Institut für Biochemie und Biologie
AG Bioinformatik

Analysing concerted criteria for local dynamic properties of metabolic systems

Dissertation
zur Erlangung des akademischen Grades
"doctor rerum naturalium"
(Dr. rer. nat.)
in der Wissenschaftsdisziplin "Bioinformatik"

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Potsdam

von
Dorothee Girbig

Potsdam, den 04.03.2014

Published online at the
Institutional Repository of the University of Potsdam:
URL http://opus.kobv.de/ubp/volltexte/2014/7201/
URN urn:nbn:de:kobv:517-opus-72017
http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-72017

Abstract

Metabolic systems tend to exhibit steady states that can be measured in terms of their concentrations and fluxes. These measurements can be regarded as a phenotypic representation of all the complex interactions and regulatory mechanisms taking place in the underlying metabolic network. Such interactions determine the system's response to external perturbations and are responsible, for example, for its asymptotic stability or for oscillatory trajectories around the steady state. However, determining these perturbation responses in the absence of fully specified kinetic models remains an important challenge of computational systems biology.

Structural kinetic modeling (SKM) is a framework to analyse whether a metabolic steady state remains stable under perturbation, without requiring detailed knowledge about individual rate equations. It provides a parameterised representation of the system's Jacobian matrix in which the model parameters encode information about the enzyme-metabolite interactions. Stability criteria can be derived by generating a large number of structural kinetic models (SK-models) with randomly sampled parameter sets and evaluating the resulting Jacobian matrices. The parameter space can be analysed statistically in order to detect network positions that contribute significantly to the perturbation response. Because the sampled parameters are equivalent to the elasticities used in metabolic control analysis (MCA), the results are easy to interpret biologically.

In this project, the SKM framework was extended by several novel methodological improvements. These improvements were evaluated in a simulation study using a set of small example pathways with simple Michaelis Menten rate laws. Afterwards, a detailed analysis of the dynamic properties of the neuronal TCA cycle was performed in order to demonstrate how the new insights obtained in this work could be used for the study of complex metabolic systems.

The first improvement was achieved by examining the biological feasibility of the elasticity combinations created during Monte Carlo sampling. Using a set of small example systems, the findings showed that the majority of sampled SK-models would yield negative kinetic parameters if they were translated back into

kinetic models. To overcome this problem, a simple criterion was formulated that mitigates such infeasible models and the application of this criterion changed the conclusions of the SKM experiment.

The second improvement of this work was the application of supervised machine-learning approaches in order to analyse SKM experiments. So far, SKM experiments have focused on the detection of individual enzymes to identify single reactions important for maintaining the stability or oscillatory trajectories. In this work, this approach was extended by demonstrating how SKM enables the detection of ensembles of enzymes or metabolites that act together in an orchestrated manner to coordinate the pathway‘s response to perturbations. In doing so, stable and unstable states served as class labels, and classifiers were trained to detect elasticity regions associated with stability and instability. Classification was performed using decision trees and relevance vector machines (RVMs). The decision trees produced good classification accuracy in terms of model bias and generalizability. RVMs outperformed decision trees when applied to small models, but encountered severe problems when applied to larger systems because of their high runtime requirements. The decision tree rulesets were analysed statistically and individually in order to explore the role of individual enzymes or metabolites in controlling the system's trajectories around steady states.

The third improvement of this work was the establishment of a relationship between the SKM framework and the related field of MCA. In particular, it was shown how the sampled elasticities could be converted to flux control coefficients, which were then investigated for their predictive information content in classifier training.

After evaluation on the small example pathways, the methodology was used to study two steady states of the neuronal TCA cycle with respect to their intrinsic mechanisms responsible for stability or instability. The findings showed that several elasticities were jointly coordinated to control stability and that the main source for potential instabilities were mutations in the enzyme α-ketoglutarate dehydrogenase.

Zusammenfassung

Metabolische Systeme neigen zur Ausbildung von Fließgleichgewichten, deren Konzentrationen und Reaktionsflüsse experimentell charakterisierbar sind. Derartige Messungen bieten eine phänotypische Repräsentation der zahlreichen Interaktionen und regulatorischen Mechanismen des zugrundeliegenden metabolischen Netzwerks. Diese Interaktionen bestimmen die Reaktion des Systems auf externe Perturbationen, wie z.B. dessen asymptotische Stabilität und Oszillationen. Die Charakterisierung solcher Eigenschaften ist jedoch schwierig, wenn kein entsprechendes kinetisches Modell mit allen Ratengleichungen und kinetischen Parametern für das untersuchte System zur Verfügung steht.

Die strukturelle kinetische Modellierung (SKM) ermöglicht die Untersuchung dynamischer Eigenschaften wie Stabilität oder Oszillationen, ohne die Ratengleichungen und zugehörigen Parameter im Detail zu kennen. Statt dessen liefert sie eine parametrisierte Repräsentation der Jacobimatrix, in welcher die einzelnen Parameter Informationen über die Sättigung der Enzyme des Systems mit ihren Substraten kodieren. Die Parameter entsprechen dabei den Elastizitäten aus der metabolischen Kontrollanalyse, was ihre biologische Interpretation vereinfacht. Stabilitätskriterien werden durch Monte Carlo Verfahren hergeleitet, wobei zunächst eine große Anzahl struktureller kinetische Modelle (SK-Modelle) mit zufällig gezogenen Parametermengen generiert, und anschließend die resultierenden Jacobimatrizen evaluiert werden. Im Anschluss kann der Parameterraum statistisch analysiert werden, um Enzyme und Metabolite mit signifikantem Einfluss auf die Stabilität zu detektieren.

In der vorliegenden Arbeit wurde das bisherige SKM-Verfahren durch neue methodische Verbesserungen erweitert. Diese Verbesserungen wurden anhand einer Simulationsstudie evaluiert, welche auf kleinen Beispielsystemen mit einfachen Michaelis Menten Kinetiken basierte. Im Anschluss wurden sie für eine detaillierte Analyse der dynamischen Eigenschaften des Zitratzyklus verwendet.

Die erste Erweiterung der bestehenden Methodik wurde durch Untersuchung der biologischen Machbarkeit der zufällig erzeugten Elastizitäten erreicht. Es konnte

gezeigt werden, dass die Mehrheit der zufällig erzeugten SK-Modelle zu negativen Michaeliskonstanten führt. Um dieses Problem anzugehen, wurde ein einfaches Kriterium formuliert, welches das Auftreten solcher biologisch unrealistischer SK-Modelle verhindert. Es konnte gezeigt werden, dass die Anwendung des Kriteriums die Ergebnisse von SKM Experimenten stark beeinflussen kann.

Der zweite Beitrag bezog sich auf die Analyse von SKM-Experimenten mit Hilfe überwachter maschineller Lernverfahren. Bisherige SKM-Studien konzentrierten sich meist auf die Detektion individueller Elastizitäten, um einzelne Reaktionen mit Einfluss auf das Stabilitäts- oder oszillatorische Verhalten zu identifizieren. In dieser Arbeit wurde demonstriert, wie SKM Experimente im Hinblick auf multivariate Muster analysiert werden können, um Elastizitäten zu entdecken, die gemeinsam auf orchestrierte und koordinierte Weise die Eigenschaften des Systems bestimmen. Sowohl Entscheidungsbäume als auch Relevanzvektormaschinen (RVMs) wurden als Klassifikatoren eingesetzt. Während Entscheidungsbäume im allgemeinen gute Klassifikationsergebnisse lieferten, scheiterten RVMs an ihren großen Laufzeitbedürfnissen bei Anwendung auf ein komplexes System wie den Zitratzyklus. Hergeleitete Entscheidungsbaumregeln wurden sowohl statistisch als auch individuell analysiert, um die Koordination von Enzymen und Metaboliten in der Kontrolle von Trajektorien des Systems zu untersuchen.

Der dritte Beitrag, welcher in dieser Arbeit vorgestellt wurde, war die Etablierung der Beziehung zwischen SKM und der metabolischer Kontrollanalyse. Insbesondere wurde gezeigt, wie die zufällig generierten Elastizitäten in Flusskontrollkoeffizienten umgewandelt werden. Diese wurden im Anschluss bezüglich ihres Informationsgehaltes zum Klassifikationstraining untersucht.

Nach der Evaluierung anhand einiger kleiner Beispielsysteme wurde die neue Methodik auf die Studie zweier Fließgleichgewichte des neuronalen Zitratzyklus angewandt, um intrinsische Mechanismen für Stabilität oder Instabilität zu finden. Die Ergebnisse identifizierten Mutationen im Enzym α-ketoglutarate dehydrogenase als wahrscheinlichste Quelle für Instabilitäten.

Allgemeinverständliche Zusammenfassung

Der zelluläre Stoffwechsel enthält eine Vielzahl von Enzymen, welche spezifische biologische Substanzen (auch Metabolite genannt) ineinander umwandeln. Die Menge und Aktivität der einzelnen Enzyme, zusammen mit der Verfügbarkeit an energiereichen Ressourcen, bestimmt die Funktionalität jeder einzelnen Zelle. Metabolitkonzentrationen können experimentell bestimmt werden, wobei es ist eine zentrale Aufgabe der aufstrebenden Disziplin der Systembiologie ist, aus derartigen Messungen das zugrundeliegende Reaktionsnetzwerk zu rekonstruieren, sowie sein Verhalten unter verschiedenen experimentellen Bedingungen vorherzusagen. Hierzu werden in der Regel detaillierte kinetische Modelle konstruiert, die das Verhalten jedes Enzyms genau beschreiben.

Falls das Wissen über einzelne Enzyme nicht für den Bau eines detaillierten kinetischen Modells ausreicht, bietet die verwandte Methodik der strukturellen kinetischen Modellierung (SKM) die Möglichkeit, Aussagen über wichtige Positionen im Reaktionsnetzwerk zu treffen. Hierbei werden experimentell gemessene Daten mit bekannten Informationen über die Umwandlungszahlen jedes Enzyms verbunden. Das zugrundeliegende mathematische Verfahren liefert dann Informationen über die Antwort des Netzwerkes auf äußere Störungen, sowie über diejenigen Netzwerkkomponenten, welche auf diese Antwort maßgeblichen Einfluss haben.

Die vorliegende Arbeit erweitert die SKM Methodik um einige grundlegende Verbesserungen. Zum Beispiel wird ein Filterkriterium hergeleitet, das die zufällige Simulation biologisch unrealistischer Szenarien verhindert, welche bei bisherigen SKM Experimenten ein Problem war. Des Weiteren wird gezeigt, wie mit Hilfe komplexer statistischer Verfahren Netzwerkinteraktionen detektiert werden können, welche gemeinsam in koordinierter Weise Einfluss auf das Systemverhalten ausüben.

Acknowledgements

This work would not have been possible without Prof. Joachim Selbig from the University of Potsdam. Thank you for always encouraging me to pursue my own ideas, but also for providing valuable scientific support when necessary.

My thanks also go to the Max Planck Institute for Molecular Plant Physiology, which enabled me to conduct this work as a member of the International Max Planck Research School Primary Metabolism and Plant Growth.

The bioinformatics groups at the MPI, as well as at the University of Potsdam, not only provided a great working atmosphere, but also friendships that made my time in these groups greatly enjoyable.

I also want to thank my family for their numerous support. My biggest thanks go to my daughter Sophia for all her patience and cheerfulness, as well as to Liam Childs, for being a great partner in my private as well as in my scientific life.

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und unter keinen anderen als den von mir angegebenen Quellen und Hilfsmitteln verfasst habe.

Ferner erkläre ich, dass ich bisher weder an der Universität Potsdam noch anderweitig versucht habe, eine Dissertation einzureichen oder mich einer Doktorprüfung zu unterziehen.

Potsdam, 3. März 2014

_____

Dorothee Girbig

Contents

1. Introduction

In the last decade, biological research has been revolutionized by the advent of high-throughput technologies which enable the measurement of the components of a cell on a large scale. This gave rise to the discipline of systems biology which aims at integrating and interpreting this vast body of information on a system level in order to understand the underlying processes which take part in the cell (Dunn *et al*, 2010; Nägele and Weckwerth, 2012). As a consequence, elucidating the way in which individual components interact in a biological network and coordinate its function has become one of the major goals of systems biology (Cornish-Bowden, 2006; Bruggeman and Westerhoff, 2007; Yuan *et al*, 2008). Metabolic networks are a subclass of biological networks that describe the enzymatically catalysed interconversion of small molecules (metabolites) in the living organism. The study of metabolism has gained considerable interest in the last decade because it provides a snapshot or 'phenotype' of the processes taking place in the cell (Stitt *et al*, 2010). In doing so, it tries to describe the ultimate responses of an organism to different environmental conditions, or to treatments within an experiment.

The progress in experimental techniques has enabled the creation of large amounts of metabolomics data (Zhang *et al*, 2011). Such datasets typically measure either the concentrations of individual metabolites or the rates of the enzymatic reactions converting them into each other. Consequently, they provide a detailed description of the metabolic state of a cell or organism in response to specific environmental or experimental conditions. However, this phenotypic representation renders it difficult to elucidate the exact mechanisms that caused the network to behave in the observed manner. To this end, sophisticated statistical analysis techniques have been proposed to analyse metabolomics datasets and to obtain information about the underlying structure and regulatory mechanisms of a metabolic network (Boccard *et al*, 2010; Kholodenko *et al*, 2012; Xia *et al*, 2012; Franceschi *et al*, 2013). However, the reconstruction of a more detailed picture of the processes taking place in a metabolic network requires mathematical modelling (Rodríguez and Infante, 2009).

1.1. Mathematical modelling of metabolic networks

Mathematical models provide formal definitions of the processes taking place in a metabolic network and allow simulations and predictions of system behaviours under different conditions. Most approaches for mathematical modelling of metabolic networks can be broadly subdivided into (1) structural modelling and (2) kinetic modelling. Both categories differ in the extent of knowledge they require about the system as well as in the types of results they are able to obtain (Steuer, 2007).

Analysing metabolic steady states by structural modelling

In general, metabolic systems are assumed to operate in steady working states. These so-called 'steady states' are characterised by constant concentrations and reaction rates. Usually, a metabolic system is assumed to maintain a particular steady state until changes in external factors require adjustments of the reaction rates, typically resulting in the transition into a new steady state. Such factors can be, for example, changes in nutrient availability, the requirement to adjust the production of certain end products, or hormonal signals that lead to changes in the abundance or in the catalytic activity of enzymes involved in a pathway.

Structural modelling typically focusses on the analysis of reaction rates in steady state scenarios without considering possible time-dependent changes in the system's components. This restriction enables it to rely solely on information about the network structure (stoichiometry). A prominent example is flux balance analysis (FBA), where the optimal distribution of steady state fluxes is computed by linear optimization (Orth *et al*, 2010).

Thanks to the rapid progress in next-generation sequencing, which allows the characterization of many of the genes and enzymes present in a cell, the structure of a metabolic network can nowadays be comprehensively reconstructed on the genome level (Thiele and Palsson, 2010). This enables the wide-spread application of structural modelling techniques. For example, it can help to predict the effects of genetic modifications in biotechnology, or to understand how certain diseases affect the cell's metabolism (Folger *et al*, 2011).

Capturing dynamic system behaviour by kinetic modelling

In contrast to structural modelling, kinetic modelling aims at describing the metabolic system by a set of ordinary differential equations (ODEs) that comprehensively capture the time-dependent behaviour of every reaction in the system (Pfau *et al*, 2011). This detailed representation enables the analysis of the dynamic properties of the network. In contrast to structural modeling, such an analysis is not restricted to steady states. For example, it allows the prediction of system responses to external stimuli, the incorporation of metabolic inhibitors or activators, and the simulation of oscillatory behaviour (Heinrich and Schuster, 1996). However, this approach relies on detailed knowledge about all enzymatic rate laws and kinetic parameters in the system, which are often difficult and labour-intensive to obtain experimentally (Li *et al*, 2012). As a consequence, kinetic models are usually small and focus on specific pathways or subsystems of a metabolic network.

Metabolic control analysis

Building upon the kinetic modelling framework, specialized methods have been developed to investigate specific aspects of a metabolic system. One of these approaches is metabolic control analysis (MCA), which aims at predicting the impact of changes in certain system parameters (e.g. enzyme concentrations) on the overall system (Fell and Sauro, 1985; Kacser and Porteous, 1987; Fell, 1997). MCA quantifies the amount by which different system parameters exhibit control on the fluxes or on the individual metabolites within a metabolic system. These quantities are captured in variables called flux control coefficients (FCCs) and concentration control coefficients. A third category of MCA variables, the so-called elasticities, quantifies the local influences of changes in metabolite concentrations on the individual reaction rates. Elasticities describe the sensitivity of the individual reaction rates to changes in metabolite concentrations and can be understood as partial derivatives of the reaction rates with respect to each metabolite in the system.

The study of FCCs elucidates the effect of changes in an enzyme's concentration on the overall fluxes through the pathway. One of the fundamental insights gained in this manner was that the control of flux is typically distributed among many different enzymes instead of just depending on a single 'rate-limiting step' (Kacser and Porteous, 1987). Furthermore, it could be shown that the FCCs directly depend on- and can be calculated from the elasticities.

1.2. Stability analysis by metabolic modelling

Besides enabling the simulation of how metabolite concentrations evolve in time, kinetic models also provide information about local dynamic properties of steady states. Such properties comprise, for example, the bifurcation structure, the possibility of oscillations, or the local asymptotic stability of the steady state (Heinrich and Schuster, 1996).

Local stability can be understood as the robustness of a steady state to small perturbations, like those arising from natural fluctuations in the concentrations of enzymes and metabolites in a cellular environment. Consequently, a stable steady state allows the fine-tuned response of the reaction rates to perturbations, eventually enabling the return to the original steady state.

Studying the local dynamic properties of steady states is of great importance because not only do they help us gain a deeper understanding of the functioning of a pathway, but they can also help us to systematically predict what type of system responses can be expected when changing certain parameters experimentally. In a kinetic model, the dynamic properties of a steady state can be derived from the eigenvalues of the corresponding Jacobian matrix. This matrix contains the partial derivatives of the reaction rates, and therefore its computation requires detailed knowledge about the kinetic rate laws, as well as their *in vivo* kinetic parameters. In practice, however, this data is often not available in sufficient accuracy. This problem of incomplete knowledge severely hampers the computation and subsequent analysis of the Jacobian matrix (Jamshidi and Palsson, 2008).

To cope with these uncertainties in kinetic parameters and rate laws, several strategies have been introduced independently of each other in recent years. These methods typically rely on stricter assumptions about the system than used in structural modelling, avoiding the necessity to construct comprehensive kinetic models. Although not offering the full functionality of a kinetic model, this trade-off enables the analysis of properties of the system not accessible by structural modelling alone. For example, methods for local stability analysis under uncertainty have been suggested by Wang *et al* (2004), Steuer *et al* (2006), or van Nes *et al* (2009). All of these methods have in common that they address uncertainties in the model parameters in a statistical manner. More precisely, instead of computing the Jacobian matrix from kinetic rate laws, its elements are sampled randomly using Monte Carlo approaches, yielding ensembles of models that can be analysed with respect to the system properties of interest.

The work of van Nes *et al* (2009) aims at estimating the probabilities of observing local stability in different network motifs in genome-scale metabolic networks. Their ultimate goal is to answer the question about whether local stability is the driving force behind the evolution of metabolic systems. Using the yeast metabolic network as a reference, they show that this hypothesis is not confirmed, and that stability does not seem to be the major selection criterion in metabolic network evolution. One drawback of their method is that it does not investigate particular realizations of steady states in terms of experimentally observed concentrations and reaction rates. Instead, it examines the general ability of a motif to yield stable steady states in principle. As a consequence, it does not account for the situation that a motif could yield a large proportion of unstable steady states in theory, but the steady state realizations actually occurring in the living cell are stable. This issue has been previously discussed in detail in the context of gene-regulatory networks by Doyle and Csete (2005). A similar generalization is performed by Wang *et al* (2004) who aim at applying MCA to the study of systems for which neither kinetic models nor experimentally obtained steady state data are available.

A very similar approach which enables the analysis of realistic and experimentally characterized steady states is the structural kinetic modelling (SKM) framework suggested by Steuer *et al* (2006). It is based on a generalised modelling framework to investigate dynamic properties of real-life systems (Gross and Feudel, 2006). The basic idea of SKM is the construction of a parameterised version of the Jacobian matrix in an observed steady state. Consequently, instead of relying on a detailed set of rate equations, together with accurate estimates of the kinetic parameters, the Jacobian matrix in the observed steady state depends only on a set of structural kinetic model (SK-model) parameters. These parameters are equivalent to the elasticities from MCA and offer a straight-forward biological interpretation. Thus, the parameters describe the influence of changes in metabolite concentrations on the reaction rates in the steady state. In enzymatic reactions, this influence depends largely on the amount of saturation of an enzyme with its metabolites (Wang *et al*, 2004). Experimental values for the steady state's elasticities are often unknown in practice. However, the statistical exploration of the parameter space by Monte Carlo sampling can indicate regions associated with different local dynamic properties of the system. This enables the identification of enzymes and metabolites with major roles in determining the system's response to the fluctuations of metabolites or enzymes within the cellular environment.

1.3. Goals of the thesis

This thesis presents a new methodology for elucidating the stabilizing and destabilizing mechanisms present in metabolic systems. So far, SKM experiments have focused on the detection of individual enzymes to identify single reactions important for maintaining the stability of a steady state (Grimbs *et al*, 2007; Bulik *et al*, 2009). Here, this approach will be extended by demonstrating how SKM enables the detection of ensembles of enzymes or metabolites that act together in an orchestrated manner to coordinate the pathway's response to perturbations. This is achieved by replacing the previously used univariate approach by supervised machine learning in order to search for multivariate patterns of elasticities associated with stability or instability (Girbig *et al*, 2012a; Girbig *et al*, 2012b). As shown in Figure 1.1, the information about stable and unstable states serve as class labels, and classifiers are trained to detect elasticity regions associated with stability and instability. Classification is performed using decision trees and relevance vector machines. Both algorithms lend themselves to this task because they preserve the original feature space, making it possible to interpret the detected patterns in a biological context.

In addition to looking for stabilizing or destabilizing patterns, it will also be investigated whether the proposed machine-learning approach can be applied to study criteria for the emergence of oscillations in metabolic systems (Heinrich and Schuster, 1996).

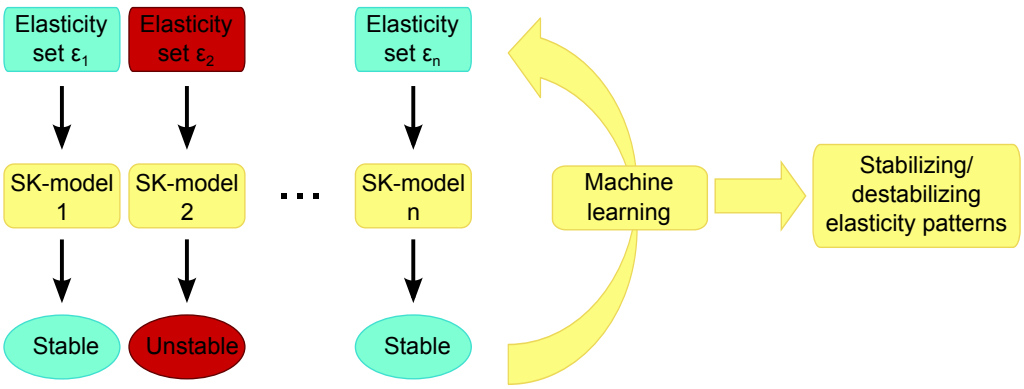


Figure 1.1.: Workflow for Monte Carlo based model generation and the subsequent detection of patterns by machine learning. A large number of SK-models is created based on randomly sampled elasticities and stable/unstable models are identified. Using the elasticities as feature vectors and the stability information as class labels, a classifier can learn those patterns with highest discriminatory power between both classes.

The application of this extended SKM approach is demonstrated on a set of small artificial pathways that enable the analytical assessment of the derived patterns. Furthermore, a real-world example will be presented in form of the neuronal tricarboxylic acid cycle (TCA cycle). The TCA cycle not only plays a central role in the energy metabolism of eukaryotic cells but it also provides metabolic precursors for biosynthetic pathways. The resulting drain of organic compounds is compensated by anapleurotic reactions in order to maintain a continuous flux through the cycle. As a consequence, maintaining the balance between biosynthetic- and energy providing- processes is crucial for mitochondrial energy homeostasis (Koopman *et al*, 2012). The proportion of flux through the cycle compared to adjacent pathways depends on enzyme and metabolite concentrations, as well as the enzymes‘ kinetic properties. As a consequence, the risk that perturbations or enzymatic mutations destabilize the cycle, potentially leading to energetic collapse, varies between enzymes. For example, kinetic models demonstrate the sensitivity of the TCA cycle to α-ketoglutarate dehydrogenase (AKGDH) activity. Indeed, dysfunction of this enzyme has been observed in neurodegenerative disorders (Berndt *et al*, 2012). This motivates the systematic analysis of the role of this and other enzymes as well as the investigation of how they coordinate their activities in order to maintain functional steady states.

1.4. Outline of the thesis

In Chapter 2 of this thesis, the mathematical background of the SKM framework is introduced in detail. Furthermore, this Chapter explains how stabilizing elasticity patterns can be derived by decision trees and RVMs. Details about the construction and analysis of the SK-models are given in Chapter 3. A part of this project was dedicated to the development of a MATLAB toolbox that enables the automated generation and analysis of SK-models. The corresponding Applications Note (Girbig *et al*, 2012b) will be provided in Chapter 4. In Chapter 5 several small example systems are introduced and analysed using the different classification strategies. The same chapter also contains several other analyses elucidating an important properties of SKM that has not been investigated in the literature so far, namely the biological feasibility of the elasticity combinations created during Monte Carlo sampling. Chapter 6 demonstrates how the proposed methodology is applied to study destabilizing mechanisms in the neuronal TCA cycle. A summary and discussion is provided in Chapter 7.

2. Mathematical Background

2.1. A short introduction to kinetic modelling of metabolic networks

Given a metabolic network with m metabolites $S_1, S_2, \dots S_m$ and n reactions $r_1, r_2, \dots r_n$, the time-dependent changes in metabolite concentrations $[\mathbf{S}] := ([S_1], S_2], \dots [S_m])$ can be described by the ODE system:

$$\frac{d[\mathbf{S}]}{dt} = \mathbf{N} \cdot \mathbf{v}([\mathbf{S}]) =: \mathbf{f}([\mathbf{S}]). \tag{2.1}$$

Here, $\mathbf{v}([\mathbf{S}]) := (v_1([\mathbf{S}]), v_2([\mathbf{S}]), \dots, v_n([\mathbf{S}]))$ describes the reaction velocities, which typically depend non-linearly on the concentrations. The stoichiometric matrix $\mathbf{N} \in \mathbb{R}^{m \times n}$ contains the molecularities of substrates and products of each reaction (Heinrich and Schuster, 1996). Consequently, the vector $\mathbf{f}([\mathbf{S}]) := (f_1([\mathbf{S}]), f_2([\mathbf{S}]), \dots f_m([\mathbf{S}]))$ summarizes the time-dependent changes of each metabolite in the network for a given set of concentrations and parameters.

A steady state is defined as a point $[\mathbf{S}]^*$ in the state space where the rate of production equals the rate of consumption for each metabolite. Hence, no net changes in the concentrations can occur and all reaction rates fulfil the mass balance equation

$$\frac{d[\mathbf{S}]}{dt} = \mathbf{N} \cdot \mathbf{v}([\mathbf{S}]^*) = \mathbf{0}. \tag{2.2}$$

Because all concentrations are fixed over time, a system can only leave a steady state in response to changes caused by external influences. Such influences can comprise, for example, the alteration of flux into the system as well as changes in enzyme concentrations. If the steady state is unstable, even small perturbations of this nature are sufficient to leave the steady state.

Depending on the underlying catalytic mechanisms, different rate laws have been developed to model enzyme catalysed reactions. One of the most wide-spread types of rate laws is the Michaelis Menten equation (Cornish-Bowden, 2004).

The Michaelis Menten equation for irreversible reactions

Given an irreversible reaction $r_{irrev} : S \rightarrow P$, the corresponding reaction rate is approximated by the equation

$$v_{irrev}([S]) = \frac{V_{max} \cdot [S]}{[S] + K_M} = \frac{V_{max} \cdot {}^{[S]}/_{K_M}}{1 + {}^{[S]}/_{K_M}}. \tag{2.3}$$

Here, V_{max} describes the maximum possible reaction rate, which is achieved when all enzymes are completely saturated by their substrate S_1. K_M is the Michaelis constant describing the substrate concentration at which half of the maximum velocity is achieved.

The Michaelis Menten equation for reversible reactions

A reversible reaction $r_{rev} : S \leftrightarrow P$ can be represented by the equation

$$v_{rev}([S],[P]) = \underbrace{\frac{V^+_{max} \cdot {}^{[S]}/_{K^{v^+}_M}}{1 + {}^{[S]}/_{K^{v^+}_M} + {}^{[P]}/_{K^{v^-}_M}}}_{v^+} - \underbrace{\frac{V^-_{max} \cdot {}^{[P]}/_{K^{v^-}_M}}{1 + {}^{[S]}/_{K^{v^+}_M} + {}^{[P]}/_{K^{v^-}_M}}}_{v^-}, \tag{2.4}$$

where S and P are the substrate and product concentrations, and v^+ and v^- indicate the forward and backward reaction rates with maximum velocities V^+_{max}, V^-_{max} and Michaelis constants $K^{v^+}_M$, $K^{v^-}_M$ (Cornish-Bowden, 2004).

2.2. Analysing local dynamic steady state properties

The response of the system in steady state to small perturbations depends on the asymptotic stability of the steady state. When it is asymptotically stable, a coordinated system response enables the return of concentrations and fluxes to the same values prior to the perturbation. If the steady state is unstable, such a return is not supported.

