Ackermann for many helpful discussions on evolutionary biology.
- ... Benjamin Volkmer for countless experiments, for sticking with me throughout the 'neverending story', and for his humor and calmness.
- ... Karin Mitosch for her Master Thesis work on a project not included in this thesis, and for her contagious enthusiasm.
- ... Judith Zaugg for her Master Thesis work on the kinetic model presented in this thesis, and also for her contagious enthusiasm.
- ... Markus Rottmar for his Master Thesis work on phenotypic bistability, and for his perseverance.
- ... all people from the Sauer Lab for being such a fun and helpful crowd.
- ... Monika Hartwig for brightening up many days.
- ... Keelin Günther for having arrived in this world.
- ... Sandra Lövenich and Sarah–Maria Fendt for being wonderful friends.
- ... many more wonderful friends across the world for enriching my life.
- ... my family for their unconditional love and support.
- ... Miriam Bauer — without you, where would I be?

List of Publications

KOTTE, O., ACKERMANN, M., AND HEINEMANN, M. On the optimality of responsive phenotypic diversification. *In preparation.*

KOTTE, O., VOLKMER, B., PANKE, S., AND HEINEMANN, M. Phenotypic bistability in *Escherichia coli*'s central carbon metabolism. *Submitted.*

KOTTE, O., ZAUGG, J.B., AND HEINEMANN, M. Bacterial adaptation through distributed sensing of metabolic fluxes. *Submitted.*

KOTTE, O., AND HEINEMANN, M. A divide–and–conquer approach to analyze underdetermined biochemical models. *Bioinformatics 25*, 4 (2009), 51925.

HÜBSCHER, J., JANSEN, A., KOTTE, O. ET AL. Living with an imperfect cell wall: compensation of femAB inactivation in *Staphylococcus aureus. BMC Genomics 8*, 307 (2007).

SEGGEWISS, J., BECKER, K., KOTTE, O. ET AL. Reporter metabolite analysis of transcriptional profiles of a *Staphylococcus aureus* strain with normal phenotype and its isogenic hemB mutant displaying the small-colony-variant phenotype. *J Bacteriol. 188*, 22 (2006), 7765-77.

Curriculum Vitae

Timo Oliver Kotte

Date of birth:	March 22, 1978
Place of birth:	Mannheim, Germany
Citizen of:	Germany and Finland
08/2005–12/2009	ETH Zürich, Switzerland Doctoral Student (PhD) Institute of Molecular Systems Biology and Institute of Process Engineering Scientific advisors: Prof. Sven Panke, Prof. Uwe Sauer, Prof. John Lygeros, and Prof. Matthias Heinemann
01/2005–07/2005	Backpacking in New Zealand
10/1998–01/2005	University of Stuttgart, Germany Degree: Dipl.-Ing. Engineering Cybernetics
09/2004–12/2004	BASF AG, Ludwigshafen, Germany Internship
08/2003–08/2004	Georgia Institute of Technology, Atlanta, USA Degree: M.Sc. Engineering Science and Mechanics
10/2001–03/2002	University of Stuttgart, Germany Research Assistant
10/2000–09/2001	Backpacking in South America
10/1997–10/1998	German Red Cross, Heppenheim, Germany Compulsory civil service
08/1997–10/1997	ABB Industrietechnik AG, Mannheim, Germany Internship
08/1988–06/1997	Lessing Gymnasium, Lampertheim, Germany Degree: Abitur

Zürich, 2009