

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

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 abgeleitet wurde. Als Modellsystem dient *Escherichia coli* 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-

Zusammenfassung

lytischen und gluconeogenetischen 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 unser 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 glukoneogenetischen 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 gluconeogenetisches 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 vereinzelt 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 unbekanntere 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 leistungsorientierten Selektion und ist daher konzeptionell neu. Zusammengefasst 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.