As explained by Heinrich and Schuster (1996), the trajectories of a perturbation $\delta[S_i]$ can be computed by the Taylor expansion of equation (2.2):

$$\frac{d(\delta[S_i])}{dt} = \sum_{k=1}^{m} \frac{\partial f_i}{\partial S_k}_{|[\mathbf{S}]=[\mathbf{S}]^*} \delta[S_k] + \frac{1}{2} \sum_{k,l=1}^{m} \frac{\partial^2}{\partial S_k \, \partial S_l} \delta S_k \delta S_l + \dots . \tag{2.5}$$

In the case of infinitesimally small perturbations, the quadratic and higher order terms can be neglected and equation (2.5) can be approximated by the linear differential equation system

$$\frac{d(\delta[S_i])}{dt} = \mathbf{J}_{[\mathbf{S}]^*} \cdot \delta[\mathbf{S}], \tag{2.6}$$

where $\mathbf{J}_{[\mathbf{S}]^*}$ is the Jacobian matrix evaluated in the steady state with elements

$$J_{i,k} := \frac{\partial f_i}{\partial [S_k]}_{|[\mathbf{S}]=[\mathbf{S}]^*} = \sum_{j=1}^{n} N_{i,j} \frac{\partial v_j}{\partial [S_k]}_{|[\mathbf{S}]=[\mathbf{S}]^*}. \tag{2.7}$$

The solutions of the linearised system (2.6) can be expressed as

$$\delta[S](t) = \sum_{i=1}^{m} c_i \mathbf{b}_i^{\lambda_i t}, \tag{2.8}$$

where $\mathbf{b}_i$ and λ_i are the eigenvectors and eigenvalues of $\mathbf{J}_{[\mathbf{S}]^*}$, and the constants c_i represent the initial perturbations (Heinrich and Schuster, 1996). Equation (2.8) shows that, only if the largest real part of the eigenvalues is negative, changes evoked by perturbations diminish over time and the steady state is asymptotically stable. The Jacobian matrix also indicates whether trajectories oscillate in the neighbourhood of a steady state. In particular, oscillations occur if the eigenvalues with the largest real parts form a complex conjugate pair (Makarov and Dong, 2001).

2.3. Metabolic control analysis

2.3.1. Flux control coefficients and elasticities

Flux control coefficients

Metabolic control analysis (MCA) (Fell, 1997) quantifies the control of an individual reaction r_j on a steady state flux F through the system by the flux control coefficient (FCC)

$$C_{e_j}^F := \frac{\partial \ln(F)}{\partial \ln(e_j)} = \frac{\partial F / F^*}{\partial e_j / e_j^*}. \tag{2.9}$$

Here, e_j is the concentration of the enzyme catalysing reaction r_j (steady state value indicated by *). FCCs describe the proportions by which control over a certain system flux is shared by the individual reactions. Consequently, all FCCs have to sum up to 1 for each flux. This is stated by the summation theorem

$$\sum_{j=1}^{n} C_{e_j}^F = 1. \tag{2.10}$$

Elasticities

In a steady state, the responses of the reaction rates to changes in any reactants or effectors are quantified by elasticity coefficients, or briefly 'elasticities'. Elasticities are defined by

$$\varepsilon_{S_i}^{v_j} := \frac{\partial \ln(v_j)}{\partial \ln([S_i])} = \frac{\partial v_j / v_j^*}{\partial [S_i] / [S_i]^*}, = \frac{\partial \mu_j}{\partial x_i}, \tag{2.11}$$

where $[S_i]^*$ is the steady state concentration of metabolite S_i and $v_j^* = v_j([\mathbf{S}]^*)$ is the rate of reaction r_j in the same steady state. $\mu_j := \frac{v_j}{v_j^*}$ and $x_i := \frac{[S_i]}{[S_i]^*}$ are defined as the reaction rates and concentrations in a system normalized to the steady state values. Elasticities are the partial derivatives of the reaction rates in this normalized system. Consequently, they describe the influence of changes in the normalized concentrations on the normalized reaction rates. In enzymatic reactions, this influence depends predominately on the amount of saturation of an enzyme with its metabolites.

The relationship between elasticities and FCCs is defined by the connectivity theorem

$$\sum_{j=1}^{n} C_{e_j}^F \cdot \varepsilon_{S_i}^{v_j} = 0, \quad i = 1, \ldots, m. \tag{2.12}$$

Deriving FCCs from elasticities

Using the simple paradigms of the summation and connectivity theorems, it is possible to uniquely identify all FCCs from the elasticities. Simple matrix methods for calculation of FCCs have been suggested by Fell and Sauro (1985) and Westerhoff and Kell (1987). Despite being easy to implement and intuitive to apply, these algorithms are restricted to specific predefined pathway topologies like linear pathways or cycles, and not generalizable to arbitrary networks. The elas-

ticities used and the FCCs produced by these methods refer to net reactions (i.e. summarizing forward and backward direction), which corresponds to the nature of experimentally accessible data.

Wang *et al* (2004) describe an alternative approach that enables the calculation of FCCs for networks of any structure and size using the following equation:

$$\mathbf{C} = -\mathbf{E} \cdot (\mathbf{N} \cdot \mathbf{V} \cdot \mathbf{E})^{-1} \cdot (\mathbf{N} \cdot \mathbf{V} \cdot \mathbf{\Pi}) + \mathbf{\Pi}, \tag{2.13}$$

where, $\mathbf{E} \in \mathbb{R}^{2n \times m}$ is matrix of elasticities, with separate values for forward and reverse reactions, $\mathbf{N} \in \mathbb{R}^{m \times 2n}$ is the corresponding stoichiometric matrix, and $\mathbf{V} \in \mathbb{R}^{2n \times 2n}$ is a matrix containing the steady state reaction rates of forward and reverse reactions on the diagonal. $\mathbf{\Pi} \in \mathbb{R}^{2n \times m}$ contains the enzyme-parameter elasticities, which describe the relative impact of changing an enzyme's concentration on the rates of its catalysed forward and reverse reaction rates. Typically, this relationship is assumed to be linear and the corresponding entries are set to 1. The matrix $\mathbf{C} \in \mathbb{R}^{2n \times n}$ contains the computed FCCS. Its entries $C^{v_i}_{E_j}$ describe the impact of perturbing the concentration of every enzyme E_j on each steady state reaction rate $v_i^{+/-}$, with reactions being split into forward and backward directions.

Summarizing FCCs for forward and reverse reactions

The FCCs computed by equation (2.13) describe the impact of all enzymes on all steady state reaction rates, separated by forward and backward directions. This high level of detail leads to an exceedingly larger number of FCCs than produced by the method of Fell and Sauro (1985) or Westerhoff and Kell (1987), making it difficult to get an impression of the control exhibited by an enzyme on the net steady state flux through a reaction. For better interpretability, FCCs can therefore be summarized over forward and reverse directions. Each entry of $\mathbf{C}$ then describes the impact of perturbing the concentration of one enzyme on the net flux through each reaction. Using definition (2.9), this can be done in the following way:

$$C^{v_i}_{v_j} = \frac{\partial \ln (v_i^+ - v_i^-)}{\partial \ln (e_j)} \tag{2.14}$$

$$= \frac{e_j^*}{v_i^{+*} - v_i^{-*}} \cdot \frac{\partial (v_i^+ - v_i^-)}{\partial e_j} = \frac{e_j^*}{v_i^{+*} - v_i^{-*}} \cdot \left(\frac{\partial v_i^+}{\partial e_j} - \frac{\partial v_i^-}{\partial e_j} \right). \tag{2.15}$$

Because of

$$\frac{\partial \ln(v)}{\partial \ln(e)} = \frac{e^*}{v*} \cdot \frac{\partial v}{\partial e} \tag{2.16}$$

$$\Leftrightarrow \frac{\partial v}{\partial e} = \frac{\partial \ln(v)}{\partial \ln(e)} \cdot \frac{v^*}{e^*} = C_e^v \cdot \frac{v*}{e^*}, \tag{2.17}$$

we can rewrite equation (2.15) as

$$C_{e_j}^{v_i} = \frac{e_j^*}{v_i^{+*} - v_i^{-*}} \cdot \left(C_{e_j}^{v_i^+} \cdot \frac{v_i^{+*}}{e_j^*} - C_{e_j}^{v_i^-} \cdot \frac{v_i^{-*}}{e_j^*} \right) \tag{2.18}$$

$$= \frac{C_{e_j}^{v_i} \cdot v_i^{+*} - C_{e_j}^{v_i^-} \cdot v_i^{-*}}{v_i^{+*} - v_i^{-*}}, \tag{2.19}$$

where v_i^{+*} and v_i^{-*} refer to the steady state fluxes through reactions r_i^+ and r_i^-.

2.3.2. Relationship between elasticities and enzyme kinetic parameters

Irreversible Michaelis Menten kinetics

If a reaction rate $v([S])$ follows irreversible Michaelis Menten kinetics, the corresponding normalized rate law $\mu := \frac{v}{v^*}$ can be derived as

$$\mu(x) = \frac{v([S])}{v([S]^*)} = \frac{v(x \cdot [S]^*)}{v([S]^*)} \tag{2.20}$$

$$= \frac{V_{max} \cdot (x \cdot [S]^*)}{(x \cdot [S]^*) + K_M^v} \cdot \frac{[S]^* + K_M^v}{V_{max} \cdot [S]^*} \tag{2.21}$$

$$= x \frac{[S]^* + K_M^v}{(x \cdot [S]^*) + K_M^v} \tag{2.22}$$

The derivative with respect to the normalized substrate x is then given by

$$\frac{\partial \mu}{\partial x} = \frac{[S]^* + K_M^v}{x \cdot [S]^* + K_M^v} - \frac{x \cdot [S]^*([S]^* + K_M^v)}{(x \cdot [S]^* + K_M^v)^2} \tag{2.23}$$

Evaluation of the derivative at the steady state (indicated by $x = 1$) provides

the substrate elasticity

$$\varepsilon_S^v := \frac{\partial \mu}{\partial x}_{|x=1} \tag{2.24}$$

$$= \frac{[S]^* + K_M^v}{S^* + K_M^v} - \frac{[S]^*([S]^* + K_M^v)}{([S]^* + K_M^v)^2} \tag{2.25}$$

$$= 1 - \frac{[S]^*}{[S]^* + K_M^v} \tag{2.26}$$

$$= \frac{1}{1 + {}^{[S]^*}/_{K_M^v}} \in (0, 1] \tag{2.27}$$

Reversely, the original kinetic parameters can be computed from the elasticities for a given steady state by

$$K_M^v = \frac{[S]^* \cdot \varepsilon_S^v}{1 - \varepsilon_S^v} \tag{2.28}$$

$$V_{max} = v^* \cdot (1 + \frac{K_M^v}{[S]^*}), \tag{2.29}$$

where v^* and $[S]^*$ indicate the reaction rate and substrate concentration in steady state.

Reversible Michaelis Menten kinetics

Repeating the procedure described above for the reversible Michaelis Menten kinetics introduced in equation (2.4) reveals the following equations for the elasticities:

$$\varepsilon_S^{v^+} = \frac{1 + \frac{[P]}{K_M^{v^-}}}{1 + \frac{[S]}{K_M^{v^+}} + \frac{[P]}{K_M^{v^-}}} \in (0, 1] \tag{2.30}$$

$$\varepsilon_P^{v^-} = \frac{1 + \frac{[S]}{K_M^{v^+}}}{1 + \frac{[S]}{K_M^{v^+}} + \frac{[P]}{K_M^{v^-}}} \in (0, 1] \tag{2.31}$$

$$\varepsilon_P^{v^+} = \varepsilon_P^{v^-} - 1 \in (\text{-}1, 0] \tag{2.32}$$

$$\varepsilon_S^{v^-} = \varepsilon_S^{v^+} - 1 \in (\text{-}1, 0] \tag{2.33}$$

The relationship to the original kinetic parameters for a given steady state looks as follows:

$$K_M^{v^+} = [S]^* \cdot \frac{1 - \varepsilon_S^{v^+} - \varepsilon_P^{v^-}}{\varepsilon_S^{v^+} - 1} \tag{2.34}$$

$$K_M^{v^-} = [P]^* \cdot \frac{1 - \varepsilon_S^{v^+} - \varepsilon_P^{v^-}}{\varepsilon_P^{v^-} - 1} \tag{2.35}$$

$$V_{max}^{v^+} = v^+([S]^*, [P]^*) \cdot \frac{1 + \frac{[S]^*}{K_M^{v^+}} + \frac{[P]^*}{K_M^{v^-}}}{\frac{[S]^*}{K_M^{v^+}}} \tag{2.36}$$

$$V_{max}^{v^-} = v^-([S]^*, [P]^*) \cdot \frac{1 + \frac{[S]^*}{K_M^{v^+}} + \frac{[P]^*}{K_M^{v^-}}}{\frac{[P]^*}{K_M^{v^-}}} \tag{2.37}$$

A closer inspection of equations (2.34) and (2.35) shows that $K_M^{v^+}$ and $K_M^{v^-}$ become negative for

$$\varepsilon_S^{v^+} + \varepsilon_P^{v^-} \leqslant 1. \tag{2.38}$$

Because Michaelis constants are given in units of concentrations and, as such, are required to be non-negative, elasticity combinations not fulfilling equation (2.38) are biologically infeasible for any given steady state.

Multiple substrates and products

Reactions with multiple substrates and products require more complex rate laws than the Michaelis Menten equation. Because of this complexity, each elasticity depends on a distinctly larger number of kinetic parameters. Consequently, expressing kinetic parameters in terms of the elasticities in a manner similar to that shown in equations (2.34)-(2.35) is an analytically challenging task.

For example, reactions with random order binding of their substrates and products are among the simplest to model mathematically (Cornish-Bowden, 2004). The rate of such a reaction

$$A + B \Leftrightarrow C + D$$

can be calculated using the following equation:

$$v = \frac{\frac{V_{max}^{v^+}[A][B]}{K_{iA}K_{mB}} - \frac{V_{max}^{v^-}[P][Q]}{K_{mP}K_{iQ}}}{1 + \frac{[A]}{K_{iA}} + \frac{[B]}{K_{iB}} + \frac{[P]}{K_{iP}} + \frac{[Q]}{K_{iQ}} + \frac{[A][P]}{K_{iA}K_{mB}} + \frac{[P][Q]}{K_{mP}+K_{iQ}}}. \tag{2.39}$$

Equation (2.39) comprises eight kinetic parameters, which can be used to cal-

culate eight elasticities. Each elasticity then describes the influence of one of the four reactants on the forward or backward reaction. When trying to derive terms that express the kinetic parameters in terms of the eight elasticities, it is necessary to solve a system of eight nonlinear equations. An attempt to solve such a system using a trial version of the MATHEMATICA language (Wolfram Research, 2012) was impossible because the program failed to converge towards an analytical solution. Consequently, a rule similar to equation (2.38) is not available to us in the same straight-forward manner.

2.4. The principles of structural kinetic modelling (SKM)

2.4.1. Computing the Jacobian matrix from elasticities

As described in Section 2.2, local dynamic properties of a steady state, like stability or oscillatory behaviour, can be derived from the Jacobian matrix evaluated in the steady state. Computation of the Jacobian matrix typically requires knowledge of all enzyme kinetic rate laws and kinetic parameters describing the reactions in the system in order to compute the partial derivatives $\frac{\partial v_j}{\partial [S_i]}$.

However, structural kinetic modelling (SKM) enables the computation of the Jacobian matrix without relying on such knowledge. Instead, it uses the following rearrangement of equation (2.7):

$$J_{i,k} = \sum_{j=1}^{n} N_{i,j} \frac{\partial v_i}{\partial [S_k]}_{|[\mathbf{S}]=[\mathbf{S}]^*} = \sum_{j=1}^{n} \frac{N_{i,j} v_j^*}{[S_k]^*} \cdot \frac{[S_k]^*}{v_j^*} \frac{\partial v_i}{\partial [S_k]}_{|[\mathbf{S}]=[\mathbf{S}]^*} \tag{2.40}$$

$$= \sum_{j=1}^{n} \lambda_{i,j} \cdot \varepsilon_{S_k}^{v_j}, \tag{2.41}$$

with $\lambda_{i,j} := N_{i,j} \cdot \frac{v_j^*}{[S_i]^*}$ and elasticities $\varepsilon_{S_k}^{v_j}$ as defined in Section 2.3.

The Jacobian matrix can thus be computed by the simple matrix multiplication

$$\mathbf{J}_{[\mathbf{S}]^*} = \mathbf{\Lambda} \cdot \mathbf{E}, \tag{2.42}$$

where the matrix $\mathbf{E}$ contains the elasticities $\varepsilon_{S_i}^{v_j},\ j = 1 \ldots, n,\ i = 1 \ldots, m$, and $\mathbf{\Lambda} \in \mathbb{R}^{m,n}$ is the matrix of normalized stoichiometric coefficients $\lambda_{i,j}$.

As already explained in Section 2.3, elasticities associated with enzyme-catalysed

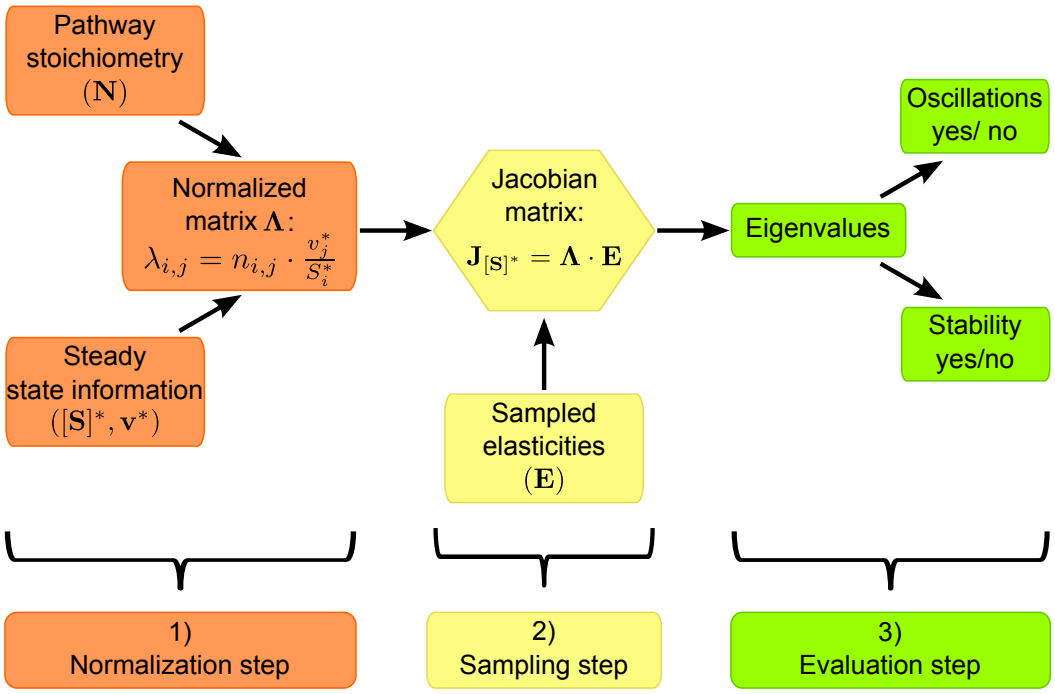


Figure 2.1.: The principles of SKM. SKM experiments can be summarized by the following steps: 1) normalization of the stoichiometric matrix with respect to steady state fluxes and concentrations; 2) random sampling of the elasticities and computation of the Jacobian matrices; 3) evaluation of the Jacobian matrix for each sampled set of elasticities with respect to local dynamic steady state properties.

reactions indicate the amount of saturation of the enzyme with a particular metabolite. Because of this implication, the sampled elasticities in SKM experiments are also called 'saturation parameters' (Steuer *et al*, 2006). In particular, the larger the absolute value of an elasticity, the less the enzyme catalysing the reaction is saturated with its metabolite (Wang *et al*, 2004). In this work, however, I will refer to the sampled variables as elasticities to emphasize their proximity to the MCA framework.

2.4.2. Monte Carlo sampling of structural kinetic models

As demonstrated by equation (2.42), the Jacobian matrix for a given steady state, for which experimental measurements are available, can be derived solely from a set of model parameters (elasticities) $\mathbf{E} \in \mathbb{R}^{n,m}$, the stoichiometric matrix $\mathbf{N}$, and the steady state measurements $[\mathbf{S}]^*$ and $\mathbf{v}^*$. While the stoichiometry, the steady state concentrations and the fluxes are experimentally accessible, the elasticities are often unknown in practice. However, due to the normalization step, they are

restricted to pre-defined intervals, from which they can be sampled in a Monte Carlo approach. This enables the creation of a large number of models followed by the exploration of the parameter space to detect regions associated with stability or instability. An overview of the SKM approach is shown in Figure 2.1.

The interval boundaries are chosen according to the type of kinetics employed by an enzyme. For example, if a reaction follows reversible Michaelis Menten kinetics, its substrate-associated elasticities are restricted to $(0, 1]$, while its product-associated elasticities lie within $(-1, 0]$ (see Section 2.3.2 for details). Further examples for other types of rate laws can be found in the supplementary information of Steuer *et al* (2006).

2.4.3. Evaluating SKM experiments

Strategies for analysing SKM experiments can be categorised into quantitative and qualitative approaches (Girbig *et al*, 2012b).

Quantitative analysis

Quantitative strategies count the proportions of stable and unstable models and derive conclusions about the general tendency of the system towards stability. Possible applications are the comparison of stability tendencies under different experimental conditions, or the assessment of the effect of certain types of metabolic regulators on stability. An example that investigates the impact of such regulators on the stability of the Calvin-Benson cycle (CBC) is given by Girbig *et al* (2012a). This approach can also serve for detecting the different types of dynamical behaviour that could emerge in the system given a suitable set of elasticities. Such a behaviour could be, for example, the presence of oscillations, which are indicated by complex conjugated eigenvalues (Section 2.2). For example, Steuer *et al* (2006) present a detailed analysis of the possible dynamics of a simple kinetic model of the CBC. They find that 5.7% of the sampled SK-models are unstable. The largest proportion of these unstable models gives rise to only one positive eigenvalue each, indicating non-oscillatory trajectories around the steady state. However, 0.6% exhibit two or more non-negative real parts, which hints at the possibility of more complex system responses.

Qualitative univariate analysis

Qualitative strategies aim at analysing the specific conditions responsible for stability or instability. If an elasticity parameter contributes to a dynamic property like stability, we can assume that this elasticity is distributed differently among those SK-models exhibiting the property and those models that do not. Therefore, the simplest approaches for detecting such elasticities are univariate in nature, which means that they compare the elasticity values leading to stable or unstable states in a pairwise manner. For example, Grimbs *et al* (2007) perform distribution comparisons using the Kolmogorov-Smirnoff test or correlation coefficients in order to detect the enzymes with most effect on the stability of Erythrocyte metabolism.

Qualitative multivariate analysis

SKM enables the identification of stable and unstable as well as oscillatory and non-oscillatory steady states for a large number of randomly generated parameter sets. When searching for model parameters with high discriminatory power between different dynamic steady state properties, previous SKM analyses used univariate comparisons in order to detect single enzymes that play 'key roles' in maintaining the stability of a steady state. However, it is a well-established fact that metabolic control is not conducted by few key enzyme alone. Instead, changes in flux distributions can also be caused by the joint orchestration of several enzymes (Kacser and Porteous, 1987). Such orchestration of enzyme activities cannot be detected by univariate tests that simply compare the distributions of elasticities between stable and unstable models. Instead, methods based on supervised classification can help identify such patterns in the elasticity space (Girbig *et al*, 2012a).

This thesis aims at exploring the possibilities and possible limitations of this approach in greater detail. Furthermore, it will be investigated whether patterns can also be derived in terms of FCCs and the performance and interpretability of such FCC patterns will be compared to those based on elasticities.

2.5. Supervised classification approaches to evaluate SKM experiments

In supervised classification, our goal is to derive an algorithm that can predict a set of discrete target values $\mathbf{T} = \{t^{(1)}, t^{(2)}, \ldots, t^{(q)}\}$ from a set of training data points $\mathbf{X} = \{\chi^{(1)}, \chi^{(2)}, \ldots, \chi^{(q)}\}$. Typically, the training data are real-valued vectors consisting of p features ($\chi^{(i)} \in \mathbb{R}^p$). After training, the derived decision function needs to be assessed using previously unknown test data points. This provides information about the generalizability of the classifier to new cases. If a classifier performs well on the training data, but poorly on the test data, it was overfitted (Rokach and Maimon, 2008, p. 49). In general, overfitting can be reduced by choosing sufficiently large training data sets that enable the classifier to derive robust classification rules during training.

In the given scenario, we aim at predicting information about the stability and oscillatory properties of a steady state based on a set of elasticities or FCCs created by an SKM experiment. After construction of the classifier, we want to be able to interpret the derived classification rules. Two algorithms lend themselves to this task because, in addition to often achieving good classification results, their derived decision functions are also easily accessible for manual interpretation. These two algorithms are called (1) decision trees and (2) relevance vector machines (RVMs). This work will explore the potential of both approaches for the evaluation of SKM experiments.

2.5.1. Decision trees

Decision trees offer an intuitive algorithm that classifies a data point by establishing a series of quantitative or qualitative criteria on its features. If all criteria are fulfilled, a particular class label is assigned. Each series of criteria is represented by a path through a tree, in which each node contains a criterion, and each leaf contains the class label associated with the path. Consequently, the decision tree can offer a complex and nonlinear representation of the class boundaries in the feature space. Different implementations exist offering various algorithms of constructing the tree (Rokach and Maimon, 2005).

One possible way to construct decision trees is the `C5.0` algorithm (Quinlan, 2013), which is a commercial version of the `C4.5` algorithm (Quinlan, 1998) with increased speed and memory efficiency that makes it well applicable for large num-

bers of training samples. One interesting feature of this algorithm is the possibility to create 'rulesets' that summarize the derived conditions for each class in an easily interpretable manner. In contrast to the classical decision tree structure, where the abundance of a feature depends on its position in the tree (for example, the feature in the root is always used), features in rulesets can be mutually exclusive. This motivates their use in finding diverse combinations of features important for different dynamic properties. The discovered rulesets describe patterns imposing coordinated thresholds on the saturation of different enzymes in the system.

2.5.2. Relevance vector machines (RVMs)

In RVMs (Tipping, 2001, Tipping *et al*, 2003) a data point χ is classified by computing its probability of belonging to a particular class $t \in \{0, 1\}$ as

$$P(t) = \left(\frac{1}{1 + e^{y(\chi)}}\right)^{t} \cdot \left(1 - \frac{1}{1 + e^{y(\chi)}}\right)^{1-t}, \tag{2.43}$$

with $y(\chi) = \sum_{i=1}^{q} \omega_i \cdot \Phi(\chi, \chi^{(i)})$. $\Phi(\chi, \chi^{(i)})$ is the value of a kernel function that provides a measure of the distance between the data point with unknown class label χ, and each of the training samples $\chi^{(i)}$. Similar to support vector machines (Schölkopf *et al*, 1999), the kernel is intended to project the data points into higher dimensional space in which they are more easily separable.

One of the most prominent features of RVMs is their ability to obtain a sparse solution in which most of the weights ω_i become infinitesimally small. This solution is obtained using a Bayesian approach that imposes a prior distribution on the weights so that their expected values are zero (Gopinath *et al*, 2013). This ensures that during model fitting, most weights will vanish. Consequently, prediction is only based on a small set of weights and their associated training vectors. Because these vectors have been determined as the most informative for prediction, they are also termed 'relevance vectors' (RVs).

The RVM algorithm has several advantages compared to other kernel based classifiers like SVMs. The prediction can take place within comparably short runtime because it only requires the computation of the few kernel values with non-zero weights. Furthermore, the decision function is easier to interpret because the RVs represent typical instances of each class. In contrast, the decision boundaries detected by SVMs often vary only little between both classes because of the proximity of the vectors representing the boundaries on both sides of the margin. A

third advantage is that the sparsity of the solution avoids an over-adaptation to the training data characteristics, which makes the classifier very robust against overfitting (Weiss and Ellis, 2006; Li *et al*, 2012; Gopinath *et al*, 2013).

3. Model generation and analysis

3.1. Model generation

In order to systematically evaluate the machine learning approaches analysed in this work, three simple artificial example models were constructed. For each of these example pathways, a kinetic model with known rate laws and kinetic parameters was created. These kinetic models were used to compute steady state concentrations and fluxes for structural kinetic modelling (SKM). Applying SKM to a kinetic model for which the 'true kinetics' are already well-defined may seem counter-intuitive at first, since one of the most distinctive advantage of SKM refers to its ability to analyse system dynamics if detailed kinetic information is not available. However, for method development this approach yields several advantages:

1. By defining small artificial pathways, we can assess the impact of changes in pathway topology on system dynamics and classification results. For this purpose, three distinct pathway structures were constructed: a linear, a branched, and a cyclic pathway.

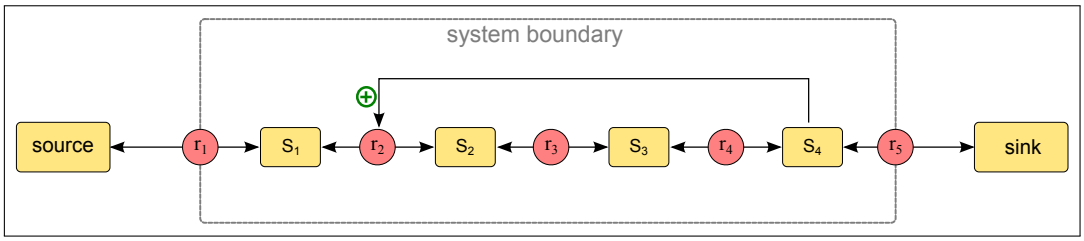
2. We can compare the derived elasticity patterns associated with certain dynamic steady state properties to the true elasticities derived from the kinetic model.

3. The statistical exploration of the elasticities can elucidate the system's response to possible mutations that would be time consuming with simulations based on the kinetic model.

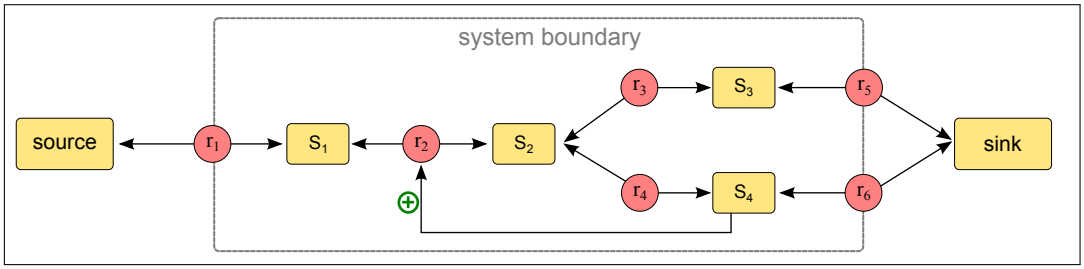
In the first part of this chapter, all kinetic models used within this work will be introduced. Afterwards, the procedure for constructing SK-models based on these steady states will be described. Finally, these SK-models will be statistically analysed in a univariate an multivariate manner.

3.1.1. Illustrative example pathways

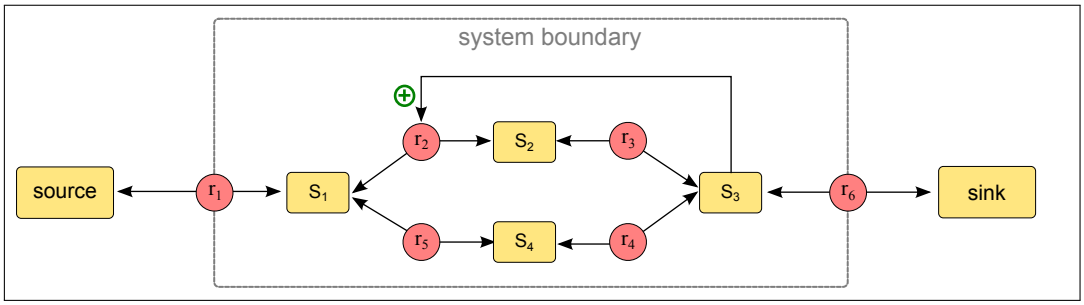
In order to account for the impact of pathway topology on the dynamic properties of steady states of a metabolic network, kinetic models were constructed using three different pathway structures (Figure 3.1).



(a) Linear pathway



(b) Branched pathway



(c) Cyclic pathway

Figure 3.1.: Pathway topologies underlying the small example models used for SKM analysis.

Kinetic parameters and rate laws were chosen as uniformly as possible in order to eliminate their influence on the outcome of the analysis. All reactions were modelled using reversible Michaelis Menten kinetics. A positive feedback term was included to increase the range of possible dynamic behaviours. Tables 3.1 and 3.2 describe the resulting differential equation systems underlying each pathway. Suitable values for the kinetic parameters were identified by a grid search with the aim of enabling the emergence of both unstable and oscillatory steady states in

each model. The kinetic parameters were chosen in such a way that the resulting steady states would enable the emergence of instabilities when analysed by SKM. For this purpose, a preliminary study was performed for each pathway. In doing so, a large number of possible combinations of kinetic parameters were selected iteratively. For each parameter set, the corresponding steady state was computed and the probabilities of stability and oscillations was estimated by a small Monte Carlo simulation using 1000 iterations. An overview of the resulting parameter values used for model creation is given in Table 3.3. The Michaelis constants of the backward reactions were chosen to be substantially lower than those of the forward reactions in order to ensure faster forward than backward rates, and, consequently, a positive net flux through each reaction.

$$\frac{d[S_1]}{dt} = \left[v_1^+ - v_1^-\right] - \left[v_2^+ - v_2^-\right]$$
$$\frac{d[S_2]}{dt} = \left[v_2^+ - v_2^-\right] - \left[v_3^+ - v_3^-\right]$$
$$\frac{d[S_3]}{dt} = \left[v_3^+ - v_3^-\right] - \left[v_4^+ - v_4^-\right]$$
$$\frac{d[S_4]}{dt} = \left[v_4^+ - v_4^-\right] - \left[v_5^+ - v_5^-\right]$$

(a) Linear pathway

$$\frac{d[S_1]}{dt} = \left[v_1^+ - v_1^-\right] - \left[v_2^+ - v_2^-\right]$$
$$\frac{d[S_2]}{dt} = \left[v_2^+ - v_2^-\right] - \left[v_3^+ - v_3^-\right] - \left[v_4^+ - v_4^-\right]$$
$$\frac{d[S_3]}{dt} = \left[v_3^+ - v_3^-\right] - \left[v_5^+ - v_5^-\right]$$
$$\frac{d[S_4]}{dt} = \left[v_4^+ - v_4^-\right] - \left[v_6^+ - v_6^-\right]$$

(b) Branched pathway

$$\frac{d[S_1]}{dt} = \left[v_1^+ - v_1^-\right] - \left[v_2^+ - v_2^-\right] + \left[v_5^+ - v_5^-\right]$$
$$\frac{d[S_2]}{dt} = \left[v_2^+ - v_2^-\right] - \left[v_3^+ - v_3^-\right]$$
$$\frac{d[S_3]}{dt} = \left[v_3^+ - v_3^-\right] - \left[v_4^+ - v_4^-\right] - \left[v_6^+ - v_6^-\right]$$
$$\frac{d[S_4]}{dt} = \left[v_4^+ - v_4^-\right] - \left[v_5^+ - v_5^-\right]$$

(c) Cyclic pathway

Table 3.1.: ODE systems used to model the pathways depicted in Figure 3.1.

All three kinetic models were implemented in MATLAB (MATLAB, 2010) using the *Systems Biology Toolbox 2 for Matlab* (Schmidt, 2013; Schmidt and Jirstrand, 2006). Steady states were computed using the *SBsteadystate* function of the Toolbox. This function attempts to solve the mass balance equation (2.2) by numerical optimization based on the Newton algorithm. Initial conditions were set to 1 $\frac{\text{mol}}{\text{l}}$ for each metabolite and pathway.

Construction of SK-models SK-models were constructed using the *MATLAB Toolbox for Structural Kinetic Modeling* (Girbig *et al*, 2012b). This toolbox, which

$v^+([S],[P])$	$\frac{V_{max}\cdot\frac{[S]}{K_M^+}}{1+\frac{[S]}{K_M^+}+\frac{[P]}{K_M^-}}$	Rate law of reaction $r_1^+, r_3^+, r_4^+, r_5^+$ and r_6^+.
$v^-([S],[P])$	$\frac{V_{max}\cdot\frac{[P]}{K_M^-}}{1+\frac{[S]}{K_M^+}+\frac{[P]}{K_M^-}}$	Rate law of reaction $r_1^-, r_3^-, r_4^-, r_5^-$ and r_6^-.
$v_2^+([S],[P],[A])$	$\frac{V_{max}\cdot\frac{S}{K_M^+}}{(1+\frac{K_A}{[A]})(1+\frac{[S]}{K_M^+}+\frac{[P]}{K_M^-})}$	Rate law of reaction r_2^+ including a mixed activation term
$v_2^-([S],[P],[A])$	$\frac{V_{max}\cdot\frac{[P]}{K_M^-}}{(1+\frac{K_A}{[A]})(1+\frac{[S]}{K_M^+}+\frac{[P]}{K_M^-})}$	Rate law of reaction r_2^- including a mixed activation term

Table 3.2.: Kinetic rate laws used for the ODE-models described in Table 3.1. Reactions were modelled using reversible Michaelis-Menten kinetics (Bisswanger, 2008). A positive feedback term representing mixed activation was included for reaction v_2 (Cornish-Bowden, 2004).

K_M^+	3 $\frac{mol}{l}$	Michaelis constant of the forward reactions
K_M^-	30 $\frac{mol}{l}$	Michaelis constant of the backward reactions
$V_{max}^{v^+/-}$	1 $\frac{mol}{l\cdot sec}$	Maximum forward/backward velocity of each reaction
K_A	1 $\frac{mol}{l}$	Activation constant of reactions r_2^+ and r_2^-
source	1 $\frac{mol}{l}$	external constant source metabolite
sink	1 $\frac{mol}{l}$	external constant sink metabolite

Table 3.3.: Parameters used for creating kinetic models of the small example systems in Figure 3.1.

is described in detail in the subsequent chapter of this work, enables the straightforward construction of SK-models for a given steady state.

Elasticities were sampled from a set of 100 discrete and evenly spaced values that ranged from 0.01 to 1 in steps of 0.01. The goal of the discretization was to make the predicted decision tree patterns better reproducible over different trees.

Until now, SKM experiments described in the literature did not take into account whether the sampled elasticities were associated with biologically feasible, i.e. non-negative, kinetic parameters (see Section 2.3.2 for details). In order to assess the proportion of biologically infeasible models, as well as their impact on the results of the experiment, the sampling procedure was performed twice: once with and once without filtering out those elasticity combinations that would result in biologically

infeasible models. For biologically feasible models, the following filtering criterion had to be fulfilled for all reactions r_j:

$$\varepsilon_S^{v_j^+} + \varepsilon_P^{v_j^-} > 1 \tag{3.1}$$

Here, $\varepsilon_S^{v_j^+}$ is the elasticity describing the substrate influence on the forward reaction v_j^+, and $\varepsilon_P^{v_j^+}$ describes the corresponding influence of product perturbations on the reverse reaction v_j^-.

For each set of biologically feasible elasticities, a complementary set of flux control coefficients was computed using equation (2.13).

3.1.2. SK-models of mitochondrial energy metabolism

The TCA cycle plays a central role in energy metabolism and in providing precursors for biosynthetic pathways. Consequently, this pathway relies on mechanisms that guarantee a stable flux through the cycle despite the perturbations occurring at the branch points to adjacent pathways. Because of its importance for cell metabolism, the intrinsic stabilizing mechanisms of the TCA cycle and the adjacent reactions for the respiratory chain and adenosine triphosphate (ATP) synthesis were investigated by SKM. After sampling, the elasticities were analysed by univariate comparisons, decision trees and RVMs. This analysis was intended to serve as an example of how these methods perform, and what types of information they can provide, when applied to a metabolic system of realistic size and complexity.

Kinetic model

The kinetic model by Berndt *et al* (2012) served as a reference for network stoichiometry and steady-state information. This model describes the electron transfer reactions taking place in the respiratory chain complexes I and III in great detail. As a result, it contains 96 state variables for the different redox states of complex I and 48 state variables for complex III.

Steady state computation

Steady state concentrations and fluxes were computed by numerical integration using the `ode15s` solver. In each integration step, reaction rates were computed

using the MATLAB implementation of Berndt *et al* (2012). Initial values were also provided by the authors in their original code. Integration was performed over 1000 seconds until all concentration changes were below $2.5 \cdot 10^{-8}$ $\frac{mol}{l}$.

Two steady states were computed for SKM analysis which represented different degrees of workload imposed on the cell. In the neuron, an increased work load implies an increase in ATP demand by the cell in order to restore the membrane potential after neuron triggering. The first state (also called reference state) represented a situation where cytosolic O_2 consumption was 50% of its maximum value. This corresponded to a moderate work amount of ATP consumption by the cell (Berndt *et al*, 2012). In contrast, the second state resembled a phenomenon called 'gamma oscillations' (Fell and Axmacher, 2011; Singer, 2013; Brittain and Brown, 2014; Hanslmayr and Staudigl, 2014), which represented a scenario of strong workload, resulting in increased consumption of ATP and O_2. This steady state was computed by setting cytosolic O_2 consumption to 90% of the maximum value.

Construction of SK-Models

The stoichiometric matrix of mitochondrial energy metabolism was constructed according to ODE system in Table 3.4. Instead of including all detailed electron transfer reactions of the kinetic model, complex I and III were reduced to one reaction each. More precisely, complex I was represented by the rate of flavin reduction, and complex III was represented by the rate of cytochrome C (CytC) reduction in the original kinetic model. All 14 regulatory interactions listed by Nelson and Cox (2004) for the TCA cycle (Table 3.5) were taken into account by including elasticities that represented the different types of feedback. Those elasticities representing activating influences were sampled from the interval $(0, 1]$ and those for inhibitors were sampled from [-1, 0). The resulting SK-model covered 24 metabolites, 20 reactions and 71 elasticities.

Elasticities of the reactions with a single substrate and product (Aconitase (ACON), Fumarase (FU) and H^+-Leak (H-leak)) were filtered for biological feasibility using the criterion derived in equation (3.1). In the remaining models, the FCCs were checked in order to detect whether the effect of pyruvate import on pyruvate dehydrogenase (PDH) was positive. This criterion was used as an additional indicator of the biological plausibility of the sampled models. Only the remaining models after both filtering steps were used for the subsequent univariate and multivariate analyses.

$$
\begin{aligned}
\frac{d[PYR_m]}{dt} &= v_{PYR/H\text{-}sym} - v_{PDH} \\
\frac{d[H_m]}{dt} &= v_{P/H\text{-}sym} + v_{PYR/H\text{-}sym} - v_{K/H\text{-}anti} - v_{Na/H\text{-}anti} + v_{H\text{-}leak} + \\
&\quad v_{CITS} - 5\ v_{CI} - 2\ v_{CIII} - 4\ v_{CIV} + 3\ v_{ATP\text{-}syn} \\
\frac{d[ACCOA_m]}{dt} &= v_{PDH} - v_{CITS} \\
\frac{d[CIT_m]}{dt} &= v_{CITS} - v_{ACON} \\
\frac{d[ICIT_m]}{dt} &= v_{ACON} - v_{IDH} \\
\frac{d[AKG_m]}{dt} &= v_{IDH} - v_{AKGDH} \\
\frac{d[SUCCOA_m]}{dt} &= v_{AKGDH} - v_{SCS} \\
\frac{d[SUCC_m]}{dt} &= v_{SCS} - v_{SDH} \\
\frac{d[FUM_m]}{dt} &= v_{SDH} - v_{FU} \\
\frac{d[MAL_m]}{dt} &= v_{FU} - v_{MDH} \\
\frac{d[OXA_m]}{dt} &= v_{MDH} - v_{CITS} \\
\frac{d_t[COA_m]}{dt} &= v_{CITS} - v_{PDH} + v_{SCS} - v_{AKGDH} \\
\frac{d[NADH_m]}{dt} &= v_{PDH} + v_{IDH} + v_{AKGDH} + v_{MDH} - v_{CI} \\
\frac{d[NAD_m]}{dt} &= -v_{PDH} - v_{IDH} - v_{AKGDH} - v_{MDH} + v_{CI} \\
\frac{d[QH_2]}{dt} &= v_{SDH} + v_{CI} - v_{CIII} \\
\frac{d[Q]}{dt} &= -v_{SDH} - v_{CI} + v_{CIII} \\
\frac{d[CytC^{(red)}]}{dt} &= 2\ v_{CIII} - 2\ v_{CIV} \\
\frac{d[CytC^{(ox)}]}{dt} &= -2\ v_{CIII} + 2\ v_{CIV} \\
\frac{d[P_m]}{dt} &= v_{P/H\text{-}sym} - v_{SCS} - v_{ATP\text{-}syn} \\
\frac{d[ADP_m]}{dt} &= v_{ATP/ADP\text{-}anti} - v_{SCS} - v_{ATP\text{-}syn} \\
\frac{d[ATP_m]}{dt} &= -v_{ATP/ADP\text{-}anti} + v_{SCS} + v_{ATP\text{-}syn} \\
\frac{d[P_c]}{dt} &= -v_{P/H\text{-}sym} + v_{ATPuse} \\
\frac{d[ADP_c]}{dt} &= -v_{ATP/ADP\text{-}anti} + v_{ATPuse} \\
\frac{d[ATP_c]}{dt} &= v_{ATP/ADP\text{-}anti} - v_{ATPuse}
\end{aligned}
$$

Table 3.4.: ODE system describing the stoichiometry used for the SK-model of mitochondrial energy metabolism. The subcellular location of a metabolite X is indicated by the subscripts X_m (mitochondrial matrix) and X_c (cytosol). All abbreviations are explained in Appendix A.

Reaction	Metabolite	Type of regulation
PDH	ATP_m	I
PDH	$ACCOA_m$	I
PDH	$NADH_m$	I
PDH	COA_m	A
PDH	NAD_m	A
CITS	$SUCCOA_m$	I
CITS	$NADH_m$	I
CITS	CIT_m	I
CITS	ATP_m	I
CITS	ADP_m	A
IDH	ATP_m	I
IDH	ADP_m	A
AKGDH	$SUCCOA_m$	I
AKGDH	$NADH_m$	I

Table 3.5.: Regulatory interactions included into the SK-model of mitochondrial energy metabolism. The letter 'I' stands for inhibitory relationships, whereas 'A' describes activating influences.

3.2. Model analysis

3.2.1. Quantitative analysis

For each steady state, the numbers of stable and unstable SK-models were counted in the unbalanced data set of 10^4 randomly sampled models with and without filtering for biological feasibility. Oscillatory and non-oscillatory models were analysed in a similar manner.

3.2.2. Univariate qualitative analysis

In order to determine the elasticities associated with stability and oscillations, elasticity distributions of the corresponding SK-models were compared using the Kolmogorov-Smirnoff test. Those elasticities whose distributions significantly differed between both classes were extracted. If not stated otherwise, the significance level α was set to 0.01.

3.2.3. Multivariate qualitative analysis

Using stability or oscillations as class labels, classifiers were trained in order to detect discriminating patterns in the parameter space. Focusing only on the biologically feasible models, each set of randomly sampled elasticities or FCCs derived from the elasticities served as a feature vector for classifier training. As a first step, the impact of training data size on classification performance was assessed. In doing so, balanced datasets of increasing sample sizes were created by repeatedly sampling model parameters until equal numbers of stable and unstable, as well as oscillatory and non-oscillatory, cases were obtained. For each sample size, classifiers were trained on five independently created training data sets. In order to test the generalizability of the derived decision function, the performance of each classifier was evaluated on a separately sampled balanced test set.

Classification accuracy on training and test data were assessed by the balanced error rate (BER)

$$BER = 0.5 \cdot \left(\frac{FP}{P} + \frac{FN}{N} \right), \tag{3.2}$$

where FP and FN are the numbers of wrong positives and negative predictions, and P and N are the total numbers of positive or negative labels in the data set. Here, the 'positive' class is assumed to consist of those models which are stable or oscillating, and the 'negative' class consists of unstable or non-oscillating models. However, due to the symmetry of equation (3.2), the reverse definition would be possible as well.

Decision tree training was performed using the `C5.0` library in `R` (R Core Team, 2012) in the `RULES` mode. After training, the generalizability of each of the obtained decision tree rulesets was assessed by the Laplace ratio

$$L_k = \frac{h_k - e_k + 1}{h_k + 2}. \tag{3.3}$$

Here, k is the ruleset index, h_k is the number of 'hits', which are defined as the test samples meeting the conditions given by the k^{th} ruleset, and e_k is the number errors made when applying the ruleset to these test samples (Quinlan, 2013). In total, 10^5 test samples were used for the assessment. Since our aim was to derive reliable conditions for stability and instability, we selected only those rulesets with Laplace ratio $\geqslant 0.95$ for further analyses.

RVM training was performed using the `SparseBayes Software`, which is a freely available MATLAB implementation of the RVM learning algorithm (Tip-

ping, 2013). For each training data size, five independent training sets were created and RVMs were trained on each of them. Following the author's example (Tipping *et al*, 2003), the Gaussian kernel was chosen for RVM training. It required a variance parameter σ which was determined by grid search. In doing so, each parameter was evaluated on a set of $50,000$ test samples. The final classification accuracy was computed on a holdout dataset of $50,000$ samples using the best performing sigma from each grid search.

4. A MATLAB toolbox for structural kinetic modelling*

Abstract

Summary: Structural kinetic modelling (SKM) enables the analysis of dynamical properties of metabolic networks solely based on topological information and experimental data. Current SKM-based experiments are hampered by the time-intensive process of assigning model parameters and choosing appropriate sampling intervals for Monte-Carlo experiments. We introduce a toolbox for the automatic and efficient construction and evaluation of structural kinetic models (SK-models). Quantitative and qualitative analysis of network stability properties is performed in an automated manner. We illustrate the model building and analysis process in detailed example scripts that provide toolbox-implementations of previously published literature models.

Availability: The source code is freely available for download at http://bioinformatics.uni-potsdam.de/projects/skm.

4.1. Introduction

SKM enables the analysis of dynamical features of metabolic systems in steady states, without requiring the knowledge necessary for the construction of kinetic models, such as kinetic parameters and reaction rates. Instead, these properties are derived solely from topological information and experimentally measurable steady state data. In doing so, the SKM algorithm derives a 'parameterized'

*Original article:
Dorothee Girbig, Joachim Selbig and Sergio Grimbs
A MATLAB toolbox for structural kinetic modeling
BIOINFORMATICS APPLICATIONS NOTE
Vol. 28 no. 19 2012, pages 2546-2547

version of the system's Jacobian matrix, in which model parameters encode the partial derivatives of the reaction rates around the steady state (Steuer *et al*, 2006). Once the Jacobian matrix is computed for a given set of parameters, the evaluation of its eigenvalues indicates whether the steady state is stable. Here, a simple normalization step enables the restriction of the parameter values to predefined sampling intervals (for example, $(0, 1]$ for classical enzyme kinetics). This enables the combination of SKM with a Monte-Carlo approach (Steuer *et al*, 2006) in which large numbers of SK-models are created using randomly sampled parameters. The resulting Jacobian matrices can then be evaluated quantitatively (by counting the proportions of stable and unstable models) or qualitatively (by analysing the conditions that lead to such stability or instability). Qualitative SKM-analysis can be performed by pairwise comparisons of the model parameters leading to stable or unstable states (Grimbs *et al*, 2007) or by machine learning approaches that search for patterns in the parameter space (Girbig *et al*, 2012a).

The SKM-experiments presented so far used customized algorithms in which the SK-models had been constructed manually 'from scratch' for each pathway (Steuer *et al*, 2006; Grimbs *et al*, 2007; Steuer *et al*, 2007; Reznik and Segrè, 2010). While this might be sufficient for small systems like in the mentioned examples, the construction of SK-models for larger systems, or even systems of genomic scale is not feasible manually. However its potential to be applied to large-scale systems is a major advantage of SKM compared to kinetic modelling. Because it does not rely on detailed kinetic knowledge, it is well-suited for the investigation of large metabolic systems for which only limited or uncertain information about the individual reaction mechanisms is available.

Here we present a MATLAB toolbox that enables the automated construction and evaluation of SK-models. Models can be constructed from a minimal input consisting only of the stoichiometric matrix $\mathbf{N}$, steady state concentrations $[\mathbf{S}]^*$ and the steady state fluxes $\mathbf{v}^*$, with the experimental data being obtained from metabolomics and isotope tracing experiments. Model parameters can be derived automatically based on the information in $\mathbf{N}$. The user can also assign additional model parameters (for example to describe regulatory interactions) or manually manipulate the suggested parameter positions and intervals.

We illustrate the model building and analysis process in example scripts which demonstrate the construction of previously published literature models (Steuer *et al*, 2006; Girbig *et al*, 2012a) using the toolbox.

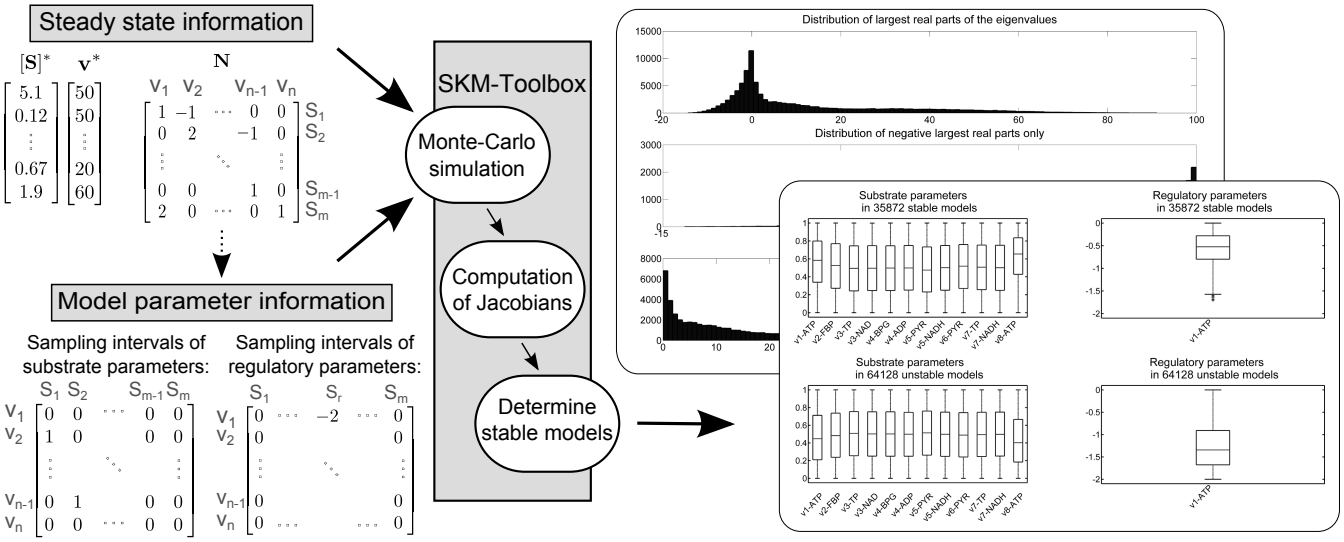


Figure 4.1.: SK-model building and evaluation using the SKM-toolbox. Required input arguments are the stoichiometric matrix $\mathbf{N}$, steady state concentrations $\mathbf{S}^0$, and the fluxes $\mathbf{v}^0$. Model parameters are assigned according to the information in $\mathbf{N}$. They can be manually adjusted (i.e. by adding allosteric regulators) before starting the Monte-Carlo-simulation. The resulting distributions of model parameters and eigenvalues for stable/unstable models are displayed automatically.

4.2. Features

The key functionalities of the toolbox can be summarized as follows:

- SK-models can be constructed from a minimum required input which consists only of $\mathbf{N}$, $[\mathbf{S}]^*$ and $\mathbf{v}^*$.
- Information about the model components and their stoichiometries can be efficiently imported from SBML files.
- The program is flexible to modifications of the model parameters. This can be achieved by either manually modifying the automatically determined parameters, or by building parameter matrices 'from scratch'.
- MATLAB functions for the quantitative and qualitative analysis of the resulting models are provided.

The most labour-intensive step in the construction of SK-models for Monte-Carlo experiments consists of choosing the model parameters' network positions, and assigning appropriate sampling intervals. The sampling intervals depend on the type of kinetic rate law assumed for the reactions. For example, the interval

$(0, 1]$ serves for modelling enzyme-substrate interactions in enzymatic reactions while $[-n, 0)$ models the impact of an allosteric inhibitor with Hill coefficient n.

Internally, the toolbox uses a MATLAB `struct` object to store network positions of model parameters that describe different types of interactions. If not provided as an input argument for the toolbox, the `struct` will be automatically created based on the stoichiometric coefficients in $\mathbf{N}$. The toolbox also enables the generation of a template `struct` for manual modification by the user (for example by including regulatory interactions) prior to the start of the program.

After Monte-Carlo simulation, the eigenvalues of each Jacobian matrix, as well as an indicator of the stability of each underlying model are returned. This information can be further analysed by additional toolbox functions, such as pairwise comparisons between stable and unstable models. It can also be converted into input for the decision tree algorithms `C4.5` or `C5.0` (Quinlan, 2013), or analysed manually with respect to specific questions posed by the user. For instance, the example script for the simplified glycolysis model of Steuer *et al* (2006) demonstrates how to reproduce the results in the original publication with the toolbox. Using this system as an example, Figure 4.1 provides an overview of the model building and evaluation process.

4.3. Availability and implementation

The SKM-toolbox was developed under MATLAB version 7.11 (release R2010b). The SBML import requires the freely available LibSBML package (Bornstein *et al*, 2008).

4.4. Summary

The proposed toolbox helps to overcome a major bottleneck of SKM-experiments, namely the time-intensive assignment of the model parameters. Furthermore, it provides a unifying framework for publishing and sharing SK-models. With the increasing availability of genome-scale reconstructions of metabolic networks, as well as the fast progress in experimental methods measuring concentrations and fluxes in these networks, our toolbox can assist in applying SKM to larger and more complex systems than attempted so far.

5. Results obtained for the small example pathways

This chapter describes the results of analysing the small example pathways introduced in Section 3.1.1. It will start with a brief description of the steady states exhibited by each pathway, and will then continue with the outcomes of the various univariate and multivariate SKM experiments.

5.1. Steady states

Figure 5.1 shows the steady state concentrations for each pathway and metabolite. In all pathways, the concentration of S_1 clearly exceeded the other three metabolites.

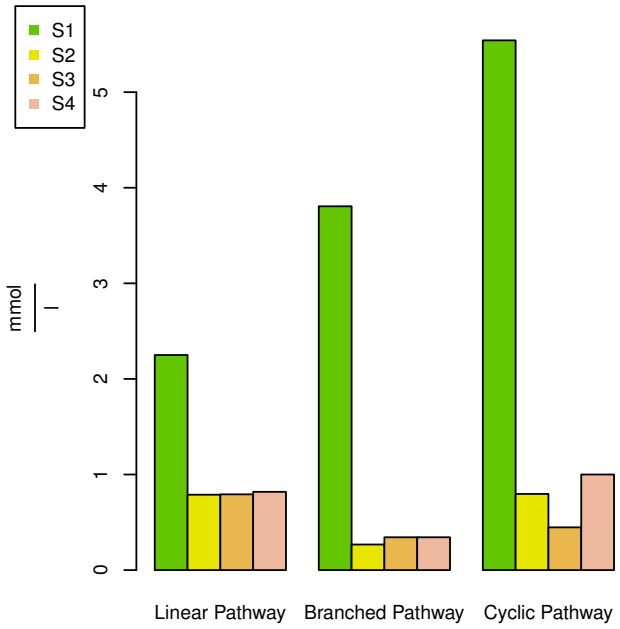


Figure 5.1.: Concentrations of metabolites S_1 - S_4 in the steady states computed for the small kinetic example models. Different pathway topologies had strong influences on the metabolite concentrations in steady state.

A possible reason for the high abundance of S_1 might be that the external *source* metabolite had fixed concentration and resulted in a high forward reaction rate (Figure 5.2). This high forward rate needed to be compensated by a high concentration of its product S_1 in order to adapt the overall rate of r_1 to the rates of the subsequent reactions in steady state. Reaction r_2, which was activated by S_4 (linear and branched pathway), or S_3 (cyclic pathway), showed lower rates in the forward and backward direction than the other reactions, indicating a weak influence of the activation term in the given steady states.

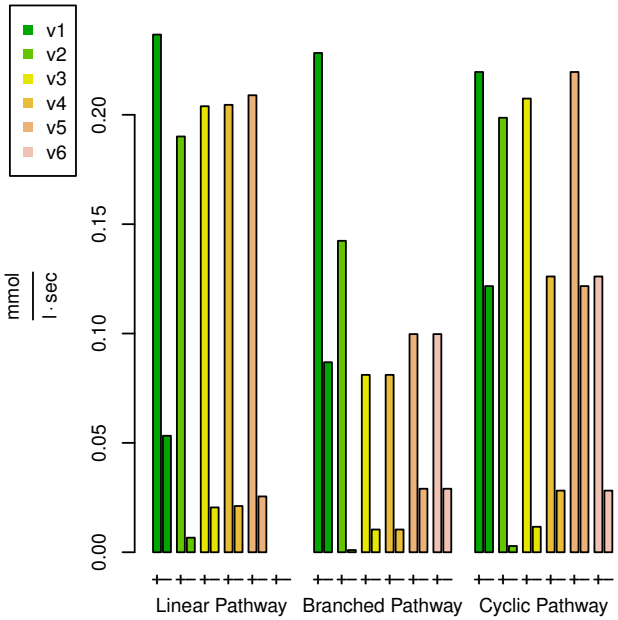


Figure 5.2.: Reaction rates of the steady states computed for the small kinetic example models with different pathway structures

Elasticities for each steady state are depicted in Table 5.1. Because of the homogeneous choice of kinetic parameters, many substrate- or product associated elasticities were similar within each pathway (e.g. $\varepsilon_{S_2}^{v_3^+} = \varepsilon_{S_3}^{v_4^+} = \varepsilon_{S_4}^{v_4^+} = 0.79$ in the linear pathway, or $\varepsilon_{S_2}^{v_3^+} = \varepsilon_{S_2}^{v_4^+} = 0.92$ and $\varepsilon_{S_3}^{v_5^+} = \varepsilon_{S_4}^{v_6^+} = 0.9$ in the branched pathway). With exception of r_2, substrate associated elasticities were all close to, or larger than, 0.8, indicating a weak degree of saturation of the corresponding enzymes by their substrates. Elasticities describing the influences of products on the backward reactions were all close to 1, showing that the corresponding enzymes were even less saturated by their products than they were by their substrates. These high elasticities allowed for a fast response to perturbations by all enzymes downstream of r_2 in each pathway. Elasticities associated with the feedback term ($\varepsilon_{S_4}^{v_2^{+/-}}$ or

$\varepsilon_{S_3}^{v_2^{+/-}}$) had slightly lower values, showing that the positive feedback adapted more slowly to changes in the activating metabolite.

	v_1^+	v_1^-	v_2^+	v_2^-	v_3^+	v_3^-	v_4^+	v_4^-	v_5^+	v_5^-
S_1	-0.05	0.95	0.58	-0.42						
S_2			-0.01	0.99	0.79	-0.20				
S_3					-0.02	0.98	0.79	-0.20		
S_4			0.55	0.55			-0.02	0.98	0.79	-0.21

(a) Steady state elasticities in the linear pathway.

	v_1^+	v_1^-	v_2^+	v_2^-	v_3^+	v_3^-	v_4^+	v_4^-	v_5^+	v_5^-	v_6^+	v_6^-
S_1	-0.09	0.91	0.44	-0.55								
S_2				1.00	0.92	-0.08	0.92	-0.08				
S_3					-0.01	0.99			0.90	-0.10		
S_4			0.74	0.74			-0.01	0.99			0.90	-0.10

(b) Steady state elasticities in the branched pathway.

	v_1^+	v_1^-	v_2^+	v_2^-	v_3^+	v_3^-	v_4^+	v_4^-	v_5^+	v_5^-	v_6^+	v_6^-
S_1	-0.12	0.88	0.35	-0.64					-0.12	0.88		
S_2			-0.01	0.99	0.79	-0.21						
S_3			0.69	0.69	-0.01	0.99	0.87	-0.13			0.87	-0.13
S_4							-0.03	0.97	0.78	-0.22		

(c) Steady state elasticities in the cyclic pathway.

Table 5.1.: Elasticities in each steady state of the example models. Each entry $\varepsilon_{S_i}^{v_j}$ describes the impact that changing the normalized concentration $\frac{S_i}{S_i^*}$ has on the normalized reaction rate $\frac{v_j}{v_j^*}$.

The corresponding FCCs that were computed from the elasticities using equation (2.15) are shown in Table 5.2. As explained in Section 2.3.1, these values describe the summarized impact of each enzyme perturbation on the overall reaction rates. In the linear pathway, the control over the total flux became smaller with each of the consecutive enzymes in the reaction chain. While reaction r_1 had a large positive influence on the flux through the pathway, r_4 exhibited no control at all. The last reaction, r_5, even exhibited negative control, which implied that an

increase in this reaction would lead to a strong decrease in total flux. A possible reason for this observation could be that an acceleration of r_5 would lead to an increased consumption of its substrate S_4. Because S_4 acted as an activator of r_2, a reduction in this metabolite would lead to a reduction in the steady state reaction rate of r_2 and therefore to a reduction in overall flux.

	e_1	e_2	e_3	e_4	e_5
F^t	0.79	0.44	0.03	0.00	-0.26

(a) FCCs for the linear pathway steady state. Because of the linear structure, there exists only one steady state flux $F^t = v_1 = v_2 = v_3 = v_4 = v_5$ through the whole pathway.

	e_1	e_2	e_3	e_4	e_5	e_6
F^t	0.59	0.92	-0.23	0.24	-0.03	-0.49
F^{b_1}	0.59	0.92	0.22	-0.21	0.03	-0.55
F^{b_2}	0.59	0.92	-0.68	0.68	-0.08	-0.44

(b) FCCs for the branched pathway steady state. Because of the branched pathway structure, steady state fluxes can be divided into the flux through the upper branch $F^{b_1} = v_3 = v_5$, the flux through the lower branch $F^{b_2} = v_4 = v_6$, and the total flux $F^t = v_1 = v_2$.

	e_1	e_2	e_3	e_4	e_5	e_6
F^t	0.48	1.41	0.04	-0.63	-0.10	-0.21
F^{c_1}	0.41	1.41	0.04	-0.20	-0.03	-0.64
F^{c_2}	0.35	1.41	0.04	0.24	0.04	-1.07

(c) FCCs for the cyclic pathway steady state. Due to the cyclic pathway structure, steady state fluxes can be divided into the total flux through the system $F^t = v_1 = v_6$, the flux through the upper part of the cycle $F^{c_1} = v_2 = v_3$ and the flux through the lower part of the cycle $F^{c_2} = v_4 = v_5$.

Table 5.2.: FCCs for each investigated steady state. Each entry describes the impact of changing the concentration of enzyme e_j on the steady state fluxes in the normalized system.

The FCCs of the branched pathway's steady state showed that reactions r_1 and r_2 exhibited positive control on the flux through the pathway and that they influenced both branches to the same extent. The control exhibited by r_2 was much larger than that exhibited by r_1. Reactions r_3 and r_5, which were involved in the upper branch of the pathway, exhibited negative control on the flux through the lower branch (F^{b_2}). The opposite behaviour could be observed for the reactions of the lower branch, r_4 and r_6. Of all reaction involved in either of the two branches,

only r_4 exhibited positive control on the total flux through the pathway. This can be explained by the fact that r_4 was involved in the positive feedback loop exhibited by its product S_4. An increase in r_4 also increased the concentration of S_4, which led to an acceleration of r_2, and therefore also of the total flux. In contrast, similar to the linear pathway, a reduction of S_4, either by increased efflux catalysed by reaction r_6, or by an increased flux through the upper branch reactions r_3 and r_5, reduced the positive feedback and therefore decreased the total flux.

Similar to the branched pathway, FCCs of the cyclic pathway steady state also indicated a competitive coordination of fluxes through the upper and lower part of the cycle. Thus, increasing the rates through r_4 or r_5 slowed down the steady state flux through the upper part of the cycle (F^{c1}), and hence, through the whole system (F^t).

Table 5.3 shows the eigenvalues with maximum real parts derived from the Jacobian matrices of each example model in steady state. As indicated by the complex conjugate pair of eigenvalues, the linear pathway was the only example system that exhibited oscillations around the steady state. The smallest maximum real part belonged to the steady state of the branched pathway. Consequently, this steady state enabled the fastest response to perturbations.

	Linear pathway	Branched pathway	Cyclic pathway
λ_1	$-0.34 + 0.14i$	-0.79	-0.04
λ_2	$-0.34 - 0.14i$	-0.30	-0.06
Stable	yes	yes	yes
Oscillating	yes	no	no

Table 5.3.: The two eigenvalues with largest real parts associated with each steady state.

The results described so far referred to the steady state properties corresponding to the underlying kinetic models. If the same steady states had been produced by enzymes with different kinetic parameters to those used in the kinetic models, different elasticities and FCCs could have emerged. As a result, the responses to perturbations could have been of different nature to those presented in Table 5.3. We will next attempt to obtain a more comprehensive view of the possible types of perturbation responses that can arise from different enzyme properties and associated elasticities. The results of the corresponding SKM experiments will be presented in the next Section.

5.2. Quantitative analysis of local dynamic steady state properties

Out of 10,000 SK-models sampled for the linear pathway, only 907 (9.07 %) were biologically feasible in the sense that the randomly sampled elasticities led to non-negative kinetic parameter values (see Section 2.3.2 for details). For the branched and cyclic pathways, only 891 (8.91 %) and 944 (9.44 %) feasible models remained. These low numbers show that it is important to account for biological feasibility when performing SKM experiments.

In order to assess whether filtering for feasible models can impact the results of SKM experiments, probabilities of the different types of dynamic steady state properties were compared before and after filtering (Figure 5.3).

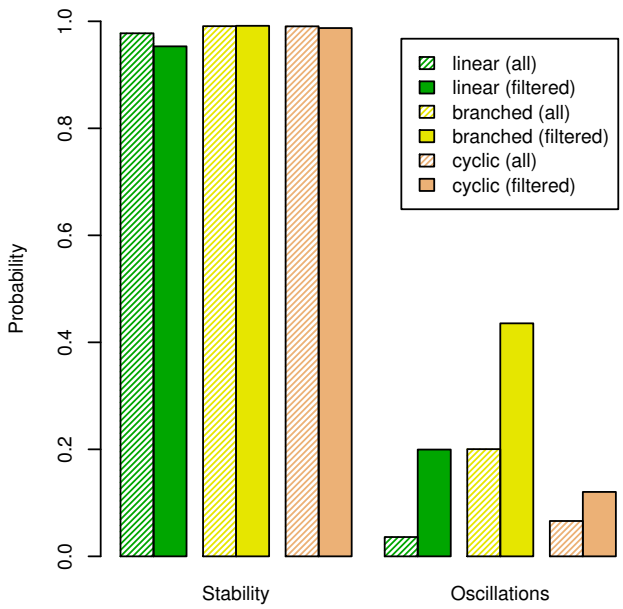


Figure 5.3.: Distribution of dynamic steady state properties in each pathway. The probability of observing oscillations increases after filtering for biologically feasible kinetic parameters.

While the proportions of stable models were hardly affected by the filtering, the chance to observe oscillations strongly increased in each pathway. Consequently, there seemed to be a stronger trend towards oscillatory behaviour among the biologically realistic models compared to the total space of randomly sampled models with arbitrary elasticity combinations.

A summary of the observed dynamic properties in each pathway is given in

Table 5.4. In all three pathway topologies, the majority of models resulted in stable steady states. The few observed instabilities almost exclusively occurred in non-oscillatory models. Oscillations only emerged around stable steady states in the given scenarios. The strongest tendencies towards oscillatory behaviour could be observed in the branched pathway, with 20% of oscillating models observed before filtering and over 44% after.

	Linear pathway		Branched pathway		Cyclic pathway	
	all	filtered	all	filtered	all	filtered
Total stability (%)	0.9776	0.9532	0.9909	0.9916	0.9906	0.9874
Total oscillations (%)	0.0361	0.1996	0.2004	0.4355	0.0661	0.1204
Stable/ with oscillations (%)	0.0361	0.1996	0.2002	0.4352	0.0661	0.1202
Stable/ no oscillations (%)	0.9415	0.7536	0.7907	0.5564	0.9245	0.8672
Unstable/ with oscillations (%)	0.0000	0.0000	0.0002	0.0003	0.0000	0.0002
Unstable/ no oscillations (%)	0.0224	0.0468	0.0089	0.0081	0.0094	0.0124

Table 5.4.: Detailed list of proportions of possible steady state properties in each pathway determined by Monte Carlo sampling.

A more detailed view of the possible dynamic properties per pathway is provided in Figure 5.4, which shows the distributions of the real and imaginary parts of the eigenvalues in the filtered models. In all three pathways, the majority of SK-models led to negative maximum real parts (located left of the origin in the 2D histograms). Imaginary parts that differed from zero were all located in this region as well. In contrast, eigenvalues with positive real parts only rarely displayed imaginary parts different from zero. This fits to the observation that oscillations were almost exclusively encountered around stable steady states.

5.3. Univariate search for discriminating features

Figure 5.5 shows an exemplary illustration of elasticity distributions in linear pathway models. After filtering for biological feasibility, most of the remaining elasticities in each pathway were located above the 50% quantile of the sampling interval

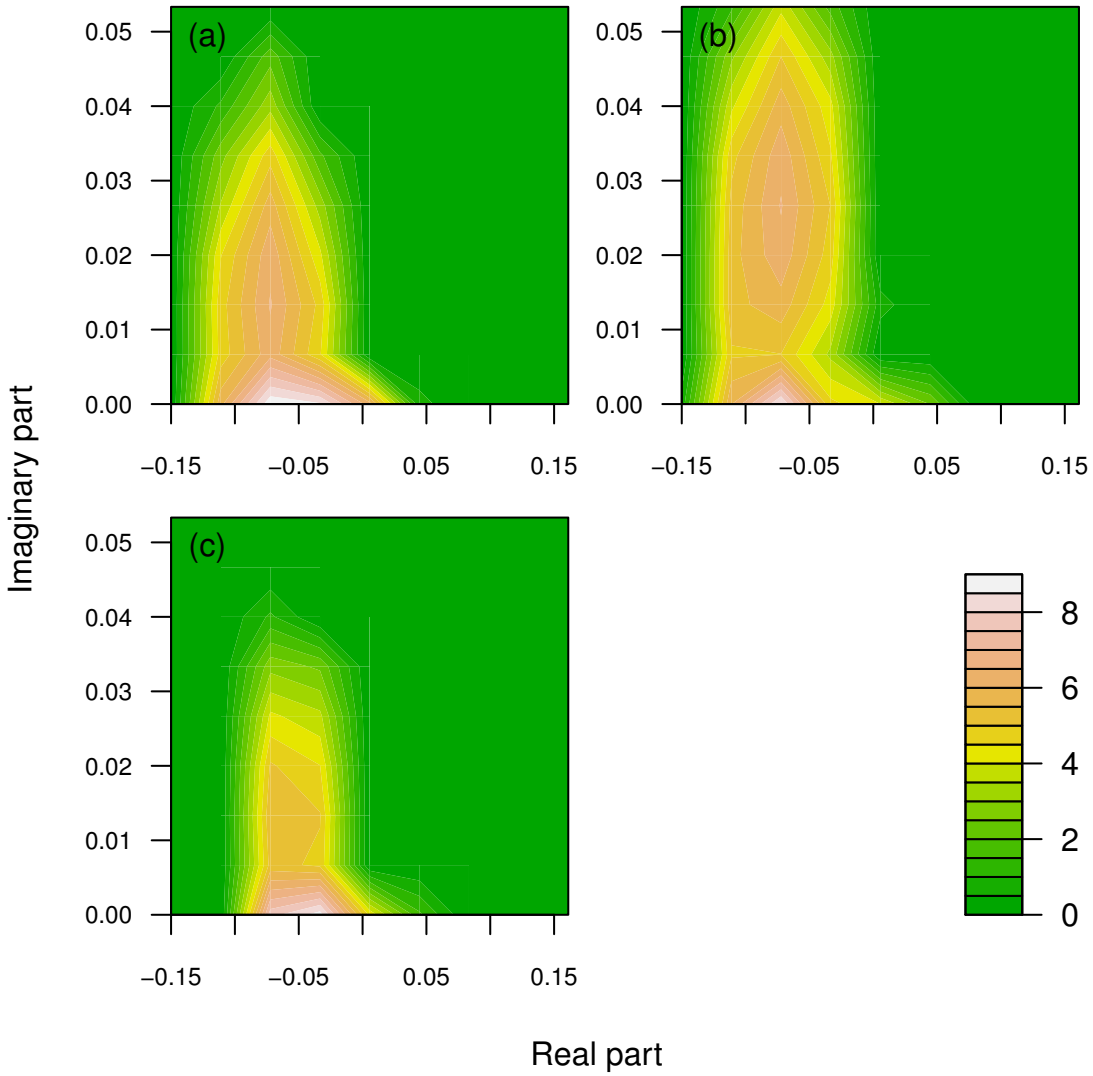


Figure 5.4.: 2D histograms displaying the distributions of eigenvalues with maximum real parts computed by SKM. The number of points in a specific bin, n, has been transformed to log(n+1). Adding the +1 in the logarithm ensures that a value of 0 (dark green) is equivalent to a frequency of 0. a): linear pathway; (b): branched pathway; (c): cyclic pathway.

$(0, 1]$. The reason for this behaviour was that the filtering criterion given in equation (3.1) favoured large elasticity values. Despite this general trend towards large values, several individual elasticities could be observed that favoured small values for a specific class. For example, $\varepsilon_{S_4}^{v_5^+}$ was located exclusively below 0.4 for unstable linear pathway models (Figure 5.5). In contrast, stable models could be realized with elasticities from the whole range $(0, 1]$. This can be interpreted in the following way: For any value in this elasticity, it was possible to find an appropriate

combination of values in the other elasticities that made the steady state stable. However, unstable states were only possible if the efflux reaction r_5 responded slowly to perturbations in its substrate S_4, regardless of the elasticities at other network positions.

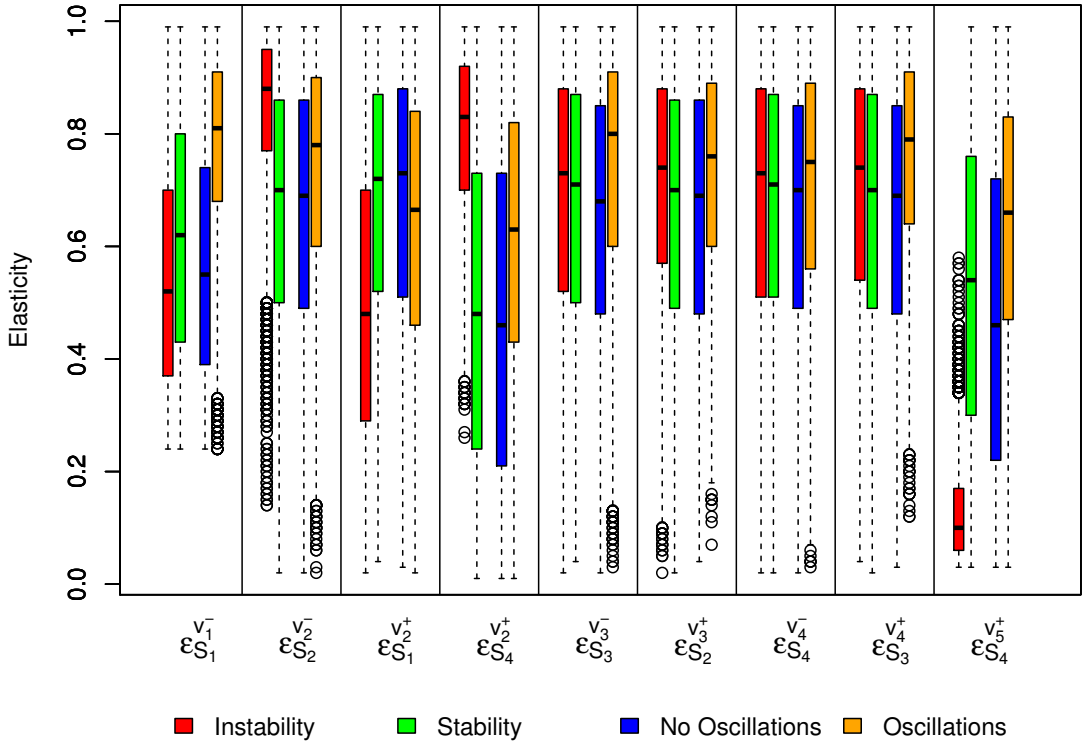
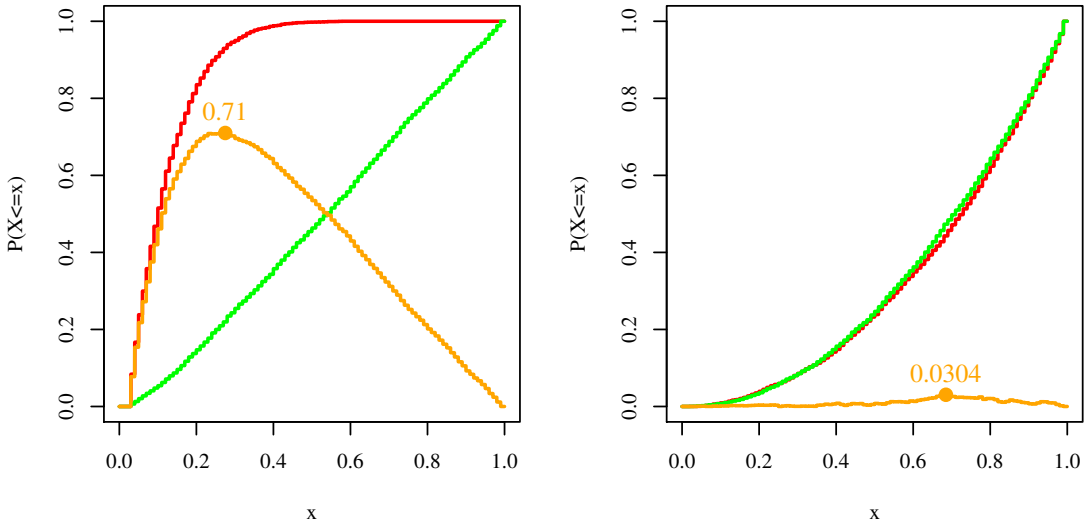


Figure 5.5.: Class specific distribution of elasticities in the linear pathway. Elasticities associated with the influx reaction r_1, the regulated reaction r_2, as well as the efflux reaction r_5 differ most strongly between each class.

In order to systematically search for elasticities that played dominant roles in determining different dynamic properties, the elasticity distributions of opposing classes (stable/unstable and oscillating/non-oscillating) were compared by the Kolmogorov-Smirnov test. This test enables the comparison of two empirical distribution functions $F_1(x)$ and $F_2(x)$ using the test statistic $D_{ks} = \sup_x |F_1(x) - F_2(x)|$, which describes the maximum distance between the distribution functions (Darling, 1957). For better illustration, Figure 5.6 shows two examples in which the distributions of elasticities $\varepsilon^{v_5^+}_{S_4}$ and $\varepsilon^{v_4^-}_{S_4}$ were compared between stable and unstable models of the linear pathway. While the distribution functions of $\varepsilon^{v_5^+}_{S4}$ differed strongly between stable and unstable models, those of $\varepsilon^{v_4^-}_{S4}$ coincided closely in both

classes.



(a) Comparison of the distributions of $\varepsilon_{S4}^{v_5^+}$. The null hypothesis can be rejected (computed p-value: 0).

(b) Comparison of the distributions of $\varepsilon_{S4}^{v_4^-}$. The null hypothesis cannot be rejected (p-value: 0.0197).

Figure 5.6.: Illustration of the Kolmogorov-Smirnov- (KS-) test. The distribution functions of (a) $\varepsilon_{S4}^{v_5^+}$ and (b) $\varepsilon_{S4}^{v_4^-}$ were compared between stable and unstable models of the linear pathway. Stable models are shown in green, unstable models in red. The pairwise distance is shown in orange. The displayed numbers refer to the maximum distance values, which are also the test statistics of the KS-test.

The systematic comparison of the distribution functions revealed that for all three pathway topologies, almost all elasticities contributed significantly to stable or oscillatory behaviour (Tables 5.5 - 5.7). Sorting the elasticities by the test statistics that compared stable and unstable models provided the following ranking: the biggest differences were observed for the elasticities associated with the regulated reaction r_2, as well as with the efflux reaction of the regulatory metabolite (S_3 in the cyclic pathway, S_4 otherwise) (Figure 5.7). This showed that the enzymes involved in the regulatory feedback loop were most critical for controlling stability. The elasticities associated with the efflux reactions were typically located at much lower values for unstable models than for stable models for all pathway topologies (see Figure 5.5 for linear pathway data). This showed that stability was most strongly endangered when the efflux of the activator reacted slowly to perturbations. We can assume that in such a case, instead of being carried out of the system, perturbations were amplified by the positive feedback loop. This mechanism appeared to be the main reason for instabilities.

The univariate comparisons performed so far gave a first impression of the importance of the individual elasticities for stability or oscillatory behaviour. However, in most cases they could not provide quantitative thresholds for elasticities associated with these properties. Instead, they only allowed the conclusions that certains elasticity were associated with 'rather big' or 'rather small' values in a certain class. Additionally, almost all network positions contributed significantly to these properties. Consequently, dynamic properties like stability or oscillations could not be attributed to single enzymes or metabolites alone.

A more refined picture of the interplay of the different network positions can only be obtained by multivariate approaches, which examine the impact of ensembles of elasticities that act together to cause instabilities or oscillations. In the following, we will demonstrate how machine learning approaches can be used for this task, and how they help to discover quantitative thresholds for elasticities associated with the dynamic properties of a steady state.

Stability		Oscillations	
Elasticity	Distance	Elasticity	Distance
$\varepsilon_{S_4}^{v_5^+}$	**0.71**	$\varepsilon_{S_1}^{v_1^-}$	**0.4242**
$\varepsilon_{S_4}^{v_2^+}$	**0.5086**	$\varepsilon_{S_4}^{v_5^+}$	**0.2652**
$\varepsilon_{S_2}^{v_2^-}$	**0.3598**	$\varepsilon_{S_4}^{v_2^+}$	**0.229**
$\varepsilon_{S_1}^{v_2^+}$	**0.3016**	$\varepsilon_{S_3}^{v_4^+}$	**0.1928**
$\varepsilon_{S_1}^{v_1^-}$	**0.1436**	$\varepsilon_{S_3}^{v_3^-}$	**0.1856**
$\varepsilon_{S_2}^{v_3^+}$	**0.0772**	$\varepsilon_{S_2}^{v_2^-}$	**0.1478**
$\varepsilon_{S_3}^{v_4^+}$	**0.065**	$\varepsilon_{S_2}^{v_3^+}$	**0.137**
$\varepsilon_{S_3}^{v_3^-}$	**0.0358**	$\varepsilon_{S_4}^{v_4^-}$	**0.092**
$\varepsilon_{S_4}^{v_4^-}$	0.0304	$\varepsilon_{S_1}^{v_2^+}$	**0.091**

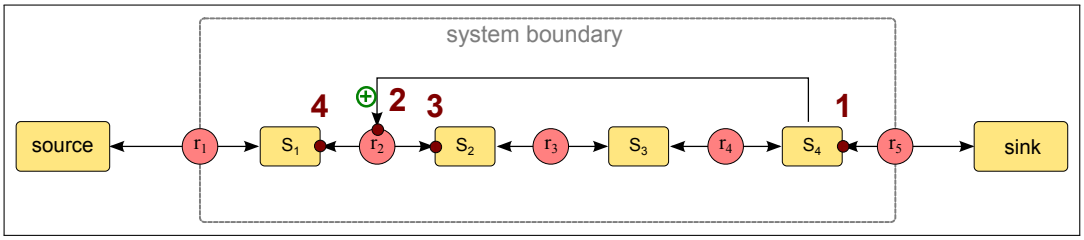
Table 5.5.: Distances between the elasticity distributions of the opposing classes for the linear pathway, determined by the Kolmogorov-Smirnov test. Bold values indicate significant differences (significance level: $\alpha = 0.01$).

Stability		Oscillations	
Elasticity	Distance	Elasticity	Distance
$\varepsilon^{v_2^+}_{S_4}$	**0.6364**	$\varepsilon^{v_2^+}_{S_4}$	**0.475**
$\varepsilon^{v_6^+}_{S_4}$	**0.617**	$\varepsilon^{v_2^-}_{S_2}$	**0.3416**
$\varepsilon^{v_2^-}_{S_2}$	**0.565**	$\varepsilon^{v_6^+}_{S_4}$	**0.233**
$\varepsilon^{v_2^+}_{S_1}$	**0.4946**	$\varepsilon^{v_4^+}_{S_2}$	**0.1252**
$\varepsilon^{v_3^+}_{S_2}$	**0.2322**	$\varepsilon^{v_2^+}_{S_1}$	**0.0728**
$\varepsilon^{v_4^+}_{S_2}$	**0.2108**	$\varepsilon^{v_1^-}_{S_1}$	**0.0498**
$\varepsilon^{v_4^-}_{S_4}$	**0.1596**	$\varepsilon^{v_5^+}_{S_3}$	**0.0424**
$\varepsilon^{v_5^+}_{S_3}$	**0.1**	$\varepsilon^{v_4^-}_{S_4}$	**0.0356**
$\varepsilon^{v_1^-}_{S_1}$	**0.0958**	$\varepsilon^{v_3^+}_{S_2}$	0.0298
$\varepsilon^{v_3^-}_{S_3}$	**0.0536**	$\varepsilon^{v_3^-}_{S_3}$	0.022

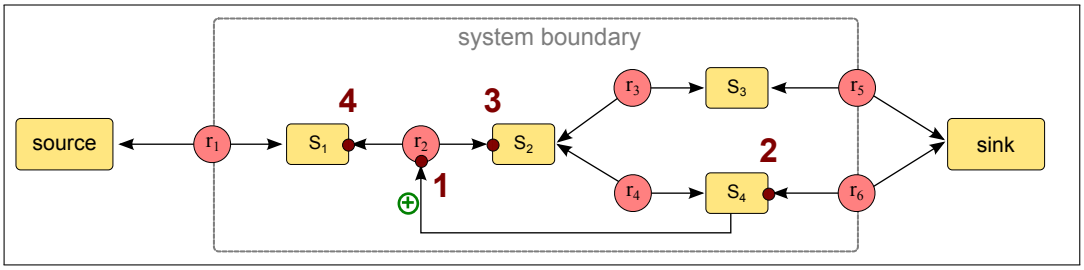
Table 5.6.: Distances between the elasticity distributions of the opposing classes for the branched pathway, determined by the Kolmogorov-Smirnov test. Bold values indicate significant differences (significance level: $\alpha = 0.01$).

Stability		Oscillations	
Elasticity	Distance	Elasticity	Distance
$\varepsilon^{v_2^+}_{S_3}$	**0.6768**	$\varepsilon^{v_2^+}_{S_3}$	**0.5548**
$\varepsilon^{v_6^+}_{S_3}$	**0.6092**	$\varepsilon^{v_2^-}_{S_2}$	**0.4332**
$\varepsilon^{v_2^-}_{S_2}$	**0.5832**	$\varepsilon^{v_4^+}_{S_3}$	**0.2254**
$\varepsilon^{v_2^+}_{S_1}$	**0.418**	$\varepsilon^{v_2^+}_{S_1}$	**0.1906**
$\varepsilon^{v_4^+}_{S_3}$	**0.304**	$\varepsilon^{v_3^+}_{S_2}$	**0.1774**
$\varepsilon^{v_4^-}_{S_4}$	**0.1214**	$\varepsilon^{v_6^+}_{S_3}$	**0.1678**
$\varepsilon^{v_3^+}_{S_2}$	**0.1114**	$\varepsilon^{v_1^-}_{S_1}$	**0.0552**
$\varepsilon^{v_1^-}_{S_1}$	**0.0716**	$\varepsilon^{v_4^-}_{S_4}$	**0.0516**
$\varepsilon^{v_5^-}_{S_1}$	**0.0482**	$\varepsilon^{v_3^-}_{S_3}$	**0.0444**
$\varepsilon^{v_5^+}_{S_4}$	**0.0334**	$\varepsilon^{v_5^-}_{S_1}$	0.02
$\varepsilon^{v_3^-}_{S_3}$	0.0302	$\varepsilon^{v_5^+}_{S_4}$	0.0112

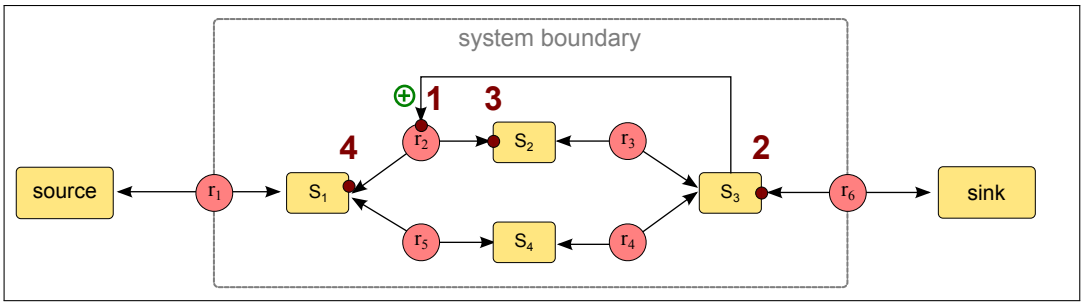
Table 5.7.: Distances between the elasticity distributions of the opposing classes for the cyclic pathway, determined by the Kolmogorov-Smirnov test. Bold values indicate significant differences (significance level: $\alpha = 0.01$).



(a) Linear pathway

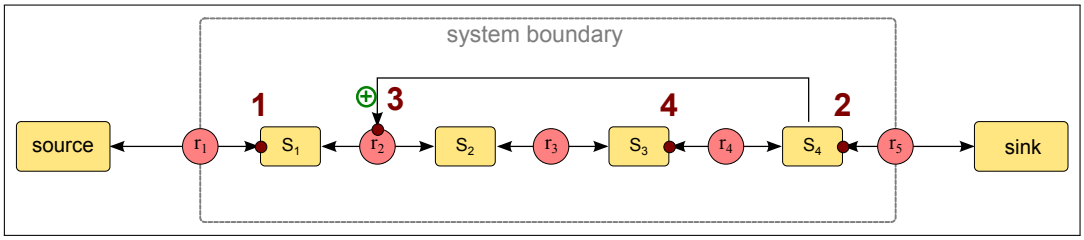


(b) Branched pathway

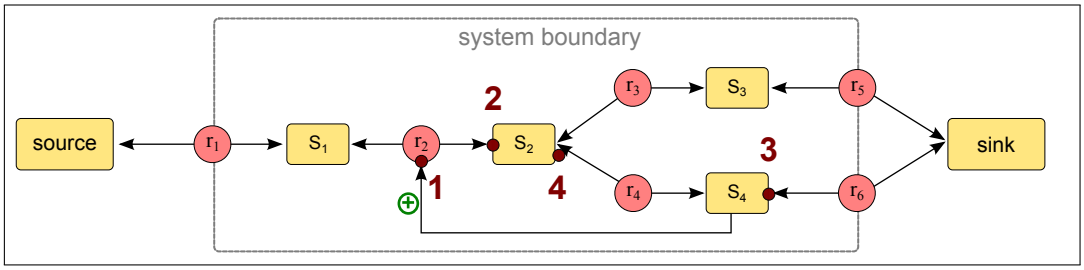


(c) Cyclic pathway

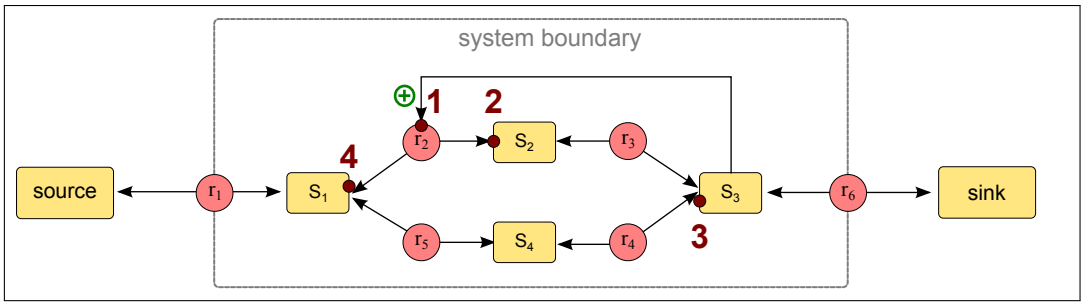
Figure 5.7.: Ranking individual elasticities with respect to their distribution differences between stable and unstable models. The ranking was determined by the Kolmogorov-Smirnov test (see Tables 5.5-5.7 for details). The four elasticities with largest impact on stability have been marked in red.



(a) Linear pathway



(b) Branched pathway



(c) Cyclic pathway

Figure 5.8.: Ranking individual elasticities with respect to their distribution differences between oscillatory and non-oscillatory models. The ranking has been determined by the Kolmogorov-Smirnov test (see Tables 5.5-5.7 for details). The four elasticities with largest impact on oscillatory behaviour have been marked in red.

5.4. Analysing stability and oscillation conditions by decision trees

5.4.1. Decision tree classification performance

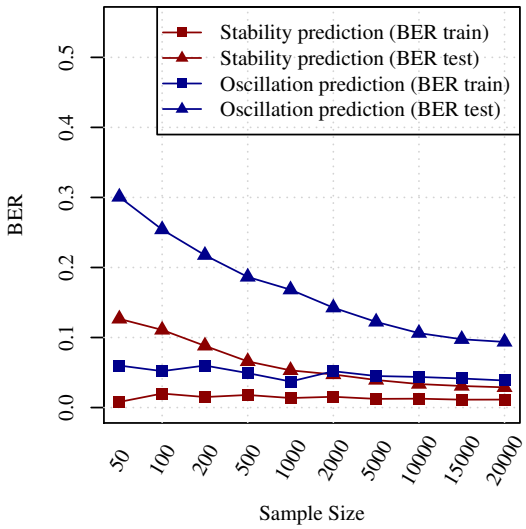
In order to assess the performance of the decision tree algorithm in classifying stable versus unstable as well as oscillatory versus non-oscillatory models, trees were trained on balanced datasets of increasing sample size as described in Section 3.2.3. Figures 5.9 - 5.11 show the mean balanced error rates (BERs) obtained for each pathway. A complete list of the mean BERs and their associated standard deviations is given in Appendix C.1.1.

Classification by elasticities

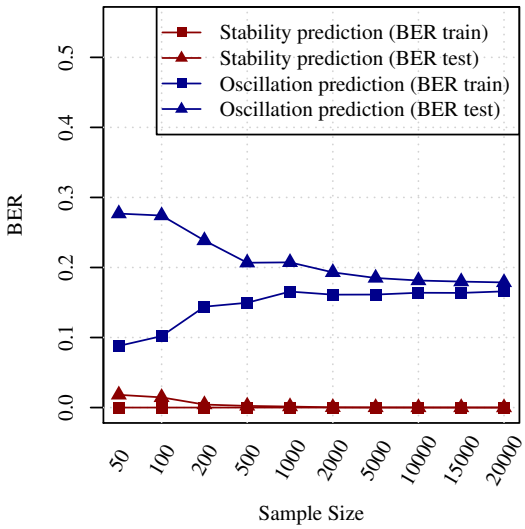
When classifying based on elasticities, training errors of stability prediction were small for all sample sizes (red squares in Figures 5.9 (a), 5.10 (a), 5.11 (a)). This indicated that the training data contained sufficient information for the decision tree algorithm to make into precise rules. In machine learning terminology, this property can be reformulated as having 'little bias'. In order to assess the generalizability of the derived rules, however, they needed to be evaluated on a separate test data set.

As could be expected, the test errors (red triangles) were higher than the training errors for small sample sizes because the small training data sets did not provide sufficient information to derive robust and well generalizable classification criteria. When increasing the sample size, however, the test errors approximated the training errors closely for each pathway structure. This indicated that the derived rulesets were well generalizable and not prone to overfitting if the training data was chosen sufficiently large.

Training errors of elasticity-based oscillation prediction tended to be slightly higher than those for stability prediction for the majority of sample sizes (blue squares in Figures 5.9 (a), 5.10 (a), 5.11 (a)), hinting at a slightly higher bias. In other words, it was more difficult for the decision tree algorithm to formulate patterns for oscillatory conditions than for stability conditions. The test errors (blue triangles) reached a plateau significantly higher than the training errors in all pathways (see Table C.1a for standard deviations)). This indicated that the derived decision tree rules were not as well generalizable and more prone to overfitting as was the case for stability prediction.

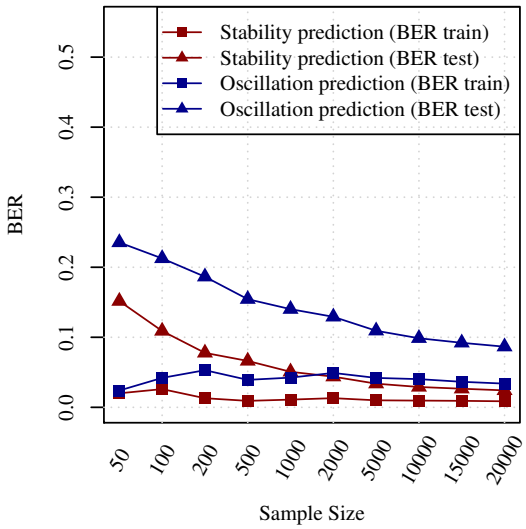


(a) BERs obtained by classifying based on elasticities.

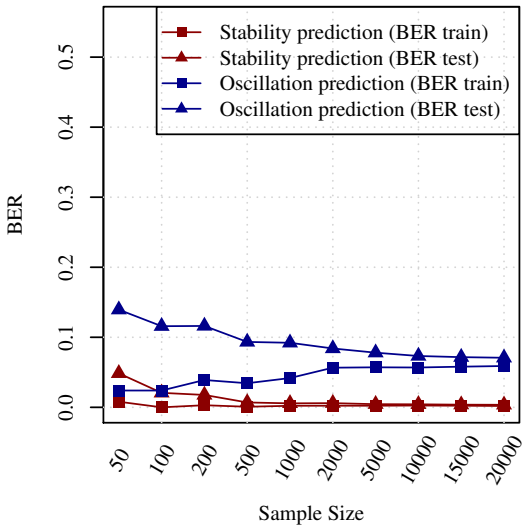


(b) BERs obtained by classifying based on FCCs.

Figure 5.9.: BERs obtained by classification of linear pathway models according to different types of features. The depicted values represent the average over the BERs obtained from five trees created for each training data size.



(a) BERs obtained by classifying based on elasticities.

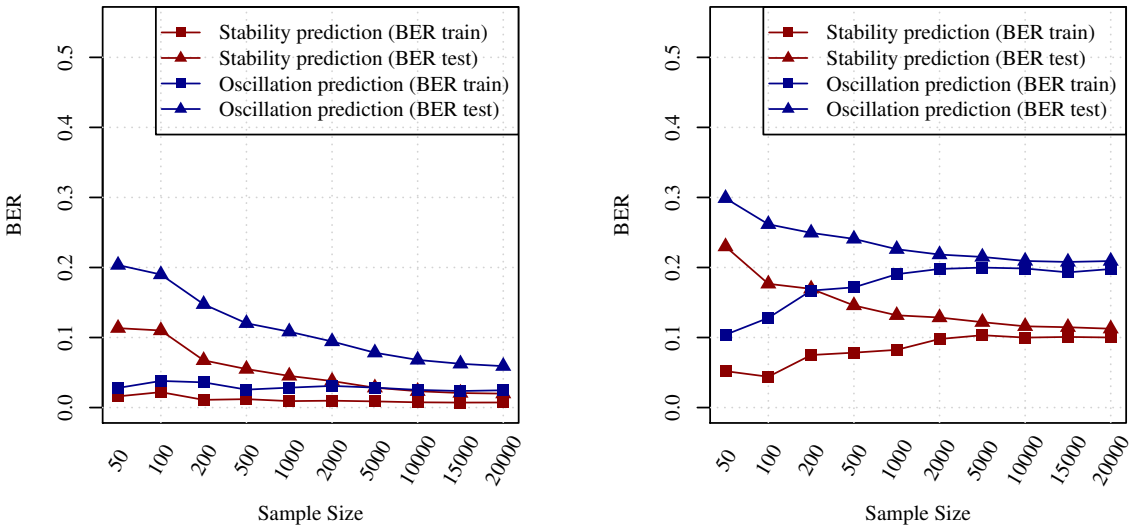


(b) BERs obtained by classifying based on FCCs.

Figure 5.10.: BERs obtained by classification of branched pathway models according to different types of features. The depicted values represent the average over the BERs obtained from five trees created for each training data size.

Classification by FCCs

The differences in performance between stability and oscillation prediction were even more prominent when the classification was based on FCCs instead of elas-



(a) BERs obtained by classifying based on elasticities.

(b) BERs obtained by classifying based on FCCs.

Figure 5.11.: BERs obtained by classification of cyclic pathway models according to different types of features. The depicted values represent the average over the BERs obtained from five trees created for each training data size.

ticities. In the linear and branched pathway, training and test errors for stability prediction both approximated zero, which hinted at a perfect or close to perfect separation based on the derived rules with almost no bias or overfitting. In other words, stability rulesets seemed to be particularly easy to derive with high generalizability (red curves in Figure 5.9 (b), 5.10 (b), 5.11 (b)). In the cyclic pathway, however, high training errors hinted at a high bias for stability prediction. We can conclude that it was difficult for the decision tree algorithm to derive stablizing FCC patterns from the training data for this system. Similar to the linear and branched pathway, the test errors closely approximated the training errors for large sample sizes. This showed that, even if deriving stabilizing rules from the training data could be problematic for the decision tree algorithm, the rules that were eventually detected were not prone to overfitting and generalized well.

In all three pathways, oscillation prediction based on FCCs was prone to high bias, as indicated by large training errors (blue curves in Figure 5.9 (b), 5.10 (b), 5.11 (b)). However, test errors approximated the training errors closely for large sample sizes, hinting at little overfitting in all cases.

Summary: Classification performance

In summary, stability prediction based on elasticities enabled the derivation of well generalizable patterns with good performances for all pathway structures. Stability prediction based on FCCs worked optimally for the linear and branched pathway, while it encountered problems for the cyclic pathway. Oscillatory patterns based on elasticities produced smaller training errors than those based on FCCs. However, the derived rules were prone to overfitting. In general, however, they still tended to outperform those derived by FCCs in terms of test errors. An exception was the branched pathway, for which the test errors for elasticity patterns and FCC patterns were comparable for large sample sizes (blue triangles in Figures 5.10 (a) and 5.10 (b)).

From the observations obtained so far, we can conclude that different pathway topologies and the resulting differences in steady state values can have profound impact on the classification performance. When evaluating an SKM experiment by supervised machine learning, it is therefore necessary to compute individual learning curves similar to the ones presented here in order to assess the quality and generalizability of the derived patterns.

5.4.2. Ruleset numbers and sizes

In order to analyse the derived decision trees, we obtained the corresponding rulesets produced by the `C5.0` algorithm. These rulesets summarized all paths through each tree in an easily readable format (Quinlan, 2013). Each ruleset was annotated with the associated numbers of correctly and incorrectly classified samples. These values were used to compute Laplace values which served as objective performance measured for each ruleset.

In general, the number of rulesets strongly increased for increasing training sample sizes (see Appendix C.1.2 for a detailed overview). However, only a relatively small proportion of rulesets exhibited sufficiently high Laplace values on the test data ($L \geqslant 0.95$) in order to be considered as 'reliable enough' for further consideration (see Tables C.4 - C.6) in the Appendix. For the following analyses, we will focus only on rulesets of those trees derived using the maximum training size of 20,000 because they offered the most comprehensive collections of rulesets with sufficient generalizability.

Rulesets derived by the `C5.0` algorithm each can have different numbers of conditions, depending on the complexity of the described path through the tree. Tables

5.8 and 5.9 show the distribution of ruleset sizes with Laplace values $\geqslant 0.95$ that were derived from five trees trained on 20,000 training samples each.

			Conditions per ruleset										
	Class label	Pathway	1	2	3	4	5	6	7	8	9	10	Sum
Stability prediction	Stability	linear	7	58	167	160	53	8	0	0	0	0	453
		branched	12	40	111	147	70	15	0	0	0	0	395
		cyclic	7	46	132	107	34	6	0	0	0	0	332
	Instability	linear	0	0	11	27	20	8	2	0	0	0	68
		branched	0	1	6	30	27	6	3	0	0	0	73
		cyclic	0	0	6	25	18	8	0	0	0	0	57
Oscillation prediction	Oscillating	linear	0	0	0	0	2	38	56	21	12	3	132
		branched	0	0	4	21	78	91	60	13	3	0	270
		cyclic	0	0	0	2	12	20	10	7	2	1	54
	Not oscillating	linear	4	53	163	280	186	91	17	0	0	0	794
		branched	6	16	125	206	110	44	12	6	0	0	525
		cyclic	9	51	198	244	131	28	4	0	0	0	665

Table 5.8.: Distribution of elasticity ruleset sizes with Laplace value $\geqslant 0.95$. The numbers have been summarized over all the five trees that were constructed for each classification problem.

The numbers and sizes of rulesets differed significantly among the classification problems as well as with the type of feature used. Using elasticities as input features led to large numbers of rulesets for stable as well as for non-oscillatory steady states, whereas rulesets for instabilities and oscillations were observed less often (Table 5.8). While it was possible to formulate reliable stabilizing or non-oscillatory patterns with only one condition each, rulesets of this size did not emerge for instabilities or for oscillatory steady states. This observation showed that forcing the system to exhibit either an unstable steady state, or oscillatory behaviour in the neighbourhood of a steady state, requires a more complex interplay of several network components than would be the case for the respective opposite behaviours.

Stability classification based on FCCs produced exactly one ruleset per tree for the linear pathway (Table 5.9). As shown in the previous section, these simple rules

were sufficient to enable optimal separation of both classes (see also Tables C.1b and C.2b in Appendix C.1.1). For the cyclic pathway, such simple and straightforward boundaries could not be derived. When classifying data obtained from the branched pathway steady state, only two rulesets could be obtained out of five trees with sufficiently large Laplace values. For the cyclic pathway, the trees produced several rulesets of varying complexities. Only for the branched pathway, reliable FCC based rulesets could be derived for oscillatory behaviour. This observation, namely that it was possible to formulate reliable criteria in terms of flux control to ensure oscillatory trajectories, hinted at the possibility that oscillations generally emerged more easily in the branched pathway structure.

			Conditions per ruleset								
	Class label	Pathway	1	2	3	4	5	6	7	8	Sum
Stability prediction	Stability	linear	5	0	0	0	0	0	0	0	5
		branched	0	2	0	0	0	0	0	0	2
		cyclic	0	6	9	14	7	3	0	0	39
	Instability	linear	5	0	0	0	0	0	0	0	5
		branched	5	7	2	0	0	0	0	0	14
		cyclic	6	0	1	0	0	0	0	0	7
Oscillation prediction	Oscillating	linear	0	0	0	0	0	0	0	0	0
		branched	2	8	17	39	10	1	1	1	78
		cyclic	0	0	0	0	0	0	0	0	0
	Not oscillating	linear	8	37	24	5	0	0	0	0	74
		branched	0	18	21	18	8	2	1	0	68
		cyclic	0	1	1	1	1	0	0	0	4

Table 5.9.: Distribution of FCC ruleset sizes with Laplace value $\geqslant 0.95$. The numbers have been summarized over all the five trees that were constructed for each classification problem.

5.4.3. Detecting the most informative elasticities

As described in the previous section, the decision trees trained using elasticities produced large numbers of rulesets that could contain many conditions each. In

order to get a first impression of the information contained in these rulesets, they were analysed using summary statistics. The focus of this analysis was two-fold:

1. Investigate the abundance of each elasticity per ruleset in order to determine the most informative candidates.

2. Compare the types of quantitative thresholds derived for each feature. In particular, the proportion of lower and upper bounds imposed on each elasticity in rulesets associated with a particular class was computed. This provided a first impression about whether low or high values were preferred for this elasticity in the given class.

Linear pathway Table 5.10 shows the frequencies of elasticities in rulesets with Laplace values $\geqslant 0.95$ obtained for the linear pathway, together with the proportion of lower bounds associated with each elasticity in percent. For example, the first entry '0.38 (98.25 %)' shows that elasticity $\varepsilon_{S_1}^{v_1^-}$ was present on average 0.38 times per ruleset, and that 98.25% of these rulesets contained conditions constraining it by a lower bound. Stability-associated rulesets contained conditions for a wide variety of elasticities, where no elasticity occurred more than 0.81 times per ruleset. In contrast, almost each instability-associated ruleset contained exactly one condition using the elasticities $\varepsilon_{S_4}^{v_2^+}$ and $\varepsilon_{S_4}^{v_5^+}$ (Table 5.10). In general, the types of boundaries imposed on each elasticity tended to be opposite between stability- and instability-associated rulesets.

The distribution of upper and lower bounds indicated that stability generally required fast perturbation responses in the influx reaction r_1 and in the efflux reaction r_5, thus enabling fast transport into and out of the system. In contrast, reactions located inside the pathway tended to adapt slowly to changes, thereby dampening perturbations.

The positive feedback strength encoded by $\varepsilon_{S4}^{v_2^+}$ was exclusively restricted by upper bounds in stability-associated rulesets. This indicated that it was favourable for stability if the feedback remained small, so that perturbations were not amplified within the system. On the other hand, $\varepsilon_{S4}^{v_2^+}$ was often restricted by lower bounds in instability-associated rulesets, whereas conditions on $\varepsilon_{S_4}^{v_5^+}$ often applied upper bounds. This showed that instabilities tended to emerge when the strength of the activating feedback term was large, accompanied by a slow response of the efflux reaction r_5 to perturbations.

When considering rulesets for oscillations in the linear pathway, we could detect four elasticities ($\varepsilon^{v_2^+}_{S_4}$, $\varepsilon^{v_3^+}_{S_2}$, $\varepsilon^{v_4^+}_{S_3}$, $\varepsilon^{v_5^+}_{S_4}$) occurring with frequencies close to or greater than 1 per ruleset. These elasticities described the control of r_2, r_3, r_4 and r_5 by their substrates S_2, S_3 and S_4. This indicated that oscillations required the fine-tuned interplay between these pathway positions and that conditions for oscillations tended to be more complex than those for instabilities, which were dominated by only the two elasticities $\varepsilon^{v_2^+}_{S_4}$ and $\varepsilon^{v_5^+}_{S_4}$. Another difference to stability/instability conditions was that the majority of oscillation-associated rulesets tended to impose lower bounds on the affected elasticities. This showed that oscillations generally required fast responses to perturbations in the majority of the system components, whereas stability or instability required more differentiated combinations of elasticity values.

Elasticity	Stability	Instability	Oscillating	Not oscillating
$\varepsilon^{v_1^-}_{S_1}$	0.38 (98.25 %)	0.62 (4.76 %)	0.86 (92.11 %)	0.58 (5.03 %)
$\varepsilon^{v_2^+}_{S_1}$	0.55 (99.19 %)	0.56 (0 %)	0.16 (80.95 %)	0.15 (64.46 %)
$\varepsilon^{v_2^+}_{S_4}$	0.68 (0 %)	**1.04** (95.77 %)	**1.08** (92.96 %)	0.54 (21.65%)
$\varepsilon^{v_2^-}_{S_2}$	0.62 (0 %)	0.71 (97.92 %)	0.37 (83.67 %)	0.47 (14.48 %)
$\varepsilon^{v_3^+}_{S_2}$	0.17 (1.3 %)	0.22 (100 %)	**0.98** (100 %)	0.38 (5.94 %)
$\varepsilon^{v_3^-}_{S_3}$	0.16 (2.78 %)	0.13 (88.89 %)	0.85 (93.75 %)	0.60 (3.16 %)
$\varepsilon^{v_4^+}_{S_3}$	0.07 (13.33 %)	0.10 (85.71 %)	**1.00** (98.48 %)	0.41 (3.7 %)
$\varepsilon^{v_4^-}_{S_4}$	0.05 (4.35 %)	0.03 (50 %)	0.73 (95.83 %)	0.54 (2.82 %)
$\varepsilon^{v_5^+}_{S_4}$	0.81 (100 %)	**1.04** (4.23 %)	**1.07** (93.62 %)	0.52 (32.60 %)

Table 5.10.: Frequency of elasticities in ruleset conditions for the linear pathway with Laplace values $\geqslant 0.95$. Values in brackets refer to the percentage of lower bounds associated with each elasticity (green: $\geqslant 90\%$ lower bounds; red: $\leqslant 10\%$ lower bounds). For example, the first entry (0.38 (98.25 %)) shows that elasticity $\varepsilon^{v_1^-}_{S_1}$ was present on average 0.38 times per ruleset, and that 98.25% of these rulesets contained conditions constraining it by a lower bound. Consequently, the elasticity was required to take on rather high values by most rulesets. Elasticities present at least 0.9 times per ruleset on average are marked in **bold**.

Branched pathway Table 5.11 shows the frequencies of elasticities in ruleset for the branched pathway with Laplace values $\geqslant 0.95$. Like for the linear pathway, no subset of particularly essential elasticities could be detected that were prominent in every stability-associated ruleset. The most abundant elasticities were $\varepsilon_{S4}^{v_6^+}$, $\varepsilon_{S4}^{v_2^+}$ and $\varepsilon_{S1}^{v_2^+}$, which both occurred in approximately 60 % of all rulesets. Elasticities belonging to reactions r_4 and r_5, which formed the branch not involved in the feedback loop, occurred only rarely per ruleset. These differences showed that, while it was favourable for stability to control the positive feedback term tightly, the remaining reactions did not need not be constrained as strongly.

Elasticity	Stability	Instability	Oscillating	Not oscillating
$\varepsilon_{S_1}^{v_1^-}$	0.09 (94.12 %)	0.04 (0 %)	0.17 (24.44 %)	0.18 (92.55 %)
$\varepsilon_{S_1}^{v_2^+}$	0.57 (99.55 %)	0.75 (1.82 %)	0.60 (98.14 %)	0.30 (18.12 %)
$\varepsilon_{S_4}^{v_2^+}$	0.57 (0 %)	**0.99** (98.61 %)	**1.24** (80.36 %)	0.79 (0 %)
$\varepsilon_{S_2}^{v_2^-}$	0.54 (0 %)	0.62 (93.33 %)	**1.20** (78.70 %)	**0.92** (1.66 %)
$\varepsilon_{S_2}^{v_3^+}$	0.51 (99.50 %)	0.51 (2.70 %)	0.11 (32.26 %)	0.08 (92.68 %)
$\varepsilon_{S_3}^{v_3^-}$	0.05 (85.71 %)	0.01 (0 %)	0.06 (13.33 %)	0.04 (76.19 %)
$\varepsilon_{S_2}^{v_4^+}$	0.33 (0 %)	0.38 (96.43 %)	0.71 (94.79 %)	0.50 (3.44 %)
$\varepsilon_{S_4}^{v_4^-}$	0.33 (0.76 %)	0.25 (94.44 %)	0.37 (86.00 %)	0.41 (4.19 %)
$\varepsilon_{S_3}^{v_5^+}$	0.10 (95.00 %)	0.08 (0 %)	0.17 (11.11 %)	0.13 (87.88 %)
$\varepsilon_{S_4}^{v_6^+}$	0.59 (100 %)	**0.92** (2.99 %)	**1.24** (41.32 %)	0.82 (99.53 %)

Table 5.11.: Frequency of elasticities in ruleset conditions for the branched pathway with Laplace values $\geqslant 0.95$. Values in brackets refer to the percentage of lower bounds associated with each elasticity (green: $\geqslant 90\%$ lower bounds; red: $\leqslant 10\%$ lower bounds). Elasticities present at least 0.9 times per ruleset on average are marked in **bold**.

With exception of $\varepsilon_{S1}^{v_2^+}$, the elasticities belonging to r_2 and r_4, which were the reactions involved in the feedback term, were almost exclusively restricted to low values. In contrast, elasticities associated with the remaining reactions r_1, r_3, r_5 and r_6 tended to be restricted to high values. These reactions either represented influx reactions, efflux reactions or reactions of the branch not involved in the feedback term. This showed that in stable steady states, we could generally assume

a fast response in the influx and efflux reactions to perturbations, while reaction r_4, which was responsible for producing the potentially destabilizing activator S_4, needed to respond slowly. Reversely, a slow response of the influx and efflux reactions, together with the fast propagation of perturbations by the positive feedback, was the major source for instabilities.

We found three elasticities with particularly high frequencies per oscillation-associated ruleset. This is one elasticity less than detected for the linear pathway. The detected elasticities were either related to the regulated reaction r_2 or to the two efflux reaction r_6. In contrast to the linear pathway, fewer elasticities were constrained by lower bounds in oscillation-associated rulesets. This indicated that oscillations in the branched pathway did not require perturbation responses that were as fast as in the linear system. This coincided with the earlier observation that the emergence of oscillations was much more likely in the branched pathway than in the linear pathway (see Table 5.4 and Figure 5.3 in Section 5.2).

Cyclic pathway Table 5.12 shows the frequency of elasticities in ruleset conditions for the cyclic pathway with Laplace values $\geqslant 0.95$. The most frequent elasticities in stabilizing rulesets were those associated with the regulated reaction r_2 ($\varepsilon_{S_1}^{v_2^+}$, $\varepsilon_{S_3}^{v_2^+}$, $\varepsilon_{S_2}^{v_2^-}$), as well as those describing the substrate influence of S_3 on its downstream reactions ($\varepsilon_{S_3}^{v_4^+}$, $\varepsilon_{S_3}^{v_6^+}$). Similar to the linear and branched pathway, stability-associated rulesets imposed lower bounds on the influx and efflux reactions v_1 and v_6, indicating that stability required fast responses to perturbations at the system boundaries. The feedback term was exclusively restricted by upper bounds, again emphasizing the requirement of strict control of the feedback strength to maintain stability.

As previously observed for the linear and branched pathway, elasticities were less homogeneously distributed in oscillatory conditions than in stability conditions. Again, the most prominent elasticities were associated with the regulated reaction r_2, as well as with the efflux of the regulator out of the system (r_3 and S_3 in this case). The majority of the most abundant elasticities was constrained by lower bounds hinting at the necessity of fast perturbation responses in both reactions.

Summary: Most informative elasticities When investigating the stabilizing rulesets derived for all three pathway structures, we could identify sets of elasticities that were more abundant than others. However, none of these elasticities was present in every ruleset. Instead, elasticities could occur in some rulesets while

Elasticity	Stability	Instability	Oscillating	Not oscillating
$\varepsilon_{S_1}^{v_1^-}$	0.06 (90.48 %)	0.07 (0 %)	0.11 (83.33 %)	0.16 (4.72 %)
$\varepsilon_{S_1}^{v_2^+}$	0.47 (100 %)	0.58 (3.03 %)	0.07 (50 %)	0.09 (87.93 %)
$\varepsilon_{S_3}^{v_2^+}$	0.57 (0 %)	**0.98** (98.21 %)	**1.22** (81.82 %)	0.61 (3.22 %)
$\varepsilon_{S_2}^{v_2^-}$	0.64 (0 %)	0.81 (97.83 %)	**1.22** (81.82 %)	0.76 (6.93 %)
$\varepsilon_{S_2}^{v_3^+}$	0.16 (0 %)	0.14 (100 %)	**0.98** (96.23 %)	0.45 (1.6 %)
$\varepsilon_{S_3}^{v_3^-}$	0.09 (9.68 %)	0.07 (100 %)	0.50 (85.19 %)	0.38 (2.76 %)
$\varepsilon_{S_3}^{v_4^+}$	0.62 (100 %)	0.84 (4.17 %)	0.89 (18.75 %)	0.76 (99.80 %)
$\varepsilon_{S_4}^{v_4^-}$	0.06 (26.32 %)	0 (-)	0.24 (7.69 %)	0.19 (88.19 %)
$\varepsilon_{S_4}^{v_5^+}$	0.02 (75 %)	0.04 (100 %)	0.04 (50 %)	0.03 (75 %)
$\varepsilon_{S_1}^{v_5^-}$	0.02 (14.29 %)	0.02 (100 %)	0.02 (100 %)	0.03 (47.06 %)
$\varepsilon_{S_3}^{v_6^+}$	0.67 (100 %)	**0.95** (0 %)	**1.04** (87.50 %)	0.36 (44.81 %)

Table 5.12.: Frequency of elasticities in ruleset conditions for the cyclic pathway with Laplace values $\geqslant 0.95$. Values in brackets refer to the percentage of lower bounds associated with each elasticity (green: $\geqslant 90\%$ lower bounds; red: $\leqslant 10\%$ lower bounds). Elasticities present at least 0.9 times per ruleset on average are marked in **bold**.

being absent in others. This emphasized that stabilizing sites were distributed over the whole system and that unfavourable values in one elasticity could be compensated by specific combinations of values at other positions. This explains the high abundance of stabilizing models observed in the Monte Carlo experiment (see Figure 5.3). The types of constraints (i.e. upper or lower bounds) imposed on elasticities in stabilizing conditions indicated that it was favourable for stability if the influx and efflux reactions responded quickly to concentration changes, while reactions within the system were preferred to respond slowly, thus preventing perturbations from being spread internally through the system.

In contrast to stabilizing rulesets, oscillatory rulesets clearly emphasized the importance of particular subsets of elasticities in each system. These elasticities were present at least once per ruleset for a given pathway. This showed that oscillations relied on combinations of elasticity values at a few specific network positions. This also explained why oscillations generally emerged less frequently than stability in

the observed steady states during Monte Carlo sampling. Almost all of these 'important' elasticities were associated with reactions or metabolites involved in the activating feedback term and were constrained by lower bounds. This indicated that the feedback term played a crucial role in maintaining oscillations in the given example pathways, and that systems with high feedback strength were more prone to exhibit oscillatory behaviour.

5.4.4. Ruleset examples

After the statistical analysis of the elasticities involved in decision tree rulesets for various types of dynamic behaviours, we now take a closer look at some particular examples of such rulesets. We start with rulesets that describe conditions under which the system is likely to be stable. For the ease of biological interpretability, we will only look at the smallest rulesets that have been derived for each pathway. At the end of this section, we will present one example of a ruleset that describes oscillatory conditions.

Stability-associated rulesets derived for the linear pathway

As previously shown in Section 5.4.2 (Table 5.8) the five trees trained on linear pathway data produced seven elasticity rulesets composed of a single condition each. Among these rulesets, four contained conditions that used elasticity $\varepsilon_{S_4}^{v_2^+}$, and the remaining three contained conditions that used $\varepsilon_{S_4}^{v_5^+}$. A summary of the associated thresholds and ruleset performances is shown in Table 5.13.

The conditions for both elasticities were well reproducible, as indicated by small standard deviations of the derived thresholds. The first four rulesets indicated that stability was secured in almost every model in which the regulatory feedback strength $\varepsilon_{S_4}^{v_2^+}$ was smaller than or equal to 0.26. Alternatively, $\varepsilon_{S_4}^{v_5^+}$, which encoded the strength by which the regulatory metabolite S_4 amplified its own efflux out of the system, was required to be larger than 0.61. Both scenarios aimed at restricting the accelerating influence of S_4 on the flux through the pathway, confirming that the positive feedback loop posed a threat on stability if it became too prominent.

The small ruleset size (only one condition necessary per ruleset) indicated that fulfilling either of the depcited conditions was sufficient for maintaining stability. This was confirmed when looking at the true elasticities for the given steady state (Table 5.1 in Section 5.1): the feedback strength $\varepsilon_{S_4}^{v_2^+}$ computed using the underlying kinetic model was equal to 0.55 and did not satisfy the first condition.

Number of rulesets	Summarized conditions	Laplace values	Equivalent kinetic conditions
4	$\varepsilon_{S_4}^{v_2^+} \leqslant 0.26 \pm 0.03$	0.9989 ± 0.0013	$\mathrm{V}_{\mathrm{max}}^{\mathrm{v}_2^{+/-}} \leqslant 0.61 \pm 0.03$ $\mathrm{K_A} \leqslant 0.29 \pm 0.05$
3	$\varepsilon_{S_4}^{v_5^+} > 0.61 \pm 0.01$	0.9998 ± 0.0001	$\mathrm{K}_{\mathrm{M}}^{\mathrm{v}_2^+} > 1.21 \pm 0.03$ $\mathrm{K}_{\mathrm{M}}^{\mathrm{v}_2^-} > 22.77 \pm 0.23$ $\mathrm{V}_{\mathrm{max}}^{\mathrm{v}_2^+} > 0.53 \pm 0.01$

Table 5.13.: Summary of stabilizing elasticity rulesets (one condition each) in the linear pathway. It is sufficient for maintaining stability to either restrict the strength of the positive feedback (first row), or to ensure fast perturbation responses of the efflux out of the system (second row).

However, $\varepsilon_{S_4}^{v_5^+}$ was equal to 0.79 and fulfilled the second condition. The kinetic parameter thresholds derived from the elasticity threshold were also met by the kinetic model (see Section 3.1.1 for details). As a result, the observed steady state was stable (Table 5.3).

Training based on FCCs produced minimally sized trees that contained exactly one ruleset with one condition each. Among the resulting five rulesets, three contained conditions for the coefficient $C_{E_3}^{J^t}$, and two for $C_{E_5}^{J^t}$ (Table 5.14). The negative threshold for $C_{E_3}^{J^t}$ showed that reaction r_3 could inhibit flux through the system for specific combinations of elasticities. The corresponding rulesets showed that restricting this FCC to values close to or above 0 is an near-optimal criterion for ensuring stability. A closer look at the number of true positives showed that they equalled the total number of stable models in the test data, leading to a recall of 1. Only few false positives (between 2 and 8 per tree) were observed, possibly because the resolution of the training data sets was not refined enough to derive thresholds that optimally separated the test data. We can summarize that instabilities could only arise due to negative impact of r_3 on the total flux, whereas stability was guaranteed if the impact of r_3 was close to 0 or positive. The rulesets in the remaining two trees referred to the acceleration of flux caused by increasing the rate of efflux reaction r_5. They indicated that stability was ensured as long as the influence of the efflux reaction on total flux was low, whereas instabilities arose if it got too large.

Number of rulesets	Summarized conditions	Laplace values
3	$C^{J^t}_{E_3} > -0.0115 \pm 0.0008$	0.9999 ± 0.0001
2	$C^{J^t}_{E_5} \leqslant \ 0.4116 \pm 0.0387$	1.0000 ± 0.0000

Table 5.14.: Summary of stabilizing FCC rulesets (one condition each) in the linear pathway. J^t refers to the steady state flux through the pathway. It is sufficient for maintaining stability to either restrict the negative impact of r_3 on total flux (first row), or to ensure strong enough acceleration of total flux by the efflux reaction r_5 (second row).

Stability-associated rulesets derived for the branched pathway

Like the linear pathway, rulesets of size one computed for the branched pathway differed widely in the elasticities used in their conditions (Table 5.15). Most rulesets referred to either the feedback strength $\varepsilon^{v_2^+}_{S_4}$ (7 rulesets), or the strength of efflux $\varepsilon^{v_6^+}_{S_4}$ (3 rulesets). Additionally, one tree contained one ruleset that constrained the effect of S_2 on the branching reaction r_4 to small values. Another tree contained a ruleset constraining the feedback strength of r_2 by its product S_2.

Thresholds on the feedback strength $\varepsilon^{v_2^+}_{S_4}$ could be assigned to two clusters, both restricting the elasticity with upper bounds of different strictness. This separation was a result of the differences in resolution in this elasticity in the randomly sampled training data sets. Even when only focusing on the stricter version (first row), the corresponding thresholds on $\varepsilon^{v_2^+}_{S_4}$ were less strict than those derived for the linear pathway. This showed that the steady state observed in the branched pathway topology could buffer perturbations more efficiently, even for larger feedback strengths. In contrast, the lower bound on efflux activation $\varepsilon^{v_6^+}_{S_4}$ was stricter than in the linear pathway, emphasizing that this network position required tighter control (third row). The condition affecting $\varepsilon^{v_4^+}_{S_2}$ (fourth row) indicated that ensuring small responses to perturbations in the entry point to the branch that exhibited the positive feedback (reaction r_4), could also restrict the feedback strength sufficiently to ensure stability.

When computing the Michaelis constants required to fulfil the condition $\varepsilon^{v_2^-}_{S_2} \leqslant 0.43$ (last row of Table 5.15) for the given steady state, they turned out to be negative. The reason for this result was that this condition imposed an upper bound on the elasticity which only allowed small values. However, according to the filterting criterion derived in equation (3.1), the Michaelis constants could only be positive if $\varepsilon^{v_2^+}_{S_1} + \varepsilon^{v_2^-}_{S_2} > 1$. Consequently, $\varepsilon^{v_2^+}_{S_1}$ must be at least greater

Number of rulesets	Summarized conditions	Laplace values	Equivalent kinetic conditions
3	$\varepsilon_{S_4}^{v_2^+} \leqslant 0.41 \pm 0.02$	0.9997 ± 0.0002	$V_{max}^{v_2^{+/-}} \leqslant 0.44 \pm 0.02$ $K_A^{v_2^{+/-}} \leqslant 0.24 \pm 0.02$
4	$\varepsilon_{S_4}^{v_2^+} \leqslant 0.64 \pm 0.01$	0.9589 ± 0.0045	$V_{max}^{v_2^{+/-}} \leqslant 0.7 \pm 0.03$ $K_A^{v_2^{+/-}} \leqslant 0.59 \pm 0.04$
3	$\varepsilon_{S_4}^{v_6^+} > 0.74 \pm 0.08$	0.9971 ± 0.0023	$K_M^{v_6^+} > 1.02 \pm 0.44$ $K_M^{v_6^-} > 24.48 \pm 2.60$ $V_{max}^{v_6^+} > 0.41 \pm 0.13$
1	$\varepsilon_{S_2}^{v_4^+} \leqslant 0.18$	0.9674	$K_M^{v_4^+} \leqslant 0.06$ $K_M^{v_4^-} \leqslant 5.60$ $V_{max}^{v_4^+} \leqslant 0.09$
1	$\varepsilon_{S_2}^{v_2^-} \leqslant 0.43$	0.9987	$K_M^{v_2^+} \leqslant -0.87$ $K_M^{v_2^-} \leqslant -0.06$ $V_{max}^{v_2^-} \leqslant 0.01$

Table 5.15.: Summary of stabilizing elasticity rulesets (one condition each) in the branched pathway.

than $1 - 0.43 = 0.57$. Instead, Table 5.1 shows that for the given steady state, $\varepsilon_{S_1}^{v_2^+} = 0.35$. The ruleset therefore only affected those SK-models for which the sampled elasticities fulfilled the described condition, which was not the case for the true steady state elasticities.

Like the linear pathway, training based on FCCs on branched pathway data produced minimally sized trees with one ruleset each. Four of these rulesets contained exactly one condition each, whereas one ruleset contained two conditions (Table 5.9). All rulesets of size one required tight control of the lower branch by the entry point reaction of the competing branch, r_3 (Table 5.16). As indicated by the negative upper bound, stability was ensured in almost all models in which an increase in the rate of r_3 led to a reduction in steady state flux through the lower branch, which contained the feedback term. On the other hand, in unstable models, this competition mechanism seemed to be disturbed so that perturbations

leading to an increase in the rate of r_3 also strengthened the positive feedback, which eventually would lead to an amplification of the flux through the whole system.

Number of rulesets	Summarized conditions	Laplace values
1	$C_{E_1}^{J^{b_1}} \leqslant 7.2942$ $C_{E_3}^{J^{b_2}} \leqslant -0.0535$	0.9628
1	$C_{E_1}^{J^{b_2}} \leqslant 2.5686$ $C_{E_5}^{J^{b_2}} \leqslant -0.01800$	0.9968

Table 5.16.: Summary of stabilizing FCC rulesets (two conditions each) in the branched pathway. J^{b_1} and J^{b_2} refer to the steady state flux through branch 1 and 2 (see Section 5.1 for details).

Stability-associated rulesets derived for the cyclic pathway

Table 5.17 summarizes the smallest stability-associated elasticity rulesets obtained for the cyclic pathway. One ruleset described the inhibition of the regulated reaction r_2 by its product S_2. Like previously observed for the branched pathway steady state, the corresponding kinetic parameters were negative so that this condition was not realizable without modifying other elasticities in the system.

Five conditions (one per tree) were derived for the activating influence of regulator S_3 on r_2, all of them restricting it with upper bounds. When looking at the threshold values, four of these conditions formed a cluster around 0.66, whereas one tree applied a stricter upper bound of 0.49. The equivalent conditions on the kinetic parameters of reaction r_2 restricted them to values below those used in the original kinetic model (1 for each parameter). The stricter ruleset (second row) constrained the corresponding kinetic parameters to a distinctly stronger extent than the other four (third row). Since this ruleset also led to the highest Laplace values, we can conclude that the system became more likely to be stable for smaller feedback strength. This observation coincided with those made in the linear and branched pathways. Interestingly, the kinetic parameters of the original kinetic model used for steady state computation lay outside of the 'stable' value ranges derived from these thresholds. This indicates that these rulesets were not comprehensive and that stabilitiy coild also be ensured for other combinations of

Number of rulesets	Summarized conditions	Laplace values	Equivalent kinetic conditions
1	$\varepsilon^{v_2^-}_{S_2} \leqslant 0.39$	0.9992	$K_M^{v_2^+} \leqslant -2.18$ $K_M^{v_2^-} \leqslant -0.33$ $V_{max}^{v_2^-} \leqslant 0.02$
1	$\varepsilon^{v_2^+}_{S_3} \leqslant 0.49$	0.9992	$V_{max}^{v_2^{+/-}} \leqslant 0.61$ $K_A \leqslant 0.43$
4	$\varepsilon^{v_2^+}_{S_3} \leqslant 0.66 \pm 0.01$	0.9647 ± 0.0093	$V_{max}^{v_2^{+/-}} \leqslant 0.91 \pm 0.04$ $K_A \leqslant 0.87 \pm 0.05$
1	$\varepsilon^{v_6^+}_{S_3} > 0.80$	0.9992	$K_M^{v_6^+} > 1.73$ $K_M^{v_6^-} > 27.38$ $V_{max}^{v_6^+} > 0.63$

Table 5.17.: Summary of stabilizing elasticity rulesets (one condition each) in the cyclic pathway.

elasticities. Instead, they fulfilled the requirements of the last ruleset, which applied a lower bound on the efflux of activator S_3. Indeed, the original steady state elasticity had a value of 0.87 which exceeded the threshold of 0.8 (Table 5.1). We can conclude that the computed steady state was stable in the given kinetic model because of the fast response of this reaction to perturbations, enabling them to be carried out of the system before they led to an amplification of flux due to the feedback term.

When training on cyclic pathway data, it was not possible to find stability-associated FCC rulesets with only a single condition that offered a classification performance as optimal as for the linear and branched pathway. Instead, seven rulesets were derived of which six contained two conditions and one contained three conditions. Each of the FCC rulesets of size two contained one condition for the control of the overall flux through the system by reaction r_5 ($C^{J^t}_{E_5}$), as well as one additional condition referring to the control exhibited by reaction r_1 (Table 5.18).

Depending on the type of ruleset, the additional condition referred to either the control that influx reaction r_1 exhibited on the total system flux, or on the flux through the upper arm of the cycle. From the derived thresholds we can

Number of rulesets	Summarized conditions	Laplace values
4	$C^{J^t}_{E_1} > 0.2212 \pm 0.0911$ $C^{J^t}_{E_5} > -0.1106 \pm 0.0162$	0.9740 ± 0.0123
2	$C^{J^{c_1}}_{E_1} > 0.1700 \pm 0.1443$ $C^{J^t}_{E_5} > -0.0896 \pm 0.0243$	0.9822 ± 0.0107

Table 5.18.: Summary of stabilizing FCC rulesets (two conditions each) in the cyclic pathway. J^{c_1} refers to the steady state flux through the upper arm of the cycle, whereas J^t refers to the flux through the whole pathway.

conclude that it was favourable for stability if the competing effect of the lower cycle arm (which included reaction r_5) on the total flux was restricted, so that the corresponding FCC did not fall too far below zero. Additionally, the influx reaction r_1 needed to sufficiently accelerate flux through the upper arm for the system to be stable.

Oscillation-associated rulesets derived for the branched pathway

Oscillation-associated rulesets were larger than those associated with stability, hinting that more fine-tuned criteria were necessary to produce oscillatory behaviour in the example systems. As an example, Table 5.19 shows a summary of elasticity-based rulesets of size three for the branched pathway.

Number of rulesets	Summarized conditions	Laplace values	Equivalent kinetic conditions
4	$\varepsilon^{v_2^-}_{S_2} > 0.9 \pm 0.1$ $\varepsilon^{v_2^+}_{S_1} > 0.76 \pm 0.14$ $\varepsilon^{v_2^+}_{S_4} > 0.24 \pm 0.1$	0.9618 ± 0.0104	$\mathrm{K}_{\mathrm{M}}^{\mathrm{v}_2^+} > 12.65 \pm 6.46$ $\mathrm{K}_{\mathrm{M}}^{\mathrm{v}_2^-} > 3.71 \pm 3.11$ $\mathrm{V}_{\mathrm{max}}^{\mathrm{v}_2^+} > 0.97 \pm 0.54$ $\mathrm{V}_{\mathrm{max}}^{\mathrm{v}_2^-} > 0.03 \pm 0.02$ $\mathrm{K_A} > 0.11 \pm 0.07$

Table 5.19.: Summary of oscillation-associated elasticity rulesets (three conditions each) for the branched pathway.

All four rulesets contained conditions for all elasticities associated with the reg-

ulated reaction r_2. All elasticities were restricted by lower bounds, with varying degrees of strictness of the associated thresholds. This coincided with the observations described previously in Section 5.4.2, which also indicated that oscillations were favoured if the individual reactions responded quickly to perturbations.

5.5. Analysing stability and oscillation conditions by RVMs

5.5.1. Classification performance

In order to assess the classification performance of RVMs, and to compare it to the performance of the decision tree classifiers, RVMs were trained on the same balanced datasets as the decision trees. Whereas decision trees had been trained up to a sample size of 20,000 samples, the longer runtime of the RVM algorithm made it necessary to stop after 1000 samples. Despite the strongly reduced sample size, training and test errors were comparable or even better than those provided by the decision trees for 20,000 samples (see summary in Table 5.20 as well as a detailed list in Appendix C.2.1). The only exceptions were observed in the FCC based classification of the cyclic pathway models.

We can conclude that in most scenarios RVMs were able to capture the class differences in the training data better than the decision trees. In other words, they were less prone to bias than the decision tree classifiers. The low test errors indicated that the derived decision functions were also well generalizable and usually not prone to overfitting. Only the elasticity based classification of oscillatory steady states revealed significantly higher test errors compared to the training errors (see Figure C.1 - C.3 in Appendix C.2.1). However, they still outperformed the decision trees. These good classification results are especially interesting considering the strongly reduced number of training samples necessary to achieve the observed performance.

When classifying based on FCCs, RVM performance for 1000 training samples was slightly below that obtained by decision trees for 20,000 training samples. Similar to decision trees, RVMs performed excellent when predicting stability of the linear and branched pathway, whereas stability of the cyclic pathway was more challenging. As shown in Figure C.3, in the latter case, errors started with 50% for small sample sizes, showing that it was not possible for the RVMs to deduce

Pathway	Classifier	Elasticities		FCCs	
		Stability	Oscillations	Stability	Oscillations
linear	DT (20,000 samples)	0.011 +- 0.001	0.039 +- 0.001	0 +- 0	0.166 +- 0.004
	RVM (1000 samples)	0.002 +- 0.001	0.015 +- 0.004	0 +- 0.001	0.16 +- 0.006
branched	DT (20,000 samples)	0.009 +- 0.001	0.034 +- 0.002	0.002 +- 0	0.059 +- 0.002
	RVM (1000 samples)	0.001 +- 0.001	0.011 +- 0.004	0 +- 0	0.061 +- 0.006
cyclic	DT (20,000 samples)	0.007 +- 0	0.025 +- 0.002	0.1 +- 0.002	0.198 +- 0.003
	RVM (1000 samples)	0.004 +- 0.002	0.009 +- 0.007	0.126 +- 0.01	0.212 +- 0.018

(a) Classification errors determined on the training data.

Pathway	Classifier	Elasticities		FCCs	
		Stability	Oscillations	Stability	Oscillations
linear	DT (20,000 samples)	0.029 +- 0.001	0.094 +- 0.001	0 +- 0	0.179 +- 0.001
	RVM (1000 samples)	0.014 +- 0.002	0.049 +- 0.003	0 +- 0	0.182 +- 0.003
branched	DT (20,000 samples)	0.024 +- 0.001	0.087 +- 0.002	0.004 +- 0	0.071 +- 0.001
	RVM (1000 samples)	0.008 +- 0.001	0.045 +- 0.001	0.006 +- 0.002	0.069 +- 0.001
cyclic	DT (20,000 samples)	0.02 +- 0.001	0.059 +- 0.001	0.113 +- 0.001	0.209 +- 0.001
	RVM (1000 samples)	0.008 +- 0	0.043 +- 0.003	0.134 +- 0.007	0.215 +- 0.003

(b) Classification errors determined on the test data.

Table 5.20.: Comparison between the BERs obtained by decision tree (DT) classification (20,000 training samples) and RVMs (1000 training samples). Despite the much smaller training data, the RVMs exceed the decision trees in most cases.

an appropriate decision function from small training datasets. Although similar problems had been encountered by decision trees, a biological interpretation of at least parts of the deduced trees had still been possible by focussing on the best performing rulesets. RVMs do not offer this possibility and therefore do not ensure that the derived RVs are indeed good representatives for each class. Because of this difficulty, the following analysis will only focus on prediction based on elasticities.

5.5.2. Decision functions and relevance vectors

When training the classifier to distinguish stable from unstable models, the majority of the returned RVs represented stable models. In contrast, fewer RVs were produced for unstable models in each training run (Table 5.21). This confirmed the previous observations that stable models allow a larger variety of elasticity combinations than unstable models.

Classification of oscillatory versus non-oscillatory models produced larger num-

bers of RVs than stability prediction. This showed that oscillations could be produced by very diverse combinations of elasticities, confirming once more that they are a more complex phenomenon than stability.

Each of the derived RVs represented one particular instance of a 'typical' class member. Because RVMs are designed to obtain sparse solutions, we can assume that the RVs differ strongly within each class in order to cover its region in the feature space as widely as possible. This assumption was confirmed by high standard deviations obtained when summarizing all RVs representing a particular class. To illustrate this point, Table 5.22 shows a summary of the RVs derived for the linear pathway. Because of the large spread within each class, we can conclude that stability or oscillatory behaviour did not depend on any single elasticity, but rather were the results of diverse combinations of elasticities.

	Class label	Pathway	Run 1	Run 2	Run 3	Run 4	Run 5	Sum
Stability prediction	Stability	linear	10	7	5	10	9	41
		branched	4	4	4	5	4	21
		cyclic	7	4	4	3	4	22
	Instability	linear	3	4	5	2	3	17
		branched	3	2	2	3	3	13
		cyclic	1	4	3	3	2	13
Oscillation prediction	Oscillating	linear	11	11	11	13	12	58
		branched	15	10	16	10	7	58
		cyclic	9	8	16	13	10	56
	Not oscillating	linear	13	16	13	12	14	68
		branched	12	16	10	15	17	70
		cyclic	16	17	7	16	23	79

Table 5.21.: Numbers of RVs obtained by training five RVMs on independent training sets of sample size 1000 each.

Figure 5.12 illustrates an example of the RVs produced during a single classification run. Elasticities as well as RVs have been clustered based in their similarity. The clustering offers a straight-forward method to look at elasticities that have to be coordinated to ensure stability or oscillations.

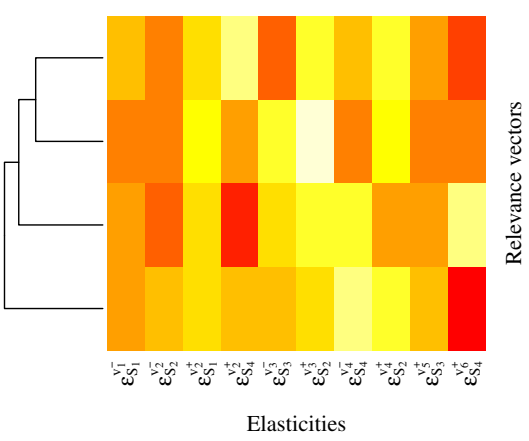
Stable models were represented by four RVs. They differed most strongly in

	Stability prediction		Oscillation prediction	
Elasticity	Stability	Instability	Oscillating	Not oscillating
$\varepsilon_{S_1}^{v_1^-}$	0.64 +- 0.23	0.54 +- 0.19	0.59 +- 0.24	0.6 +- 0.24
$\varepsilon_{S_1}^{v_2^+}$	0.68 +- 0.24	0.48 +- 0.23	0.7 +- 0.21	0.76 +- 0.22
$\varepsilon_{S_2}^{v_2^-}$	0.51 +- 0.26	0.9 +- 0.12	0.77 +- 0.23	0.52 +- 0.3
$\varepsilon_{S_4}^{v_2^+}$	0.47 +- 0.26	0.88 +- 0.08	0.64 +- 0.24	0.41 +- 0.31
$\varepsilon_{S_2}^{v_3^+}$	0.71 +- 0.21	0.55 +- 0.26	0.68 +- 0.21	0.72 +- 0.23
$\varepsilon_{S_3}^{v_3^-}$	0.69 +- 0.22	0.57 +- 0.28	0.65 +- 0.25	0.66 +- 0.23
$\varepsilon_{S_2}^{v_4^+}$	0.63 +- 0.26	0.86 +- 0.13	0.71 +- 0.23	0.62 +- 0.27
$\varepsilon_{S_4}^{v_4^-}$	0.72 +- 0.22	0.82 +- 0.2	0.71 +- 0.24	0.74 +- 0.19
$\varepsilon_{S_3}^{v_5^+}$	0.54 +- 0.25	0.37 +- 0.31	0.48 +- 0.29	0.54 +- 0.26
$\varepsilon_{S_4}^{v_6^+}$	0.57 +- 0.29	0.2 +- 0.18	0.43 +- 0.27	0.63 +- 0.3

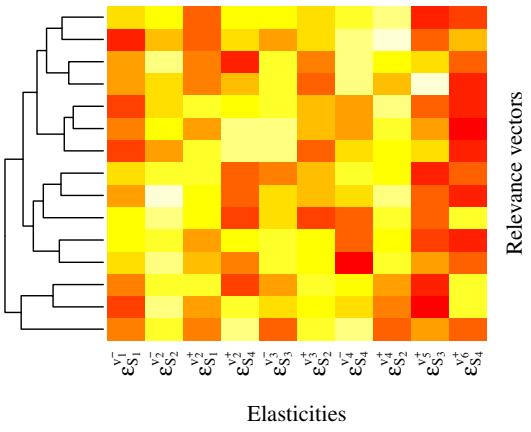
Table 5.22.: Summary of the RVs representing different classes of branched pathway models. For each elasticity, means and standard deviations were computed from all RVs derived for a particular class over all classification runs. RVs differed strongly within each class in order to cover its region in the feature space as widely as possible.

elasticity $\varepsilon_{S_4}^{v_6^+}$ and $\varepsilon_{S_4}^{v_2^+}$. A high value in $\varepsilon_{S_4}^{v_2^+}$ was accompanied by a low value in $\varepsilon_{S_4}^{v_6^+}$ and vice versa. This shows that the feedback strength exhibited by activator S_4 had to be coordinated with its efflux rate out of the system. All three RVs of the unstable class had high values in elasticity $\varepsilon_{S_1}^{v_1^-}$ which showed that instabilities could only arise if perturbations were propagated quickly at the entry point into the system. Furthermore, unstable models contained either high values in $\varepsilon_{S_1}^{v_2^+}$ and $\varepsilon_{S_4}^{v_6^+}$, or in $\varepsilon_{S_2}^{v_3^+}$ and $\varepsilon_{S_4}^{v_6}$. Interestingly, if both $\varepsilon_{S_1}^{v_2^+}$ and $\varepsilon_{S_2}^{v_3^+}$ had high values, instabilities could only arise if $\varepsilon_{S_4}^{v_6^+}$ was low.

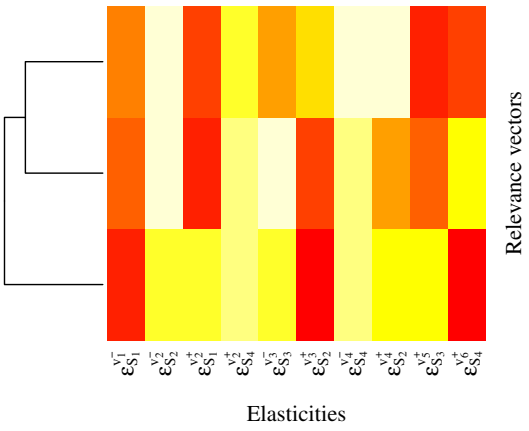
Oscillatory RVs were more diverse, but generally also required large values in $\varepsilon_{S_4}^{v_6^+}$ together with small values in $\varepsilon_{S_2}^{v_4^+}$. This similarity to the stabilizing rulesets might explain why most of the rulesets fulfilling the complex oscillation criteria also fulfilled those for stability, so that the proportion of stable oscillatory models was larger than that of unstable oscillatory ones.



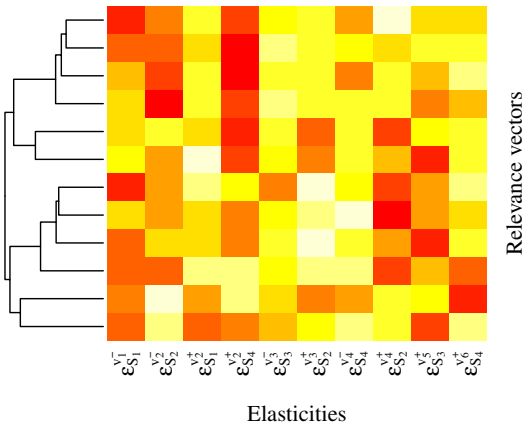
(a) RVs representing the stable class.



(b) RVs representing the oscillatory class.



(c) RVs representing the unstable class.



(d) RVs representing the non-oscillatory class.

Figure 5.12.: Examples of the RVs representing each class in branched pathway models. Elasticities and RVs have been clustered based in their similarity using the default settings of the `heatmap` function in `R`. Increased colour intensity indicates larger elasticity values.

6. The neuronal TCA cycle: a real-world example

This chapter is dedicated to elucidating local system properties of the neuronal TCA cycle. The TCA cycle is of fundamental importance for cellular energy metabolism because it is the major source for reduced nicotinamide adenine dinucleotide (NADH) and ubiquinol (QH_2), which are required for the production of adenosine triphosphate (ATP) in the mitochondrion (Figure 6.1). Its reactions are tightly controlled by allosteric feedback regulators to enable an adjustment of the steady state fluxes to varying ATP demands (Nelson and Cox, 2004). In neurons, a steady supply of ATP is crucial for restoring the cellular membrane potential after triggering an action potential. Once the system obtains a functional working state that enables it to meet the ATP demand of the cell, we can therefore expect this state to be robust against perturbations from the cytosol (Koopman *et al*, 2012). Such perturbations can arise, for example, due to neuronal activity and the resulting fast fluctuations of ATP turnover. In order to elucidate the mechanisms responsible for perturbation responses in this system, two steady states will be analysed here by using the previously presented methods for the evaluation of SKM experiments.

6.1. Steady states

Two steady states representing different scenarios in terms of cytosolic ATP demand were computed using the kinetic model developed by Berndt *et al* (2012). While state 1 (reference state) was characterised by a moderate extent of neuronal activity, state 2 described a phenomenon called 'gamma oscillations'. Gamma oscillations are defined as rhythmic brain activity with alternating epochs of enhanced and reduced neuron firing in a frequency of $30 - 100$ Hz (Fell and Axmacher, 2011; Singer, 2013). They have been associated with cognitive processing and memory formation (Brittain and Brown, 2014; Hanslmayr and Staudigl, 2014).

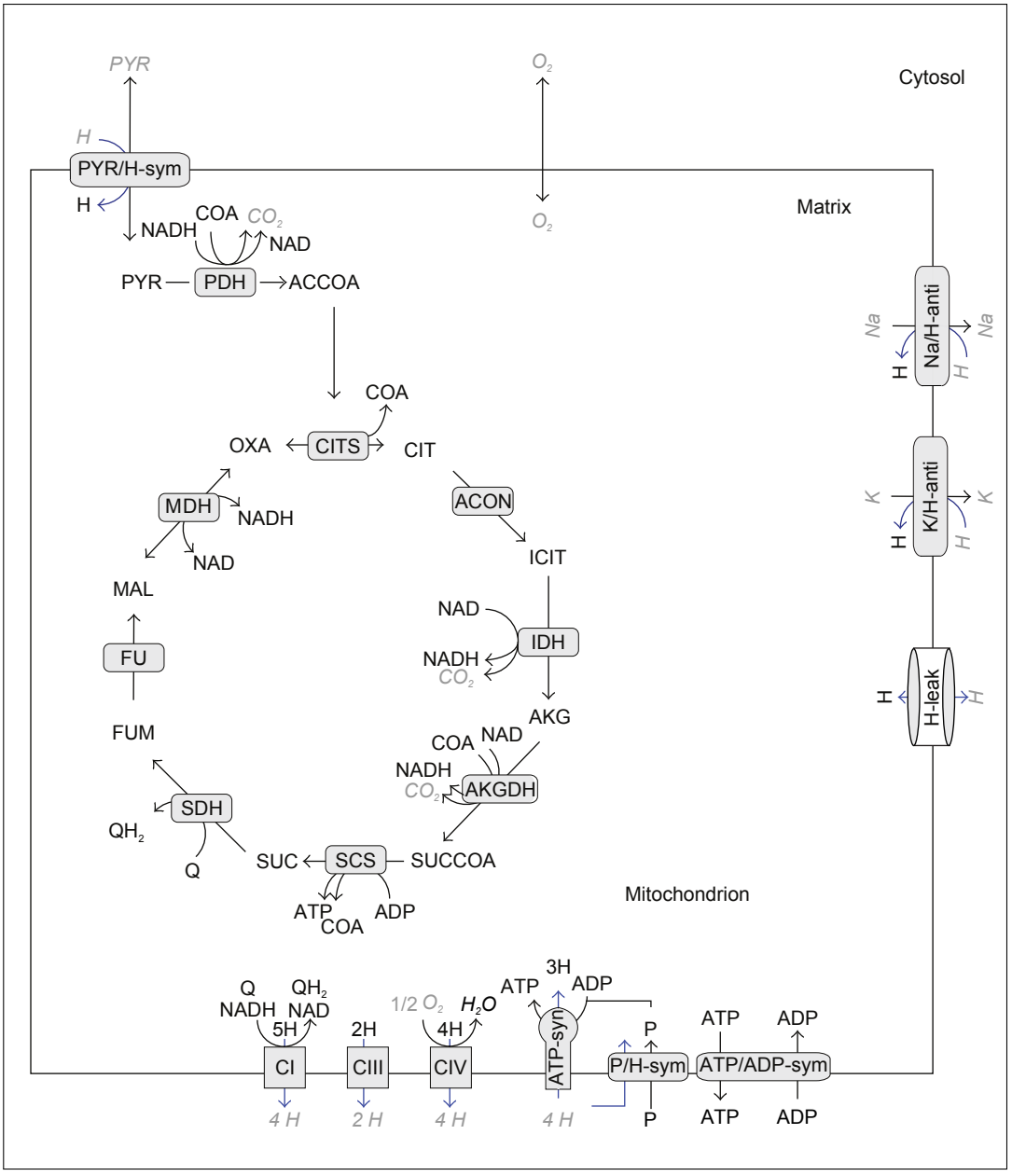


Figure 6.1.: Network model of the TCA cycle and connected pathways. In the TCA cycle, pyruvate (PYR) from the cytosol is incorporated into citrate (CIT) and converted to CO_2 by a series of oxidation steps. The energy released in this process is utilized to form the reduced metabolites NADH and QH_2. In the respiratory chain complexes CI, CII, and CIV these metabolites are oxidised, driving the import of protons into the mitochondrial matrix. As a result of this process, a proton gradient is created which then serves as a driving force for ATP synthesis. The picture has been adapted from Berndt *et al* (2012). Metabolites marked in grey were treated as constant values and not included into the ODE system shown in Table 3.4.

Figure 6.2 shows the reaction rates of both investigated steady states. Because of the high frequency of action potentials, gamma oscillations are accompanied by a large ATP demand in the cytosol. This requires higher rates of mitochondrial ATP synthesis, which in turn can only be achieved by accelerated respiratory chain and TCA cycle reactions. The computed steady states indicated that the increased fluxes through the TCA cycle led to elevated levels of the low-energy metabolites ADP, NAD, Q, as well as reduced Cytochrome C (CytC) (Figure 6.3). Reciprocally, the energy-rich metabolites ATP, NADH, QH_2 and oxidised CytC were less abundant under gamma oscillations compared to the reference state. Furthermore, the level of α-ketoglutarate (AKG) was strongly reduced, while succinate (SUC) concentration was distinctly increased under gamma oscillations.

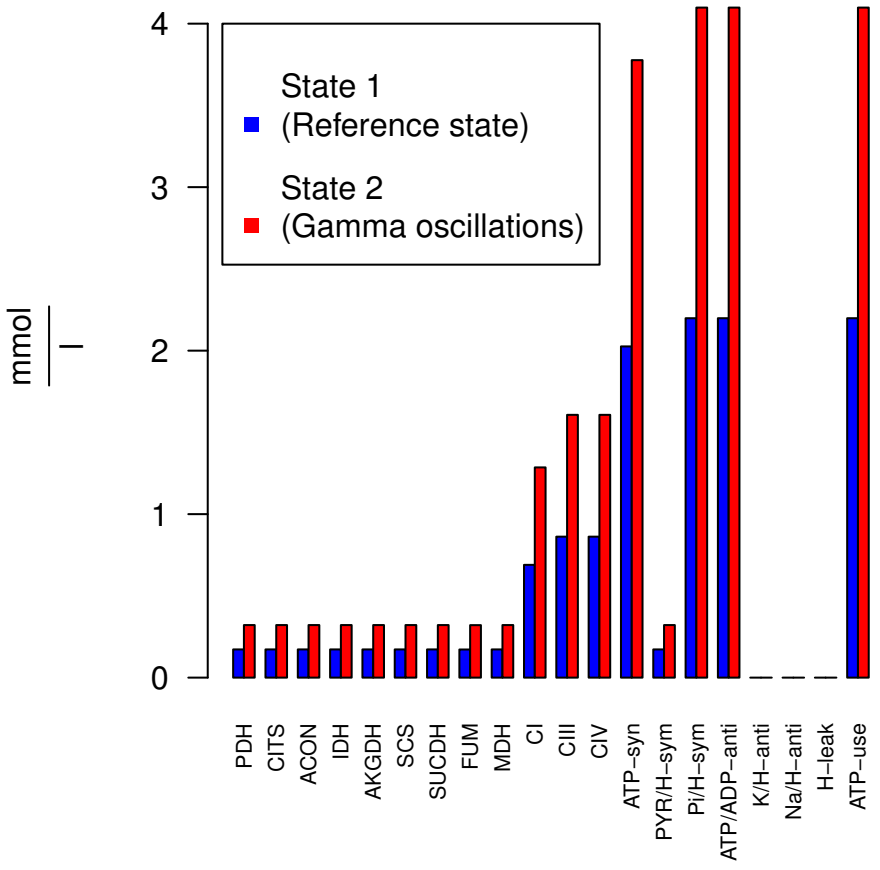


Figure 6.2.: Reaction rates of two steady states (reference state and gamma oscillations) of the neuronal TCA cycle.

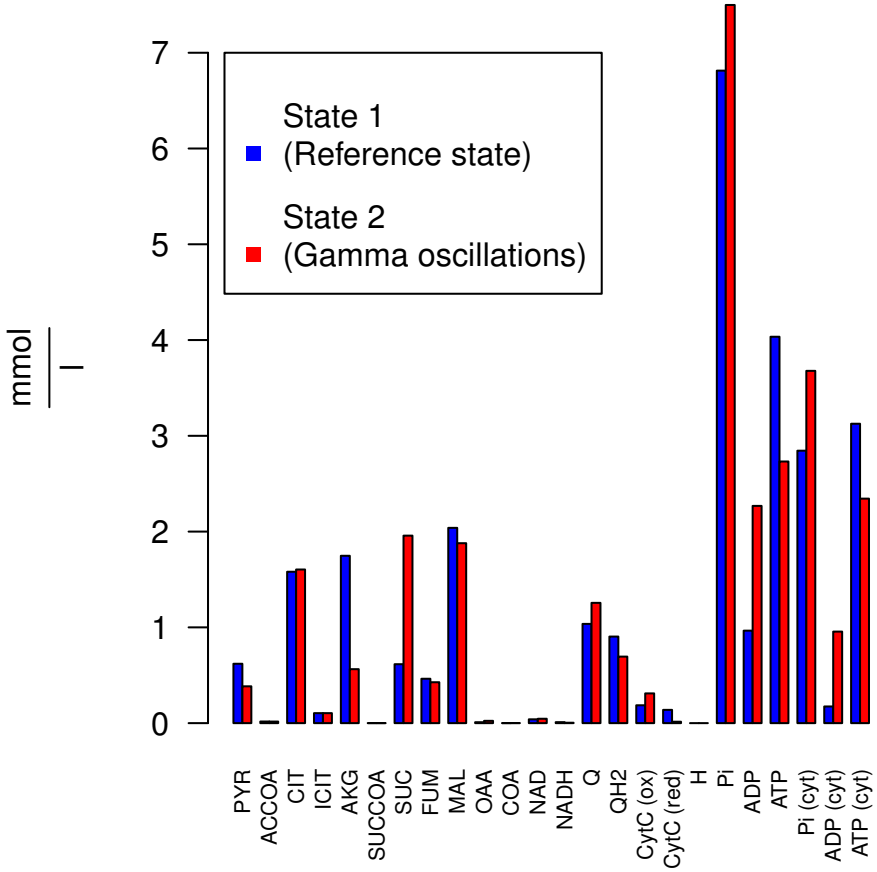


Figure 6.3.: Metabolite concentrations of the investigated steady states.

6.2. Quantitative analysis of local dynamic steady state properties

For each steady state, 10,000 SK-models were sampled and the proportions of stable and oscillatory models were determined. The system contained three reactions with only one substrate and product each (ACON, FU, H-leak). The corresponding elasticities were filtered for biological feasibility using the criterion derived in Section 2.3.2. In doing so, 76.08% of all models could be corrected for biological feasibility in these reactions within each steady state. In 63.3% of the resulting models the FCCs indicated negative control of PYR import on PDH. Consequently, an additional criterion was implemented during sampling that only admitted models in which pyruvate import exhibited positive control on the PDH.

All models showed strong tendencies towards stability (Figure 6.4). Oscillatory trajectories were only observed in between 30 - 40 % of all models that included

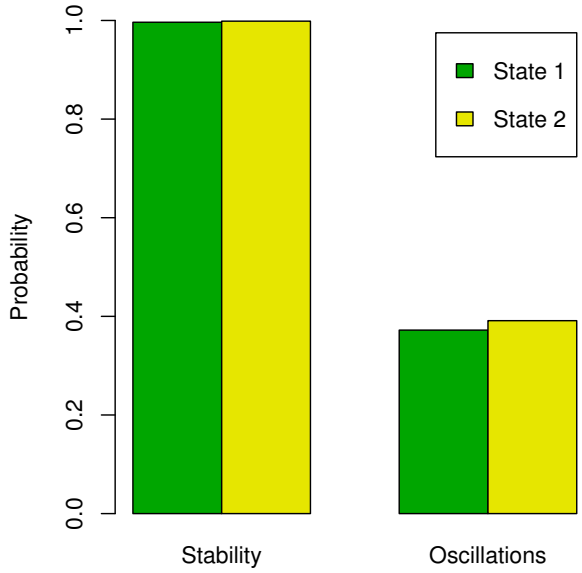


Figure 6.4.: Distribution of local dynamic properties of both steady states of the TCA cycle.

regulators. In contrast to the small example pathways in which all enzymes followed simple Michaelis Menten kinetics, the dynamic behaviour of the TCA cycle steady states was hardly affected by elasticity filtering (Table 6.1). This indicated that the affected reactions ACON, FU and H-leak did not play significant roles in controlling stability or oscillations.

Similar to elasticity filtering, removing biologically infeasible FCCs had no significant impact on stability and it led to only a minor reduction in oscillations around the reference state. This observation shows that it is generally not possible to reliably detect biologically infeasible elasticity combinations simply from the eigenvalues and emerging dynamic properties in an SKM experiment. Instead, we need to rely on prior knowledge about the expected flux control exhibited by at least some of the enzymes in the system, in order to reduce the sampling space for the elasticities and thereby increase the chances to obtain realistic SK-models.

6.3. Univariate search for discriminating features

In order to detect the network positions that exhibited the most control over the system's dynamics, elasticities and flux control coefficients were compared using the Kolmogorov-Smirnov-test. The five most influential elasticities of the reference

	No filtering		Filtered elasticities		Filtered elasticities & FCCs	
	State 1	State 2	State 1	State 2	State 1	State 2
Total stability (%)	0.998	0.998	0.996	0.999	0.996	0.998
Total oscillations (%)	0.372	0.397	0.372	0.391	0.350	0.401
Stable/ with oscillations (%)	0.372	0.397	0.372	0.391	0.349	0.401
Stable/ no oscillations(%)	0.625	0.602	0.624	0.608	0.646	0.598
Unstable/ with oscillations (%)	0.000	0.000	0.000	0.000	0.000	0.000
Unstable/ no oscillations (%)	0.002	0.002	0.003	0.001	0.004	0.002

Table 6.1.: Detailed list of proportions of possible steady state properties of the TCA cycle determined by Monte Carlo sampling. State 1: reference state; State 2: gamma oscillations.

state are shown in Table 6.2, together with the distances between the distributions of the opposing classes. All depicted elasticities were associated with p-values below 0.01.

The majorities of top-ranking elasticities that controlled stability were associated with citrate synthase (CITS). CITS marks the entry point to the TCA cycle for carbon backbones in the form of acetyl-CoA (ACCOA). Because of this important function, together with its interconnectedness to the TCA-cycle metabolites oxalacetate (OAA) and citrate (CIT), as well as with the cofactor coenzyme A (COA), it can be expected to play a crucial role in maintaining stability.

The second most influential elasticity described the control of AKGDH by its substrate AKG. The enzyme complex AKGDH is not only considered a rate-limiting step in the TCA cycle (Berndt *et al*, 2012), but also catalyses a complex reaction, which involves different types of cofactors (NAD, NADH, COA). Due to these connections, a change in this enzyme can be expected to affect the rates of other reactions in the network and to cause potential threats to stability. In the reference state, the allosteric inhibition of AKGDH by succinyl-CoA (SUCCOA) closely followed the substrate activation of CITS and AKGDH with respect to its importance in controlling stability. However, its role was less important under gamma oscillations. During gamma oscillations, the substrate activation

Stability			Oscillations		
Elasticity	Distance (state 1)	Distance (state 2)	Elasticity	Distance (state 1)	Distance (state 2)
$\varepsilon^{\mathrm{CITS}^+}_{\mathrm{OAA_m}}$	0.54	0.53	$\varepsilon^{\mathrm{AKGDH}^+}_{\mathrm{AKG_m}}$	0.39	0.14
$\varepsilon^{\mathrm{AKGDH}^+}_{\mathrm{AKG_m}}$	0.48	0.40	$\varepsilon^{\mathrm{IDH}^+}_{\mathrm{ICIT_m}}$	0.21	0.41
$\varepsilon^{\mathrm{AKGDH}^+}_{\mathrm{SUCCOA_m}}$	0.38		$\varepsilon^{\mathrm{AKGDH}^+}_{\mathrm{SUCCOA_m}}$	0.19	0.13
$\varepsilon^{\mathrm{CITS}^+}_{\mathrm{SUCCOA_m}}$	0.31	0.32	$\varepsilon^{\mathrm{CITS}^+}_{\mathrm{CIT_m}}$	0.13	
$\varepsilon^{\mathrm{CITS}^+}_{\mathrm{CIT_m}}$	0.28	0.33	$\varepsilon^{\mathrm{CITS}^+}_{\mathrm{SUCCCOA_m}}$	0.09	
$\varepsilon^{\mathrm{IDH}^+}_{\mathrm{ICIT_m}}$		0.28	$\varepsilon^{\mathrm{AKGDH}^+}_{\mathrm{COA_m}}$		0.10
			$\varepsilon^{\mathrm{IDH}^+}_{\mathrm{ADP_m}}$		0.09

Table 6.2.: The top five elasticities with largest distances between their distributions in opposing classes for the TCA cycle model, determined by the Kolmogorov-Smirnov test. All values were associated to p-values below 0.01. Elasticities reflecting allosteric inhibition are marked in red, whereas activating regulators are marked in green. State 1: reference state; State 2: gamma oscillations.

of isocitrate dehydrogenase (IDH) by isocitrate (ICIT) played a dominant role. IDH not only interacts with the cofactors NAD and NADH, but it also produces AKG, which serves as a substrate for AKGDH. We can conclude that in both steady states, elasticites associated with reactions that were located upstream of the TCA cycle and that contained the cofactors COA or NAD/NADH exhibited most control on stability. The differences in top-ranking elasticities between scenarios of moderate and high ATP demand showed that control over stability can shift with changing steady states. In both steady states, the concentrations of SUCCOA and ICIT coincided closely. However, the concentration of AKG was lower during gamma oscillations, which could be a possible explanation for greater emphasis on ICIT in controlling system properties.

Occurrence of oscillatory trajectories around the reference state was mainly controlled by the substrate activation of AKGDH by AKG. Under gamma oscillations, however, substrate activation of IDH by ICIT was most important. The test statistics of the KS-test indicated that the corresponding elasticity showed strong distribution differences between oscillatory and non-oscillatory models. In contrast, all other elasticities exhibited distance values of 0.14 or lower. Analogous to controlling stability, the enzyme IDH turned out to obtain more control over oscillations when flux through the system increased and the levels of AKG decreased under

gamma oscillations.

In summary, the Kolmogorov-Smirnov test revealed that the distributions of some elasticities associated with CITS, AKGDH, and to a more limited extent IDH, differed strongly between stable and unstable models. Three of the top-ranking elasticities represented allosteric feedback effects. However, we already observed in Chapter 5 that comparing individual elasticity distributions is only informative to a limited extent when detecting quantitative elasticity thresholds associated with a specific system property. For example, the homogeneous distributions of elasticities in stable steady states had showed that, for almost all values in any elasticity, there existed a combination of other elasticities that allow the system to maintain stability in the small example systems. Therefore, the following section will demonstrate how deriving quantitative thresholds for elasticity combinations can help in the understanding of which enzyme mutations could potentially make the system prone to instabilities.

6.4. Analysing instability conditions by multivariate pattern search

6.4.1. RVM classification performance

Due to the longer runtime of the RVM algorithm, RVMs could only be successfully trained for smaller sample sizes than those possible for decision trees. When analysing the example pathways in Chapter 5, small sample sizes had been sufficient to obtain comparable or even better RVM classification performances than could be obtained by decision trees. In contrast, the TCA cycle posed a more challenging classification problem, which made it difficult for the RVM algorithm to deduce informative and reproducible decision functions. As a result, training RVMs on the corresponding SK-models led to high bias as well as large generalization errors (Table 6.3). Because of these drawbacks, the following analysis will focus on the decision tree results only.

6.4.2. Decision tree classification performance

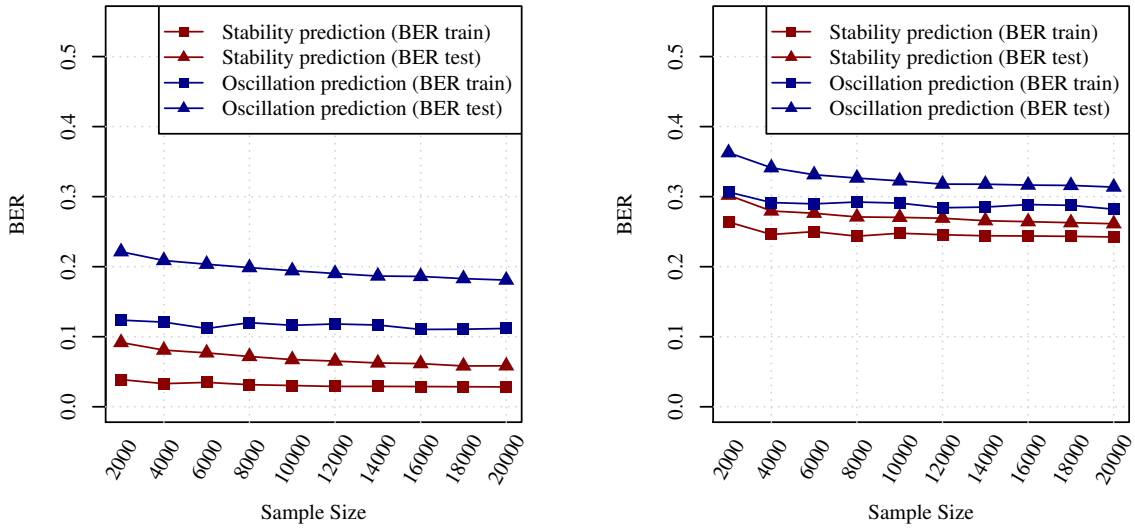
In order to assess the performance of the decision tree algorithm in classifying stable versus unstable as well as oscillatory versus non-oscillatory models, trees were trained on balanced datasets derived from SK models for the reference state. Dur-

Algorithm	Balanced training error	Balanced test error	Runtime in minutes
Decision trees	0.039 ± 0.006	0.092 ± 0.005	0.39 ± 0.01
RVMs	0.148 ± 0.013	0.294 ± 0.037	34.236 ± 1.518

Table 6.3.: Comparison of classification performance between decision trees and RVMs for stability conditions of the TCA cycle. For each algorithm, training was performed five times on independently sampled datasets comprising 2000 SK-models for the reference state of the TCA cycle. The depicted values show means and standard deviations obtained from the five replicates.

ing training, the sample sizes were gradually increased in order to check whether data availability was a limiting factor for model bias and generalizability.

Classification by elasticities



(a) BERs obtained by classifying based on elasticities.

(b) BERs obtained by classifying based on FCCs.

Figure 6.5.: BERs obtained by classification of SK-models of the TCA cycle reference state according to different types of features. The depicted values represent the average over the BERs obtained from five trees created for each training data size. Red curves: classification of stable versus unstable models; blue curves: classification of oscillatory versus non-oscillatory models. Squares: training errors; triangles: test errors.

When classifying based on elasticities, training errors of stability prediction were small for all sample sizes (see Figure 6.5 for exemplary training curves obtained for the reference state, as well as Appendix D.1.1 for details). Consistent with the observations obtained previously for the small example pathways, the training

data contained sufficient information to be translated into rules by the algorithm so that the derived trees were less prone to bias. For increasing sample sizes, the test errors approximated the training errors closely. This indicated that most rulesets generalised well with little overfitting.

Training errors for elasticity-based oscillation prediction were distinctly higher than those for stability prediction, hinting at a higher model bias. Consequently, it was more difficult for the decision tree algorithm to formulate patterns for oscillatory conditions than for stability conditions. The test errors always stayed significantly higher than the training errors, indicating that, similar to classification on the small example pathways, the derived trees were not as well generalizable as those constructed for stability prediction.

Classification by flux control coefficients (FCCs)

In contrast to the results obtained for the small example pathways, stability and oscillation prediction for TCA cycle models both suffered from high bias, which showed that it was difficult for the decision tree algorithm to derive stabilizing of oscillatory FCC patterns from the training data (Figure 6.5). The previous observation that classification performance for the linear and branched pathway structures was much better than that for the cyclic pathway already indicated that the success of stability prediction based on FCCs strongly depended on the pathway topology (Figure 5.9 - 5.11). Similarly to the cyclic pathway structure in the small example models, the complex structure of the TCA cycle model therefore introduced high bias for classification.

Consistent with the results obtained from the small example pathways, test errors closely approximated training errors for large sample sizes. This confirmed that, despite the difficulties in deriving informative FCC-based rules, those that were eventually detected generalized well to unknown test cases.

6.4.3. Ruleset numbers and sizes

Tables 6.4 and 6.5 show the distribution of ruleset sizes with Laplace values $\geqslant 0.95$ that were derived from five trees trained on 20,000 training samples each. Due to the high model bias when training on the FCC data sets, no reliable rulesets in terms of FCCs could be derived for the TCA cycle steady states.

Using elasticities as input features led to large numbers of rulesets for stable and unstable models. Despite the previously observed high abundance of stable models

		Conditions per ruleset										
	Class label	1	2	3	4	5	6	7	8	9	10	Sum
Stability	State 1	0	0	1	10	26	27	23	11	3	1	102
	State 2	0	0	1	1	16	34	24	14	3	0	93
Instability	State 1	0	3	10	21	32	21	21	7	4	1	120
	State 2	0	2	21	30	40	38	18	10	2	1	162

Table 6.4.: Distribution of elasticity ruleset sizes with Laplace value $\geqslant 0.95$ obtained by classifying stable versus unstable models. The numbers have been summarized over all the five trees that were constructed for each classification problem. State 1: reference state; State 2: gamma oscillations.

during random sampling, reliable stability rulesets were less frequent and contained larger numbers of conditions than those describing unstable steady states. Only one reliable ruleset could be derived for oscillatory trajectories around the reference state, whereas none could be detected during high neuron activity. Containing eight conditions, the detected ruleset for the reference state was comparably large.

		Conditions per ruleset										
	Class label	1	2	3	4	5	6	7	8	9	10	Sum
Oscillations	State 1	0	0	0	0	0	0	0	1	0	0	1
	State 2	0	0	0	0	0	0	0	0	0	0	0
No oscillations	State 1	0	7	19	26	17	7	3	3	0	1	83
	State 2	1	22	25	27	17	12	7	1	1	0	113

Table 6.5.: Distribution of elasticity ruleset sizes with Laplace value $\geqslant 0.95$ obtained by classifying oscillatory versus non-oscillatory models. The numbers have been summarized over all the five trees that were constructed for each classification problem. State 1: reference state; State 2: gamma oscillations.

Rulesets ensuring non-oscillatory trajectories around the reference state contained at least two conditions each. In contrast, the higher neuronal activity in state 2 not only increased the number of rulesets describing non-oscillatory trajectories, but it also enabled the emergence of one ruleset for which one condition was sufficient. This indicated that faster flux caused by gamma oscillations made it easier to formulate reliable conditions in order to ensure non-oscillatory trajectories.

To summarize, it was possible to formulate reliable destabilizing or non-oscillatory patterns with low numbers of conditions, whereas rulesets reliably describing stable or oscillatory steady states always required three or more conditions. We can conclude that in a complex system like mitochondrial metabolism, with its sophisticated interconnectedness of substrates and cofactors, changes in at least three (but often more) enzymes are generally required to reliably change local system properties while maintaining the same steady state.

6.4.4. Detecting the most informative elasticities for stability control

Table 6.6 shows the frequencies of elasticities in stabilizing rulesets with Laplace values $\geqslant 0.95$ obtained for both steady states, together with the proportion of lower bounds associated with each elasticity in percent.

Analysis of the small example pathways in the previous section had already shown that elasticities were distributed widely among stabilizing rulesets, whereas destabilizing rulesets focused on a smaller number of certain elasticities that were more prominent than the rest. Such a strong difference in ruleset content between the stable and unstable class could not be detected for the TCA cycle steady states. Moreover, there existed no single elasticity that was required in every stabilizing or destabilizing ruleset. Instead, the investigated system properties could both be invoked by diverse combinations of elasticities associated with PDH, AKGDH, CITS and IDH. Under gamma oscillations, two additional elasticities ($\varepsilon_{\mathrm{H_m}}^{\mathrm{ATPsyn}^-}$, $\varepsilon_{\mathrm{H_m}}^{\mathrm{CIII}^+}$) were included in the vast majority of rulesets. They described the extent to which ATP synthesis and Complex III (CIII) were affected by increasing proton concentration in the matrix.

These results indicated that, similar to the univariate comparisons shown in the previous section, the most informative elasticities for reference state stability were associated with reactions upstream of the TCA cycle. The high ATP demand under gamma oscillations led to a shift in control that included reactions involved in ATP synthesis. The univariate comparisons had only revealed comparably small distribution differences in $\varepsilon_{\mathrm{H_m}}^{\mathrm{ATPsyn}^-}$ (KS-test statistic of 0.13) and $\varepsilon_{\mathrm{H_m}}^{\mathrm{CIII}^+}$ (KS-test statistic of 0.09). However, the multivariate analysis revealed that these elasticities were of central importance in controlling stability under gamma oscillations. The shift towards ATP synthesizing reactions reduced the importance of allosteric inhibition of AKGDH and CITS. As a result, stabilizing rulesets for gamma oscil-

Elasticity	Stability		Instability	
	State 1	State 2	State 1	State 2
$\varepsilon^{\mathrm{AKGDH}^+}_{\mathrm{AKG_m}}$	**0.92** (87.23 %)	**0.94** (95.4 %)	0.87 (5.77 %)	0.81 (6.87 %)
$\varepsilon^{\mathrm{ATPsyn}^-}_{\mathrm{H_m}}$		**0.92** (96.51 %)		0.64 (5.77 %)
$\varepsilon^{\mathrm{CIII}^+}_{\mathrm{H_m}}$		0.69 (87.5 %)		
$\varepsilon^{\mathrm{AKGDH}^+}_{\mathrm{NADH_m(regulator)}}$	0.75 (92.11 %)		0.86 (30.1 %)	
$\varepsilon^{\mathrm{AKGDH}^+}_{\mathrm{SUCCOA_m(regulator)}}$	0.74 (77.33 %)		0.68 (1.22 %)	0.71 (26.96 %)
$\varepsilon^{\mathrm{CITS}^+}_{\mathrm{CIT_m(regulator)}}$	0.72 (39.73 %)		0.52 (31.75 %)	0.48 (32.47 %)
$\varepsilon^{\mathrm{CITS}^+}_{\mathrm{NADH_m(regulator)}}$			0.43 (96.15 %)	0.34 (98.18 %)
$\varepsilon^{\mathrm{CITS}^+}_{\mathrm{OAA_m}}$	0.36 (86.49 %)	0.55 (56.86 %)	0.42 (100 %)	0.28 (62.22 %)
$\varepsilon^{\mathrm{CITS}^+}_{\mathrm{SUCCOA_m(regulator)}}$	0.35 (2.78 %)	0.47 (88.64 %)	0.33 (2.5 %)	0.27 (93.02 %)
$\varepsilon^{\mathrm{IDH}^+}_{\mathrm{ICIT_m}}$	0.34 (11.43 %)	0.39 (13.89 %)	0.27 (84.38 %)	0.23 (56.76 %)
$\varepsilon^{\mathrm{PDH}^+}_{\mathrm{COA_m}}$	0.27 (96.43 %)	0.37 (94.12 %)	0	
$\varepsilon^{\mathrm{PDH}^+}_{\mathrm{COA_m(regulator)}}$	0.25 (68 %)	0.26 (12.5 %)	0.19 (86.96 %)	0.21 (88.24 %)
$\varepsilon^{\mathrm{PDH}^+}_{\mathrm{NADH_m(regulator)}}$	0.22 (9.09 %)	0.19 (0 %)	0.17 (95 %)	0.2 (93.75 %)
$\varepsilon^{\mathrm{PDH}^+}_{\mathrm{PYR_m}}$		0.17 (93.75 %)		

Table 6.6.: Frequency of the ten most common elasticities in ruleset conditions for the TCA with Laplace values $\geqslant 0.95$. Rulesets were derived from SK-models of two steady states: a reference state (Ref) and gamma oscillations (Gam). Values in brackets refer to the percentage of lower bounds associated with each elasticity (green: $\geqslant 90\%$ lower bounds; red: $\leqslant 10\%$ lower bounds). Elasticities present at least 0.9 times per ruleset on average are marked in **bold**. State 1: reference state; State 2: gamma oscillations.

lations contained no regulatory elasticity associated with AKGDH and only one out of three associated with CITS.

Frequencies of lower and upper bounds indicated that high values in both $\varepsilon^{\mathrm{ATP\text{-}syn}^-}_{\mathrm{H_m}}$ and $\varepsilon^{\mathrm{CIII}^+}_{\mathrm{H_m}}$ were favoured in stable models, but low values were preferred in unstable models. This implied that in many SK-models a fast response to changes in proton gradient was required for CIII and ATPsyn in order to maintain stability. Rulesets describing instabilities under gamma oscillations focused more strongly on allosteric regulation of AKGDH and CITS than stabilizing rulesets. In contrast, elasticities associated with the proton gradient were not as abundant

in destabilizing rulesets as in stabilizing ones. This showed that changes in the allosteric inhibitors of AKGDH and CITS seemed to be the bigger risk to instabilities than perturbations in the proton gradient.

The elasticity $\varepsilon^{\mathrm{AKGDH}^+}_{\mathrm{AKG_m}}$, which described substrate activation of AKGDH by AKG, was the most abundant elasticity in all rulesets. This observation agreed with the univariate comparisons in which the same elasticity had shown strong distribution differences between both classes. The high abundance of lower bounds in stabilizing rulesets showed that they tended to restrict AKGDH-associated elasticities to large values, whereas destabilizing rulesets often restricted them to small values. Therefore, a quick response of AKGDH to perturbations seemed to be favourable for stability. Reciprocally, a slow response could be a potential threat to stability of the steady states and could be a result, for example, of a mutation that reduced the Michaelis constant associated with AKG,.

Contrary to $\varepsilon^{\mathrm{AKGDH}^+}_{\mathrm{AKG_m}}$, the elasticity $\varepsilon^{\mathrm{CITS}^+}_{\mathrm{OAA_m}}$, which described the substrate activation of CITS by OAA, occurred in only $36 - 55\%$ of reliable stabilizing and destabilizing rulesets, despite the strong class-differences observed by the pairwise comparisons. This indicated that, even if the distributions differed significantly so that stable models often tended to have different values than unstable models, stability and instability were also possible with values that were predominately found in the opposing class. Both stabilizing and destabilizing rulesets for the reference state mainly restricted this elasticity by lower bounds, whereas under gamma oscillations, such a prevalence could not be detected. A similar phenomenon was observed for elasticity $\varepsilon^{\mathrm{CITS}^+}_{\mathrm{SUCCOA_m(regulator)}}$, which described the regulatory effect of SUCCOA on CITS. It was restricted by upper bounds in the reference state, but by lower bounds during gamma oscillations.

6.4.5. Simple causes of instabilities in the TCA cycle

Trees trained on SK-models for the reference state and for gamma oscillations produced five elasticity rulesets with two conditions each (Table 6.7). In each ruleset, one condition imposed a very strict upper boundary on the substrate activation of AKGDH by AKG. The second condition restricted the feedback inhibition of AKGDH by its product SUCCOA. According to these rulesets, instabilities were likely to arise if AKGDH was highly saturated by its substrate AKG and by its product SUCCOA. This indicated that the kinetic properties of AKGDH had to be tailored towards allowing fast responses of this enzyme to perturbations, and

Steady state	Summarized conditions	Laplace values
Reference state	$\varepsilon^{\mathrm{AKGDH^+}}_{\mathrm{AKG_m}} \leqslant 0.0433 \pm 0.0252$ $\varepsilon^{\mathrm{AKGDH^+}}_{\mathrm{SUCCOA_m}} > -0.13 \pm 0.0781$	0.9748 ± 0.0158
Gamma oscillations	$\varepsilon^{\mathrm{AKGDH^+}}_{\mathrm{AKG_m}} \leqslant 0.03 \pm 0$ $\varepsilon^{\mathrm{AKGDH^+}}_{\mathrm{SUCCOA_m}} > -0.24 \pm 0.01$	0.9748 ± 0.0158

Table 6.7.: Summary of destabilizing elasticity rulesets (two conditions each) in the TCA cycle.

that mutations that changed these properties were among the most hazardous for mitochondrial metabolism. The rulesets were well reproducible between both steady states. Consequently, the crucial role of AKGDH in causing this specific type of instabilities was independent of the overall flux through the system and of the concentrations of their reactants.

In order to understand the reasons for instabilities caused by the conditions in Table 6.7, the eigenvectors belonging to the largest real parts of the affected models were analysed. The eigenvectors of the Jacobian matrix can give an approximation of the system's behaviour in the neighbourhood of the steady state (see Section 2.2 in Chapter 2). Among 100,000 SK models with equal numbers of stable and unstable instances computed for the reference state, 7535 fulfilled the conditions in Table 6.7. 7367 (97.77 %) of these models indeed described unstable steady states. Under gamma oscillations, 7852 fulfilled the conditions and 7668 (97.66 %) of these models actually belonged to unstable steady states.

Analysis of the eigenvector with largest real part indicated that perturbation led to an accumulation of AKG, accompanied by a decrease in the subsequent TCA cycle metabolites (Figure 6.6). A possible explanation for this behaviour could be that the strong saturation of AKGDH by its substrate AKG prevented an acceleration of the enzyme, which would be necessary to cope with slight perturbation-induced increases in the concentrations at the entry point to the TCA cycle. Because AKGDH was allosterically inhibited by its product SUCCOA in the model, the resulting depletion of SUCCOA could be expected to decrease the inhibitory feedback and assist in accelerating the reaction. However, the conditions in Table 6.7 stated that changes in SUCCOA only weakly affected the inhibitory feedback. Due to the combination of weak activation by its product and strong inhibition by

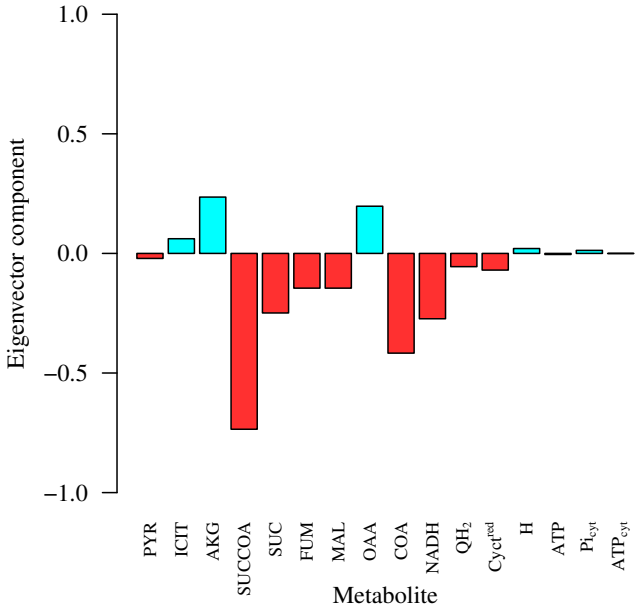


Figure 6.6.: Eigenvector belonging to the largest positive eigenvalue in unstable models of the reference state that fulfilled the instability conditions in Table 6.7. Each eigenvector component describes the time-dependent concentration changes in one metabolite in the linearised system (see Section 2.2 for details). Concentrations that increased due to a perturbation are shown in blue; decreasing metabolites are shown in red. Instabilities arise due to accumulation of AKG and a depletion in the subsequent metabolites.

its substrate, the response of the enzyme to changes in its reactants would be too slow to diminish the perturbations.

6.5. Summary: TCA cycle analysis

In this chapter, two steady states of mitochondrial metabolism were analysed with respect to their stabilizing properties. Both steady states varied in the amount of ATP production and fluxes through the system.

Monte Carlo sampling of SK-models revealed strong tendencies towards stability in both steady states. The filtering criterion derived in Chapter 3 served for correcting for kinetically infeasible elasticity combinations associated with enzymes that followed Michaelis Menten kinetics. An additional filtering step was introduced to correct for models with biologically unrealistic FCCs.

Univariate comparisons between elasticity distributions of stable and unstable models showed that elasticites associated with reactions located close to the entry

point to the TCA cycle and that contained the cofactors COA or NAD/NADH exhibited most control on stability.

Multivariate analysis was performed in order to detect quantitative criteria for instabilities and for oscillatory trajectories. For this purpose, classification was attempted using RVMs and decision trees. Although RVMs had previously demonstrated excellent performance for the analysis of small example pathway, the complexity of the mitochondrial metabolic network hampered classification in the given scenario. The direct comparison identified the decision tree algorithm as the superior approach for the analysis of SK-models in a complex biochemical pathway.

Decision tree classification enabled the detection of stabilizing and destabilizing rulesets with little bias and overfitting when elasticities were used as training data. Formulating patterns for oscillatory conditions was more difficult than for stability conditions. This indicated that oscillating trajectories around the observed steady states of mitochondrial metabolism are a complex phenomenon that requires the fine-tuned coordination of many elasticities and that it remains a challenge to describe these combinations by reliable and reproducible rules.

Closer inspection of stabilizing and destabilizing elasticity rulesets showed that, similar to the univariate comparisons, the most informative elasticities for reference state stability were associated with reactions upstream of the TCA cycle. The high ATP demand under gamma oscillations led to a shift in control that included reactions involved in ATP synthesis. The sensitivity of AKGDH to changes in its substrate AKG was the most frequent feature in stabilizing and destabilizing rulesets. A slow response of AKGDH to perturbations was detected as a potential threat to stability of the steady states.

Closer inspection of destabilizing rulesets revealed one simple and easily reproducible pattern which was a potential reason of instabilities in both investigated steady states. Eigenvector analysis indicated that in the affected models, instabilities arose due to slow responses of the enzyme AKGDH to perturbations.

In summary, using machine learning helped to unravel a more detailed picture on the potential sources of instabilities than was possible with univariate elasticity comparisons alone. For example, they hinted at a shift in the system components with biggest importance for controlling stability under increased ATP demand. The decision tree algorithm yielded quantitative thresholds that could be interpreted in a straight-forward manner. Additionally, eigenvector analysis of the models fulfilling certain ruleset criteria helped to obtain a better understanding of

the trajectories around unstable steady states.

7. Summary and discussion

In this thesis the development of new approaches for the execution and evaluation of SKM experiments was investigated. In Chapters 2 and 3, the relationships between elasticities and the kinetic parameters of the reversible Michaelis Menten rate law were inspected and it was shown how models could be checked for biological plausibility during Monte Carlo sampling of the elasticities. Furthermore, it was demonstrated how the sampled elasticities could be converted to flux control coefficients (FCCs) using the connectivity theorem in order to derive the FCCs associated with different local system properties. Additionally, a new approach for evaluating SKM experiments was presented that extended the previously used univariate elasticity comparisons by multivariate pattern search. In the cases where prior knowledge about the underlying types of rate laws existed, it was shown how the derived elasticity patterns could be calculated back into kinetic parameters. In Chapter 4, a toolbox for the efficient construction and analysis of SK-models was introduced (Girbig *et al*, 2012b). This toolbox was used for all SKM experiments presented in this thesis.

All new methodological advancements were first evaluated in a simulation study using a set of small example pathways that employed simple Michaelis Menten kinetics with homogeneous choices of kinetic parameters (Chapter 5). In doing so, the main goal was to show how SKM could be used to study the conditions for stability or for oscillatory trajectories. Afterwards, a detailed analysis of the dynamic properties of the neuronal TCA cycle was performed demonstrating how the new insights obtained in this work could be used for the study of complex metabolic systems (Chapter 6).

In the following, every topic will be recapitulated in greater detail. The opportunities, but also possible potential drawbacks for practical applications will be presented and discussed.

7.1. Investigating metabolic steady states by SKM

7.1.1. The challenge of finding biologically plausible models

In order to perform an SKM experiment, it is not necessary to specify the kinetic rate laws in detail. However, the sampling procedure requires the selection of predefined intervals for each elasticity. Often, the interval $(0, 1]$ is chosen for enzymatic reactions. This interval represents the range of possible elasticity values for reactions following Michaelis Menten kinetics (see Section 2.3.2 for analytical derivation). In this work, I demonstrated that if this assumption of Michaelis Menten kinetics holds true, only very specific combinations of elasticities allow to maintain the steady state with non-negative kinetic parameters (Section 2.3.2, equation (2.38)). Because of this, an easily applicable filtering criterion for enzymatic reactions with a single substrate and product was introduced in Section 3.1.1 (equation (3.1)), that enabled the detection of elasticity combinations associated with negative kinetic parameters.

If this filtering criterion is not taken into account during Monte Carlo sampling, the large abundance of SK-models with biologically implausible elasticity combinations can bias the output of the experiment. For example, analysis of a set of simple example pathways showed that even for small models with $9 - 11$ sampled elasticities, approximately 90 % of the resulting SK-models contained elasticity combinations that required at least one negative value in the kinetic parameters in order to enable emergence of the observed steady state. Focusing only on those models with non-negative kinetic parameters hardly changed the numbers of stable models, but strongly affected the frequency of oscillatory trajectories around the steady state.

The filtering criterion introduced in equation (3.1) offers a helpful starting point when sampling elasticities for reactions that can be assumed to follow reversible Michaelis Menten kinetics. When analysing complex biological systems, one has to keep in mind, however, that the analytical derivation of similar criteria for more complex rate laws is not possible in the same straight-forward manner. For example, the TCA cycle model analysed in Chapter 6 contained only four reaction with a single substrate and product each. Although 76.08 % of all models could be corrected for biological feasibility in these reactions, the challenge remains to detect similar criteria for the remaining reactions in order to refine the results of the Monte

Carlo experiment. So far, attempts to solve the resulting equations analytically were hampered by the complexity of the corresponding nonlinear equation system.

An interesting alternative approach could be based on random sampling of kinetic parameters instead of elasticities. The elasticities could then be calculated analytically using the partial derivatives of the rate laws in the steady state. This approach is similar to the sampling routine introduced by Wang *et al* (2004) for Michaelis-Menten kinetics. The analytical computation of elasticities requires prior knowledge about the catalytic mechanism and the corresponding type of rate law for each enzyme. However, it would offer a helpful extension of the filtering criterion introduced in this thesis for studying enzymes for which such prior knowledge about reaction mechanisms and suitable rate laws is available (Li *et al*, 2012).

Although elasticity checks were only possible for a limited number of reactions in the TCA cycle, an additional filtering criterion was introduced that was based on the FCCs belonging to each set of elasticities. In particular, it required that an increase in substrate input into the system had no negative implications for the overall flux through the TCA cycle.

Both types of filtering demonstrated how prior knowledge about a system could be incorporated (if available) in order to narrow down the combinations sampled elasticities to those that could actually occur in nature for the experimentally observed steady state. However, they should also serve as a reminder that the success rate of detecting biologically unrealistic models depends on the nature of the kinetic rate laws and that it remains a challenging task to select feasible SK-models if they contain complex enzymatic reactions.

7.1.2. Information that can be obtained from elasticities

The first intuitive approach in evaluating an SKM Monte Carlo experiment is to quantify the proportions of models with certain dynamic properties like stability of oscillations. This can provide information about the likely behaviour of a system for which the elasticities are unknown. For example, Steuer *et al* (2006) set all elasticities associated with product inhibition to equal values of increasing magnitude in order to show the general effects of product inhibition on stability. In order to understand the mechanisms involved in the emergence of certain dynamic properties, one can next attempt to infer causal relationships between the observed behaviour and the elasticities responsible for its emergence. This can be done by exploration of the sampled elasticities using supervised univariate or multivariate

statistical analyses.

Univariate comparisons of elasticities

In this thesis, elasticities were compared by the Kolmogorov-Smirnov test between stable and unstable, as well as between oscillatory and non-oscillatory models. Typically, distributions differed significantly in almost every elasticity. In order to obtain a fine-tuned picture of the individual contribution to the system's behaviour, they were ranked by the test statistics that estimated the maximum difference between the compared distribution functions.

Extension of the previously used univariate approaches for model evaluation by multivariate machine learning

In Section 5.3, the visual inspection of the elasticities obtained for the small example models showed that many elasticity distributions indeed differed strongly between stable and unstable models. However, it was also shown that most distributions widely covered the complete sampling interval. This implied that, even if an elasticity showed a strong tendency towards a specific value range in stable models, it was still possible to ensure stability if the elasticity obtained a value typically found among unstable models by an appropriate choice of the other elasticities in the system. This indicated that, for almost each value in each elasticity, a suitable set of values in the other elasticities existed that enabled the construction of a stable model.

We concluded that stability of a steady state was the result of a the fine-tuned interplay between different elasticities in the network. In order to elucidate possible combinations of elasticities responsible for stable or unstable models, the SKM framework was extended in this thesis by multivariate pattern search. In contrast to previous studies that used univariate tests to search for single, important parameters, this enabled the search for ensembles of enzymes and metabolites ensuring stability. This new approach allowed the determination of fine-tuned interactions between combinations of several enzymes and metabolites that could not be investigated by classical in vivo studies focusing only on a limited number of enzymes per experiment.

In doing so, the applicability of two supervised machine-learning approaches was evaluated. Because the goal was to obtain information about possible ways in which elasticities act together in a concerted manner, the most important aspect in

the choice of appropriate classifiers was the interpretability of the derived decision functions. This motivated the use of two classifiers, namely decision trees and RVMs.

The decision function, which forms the basis for assigning a class label to an unknown data point by a classifier, can either provide a representation of the boundaries between the different classes, or it can describe the 'typical' class centers and perform classification based on their similarities to an unknown data point.

Examples of classifiers of the first category are SVMs, perceptrons or decision trees (Hastie *et al*, 2009). In this thesis, decision trees were chosen because they offered the best interpretability of the derived decision function. In particular, the `C5.0` algorithm used in this work offered the additional feature of summarizing the paths through a tree by compact rulesets. These rulesets can be understood as 'patterns' of enzyme-metabolite relationships that mark the transition from stable to unstable steady states, or vice versa.

The second category of classifiers aims at detecting descriptions of class centres rather than boundaries. For example, the RVM algorithm provides a small set of the 'most typical' data points for each class (Tipping *et al*, 2003). These representative data points are called RVs. When classifying a new data point, the RVs then serve for estimating the probability of belonging to a particular class. A big advantage of RVMs is that they produce sparse solutions that contain as few RVs as possible. Therefore, the RV output tends to be smaller and easier to inspect visually than the large numbers of rulesets produced by decision trees. On the other hand, the quantitative rulesets returned by decision trees turned out to be more easily interpretable because they only focused on a small subset of 'important' elasticities whereas each RV contained the complete set of all elasticities in the system.

Another advantage of both classifiers compared to univariate comparisons was that they provided characteristic patterns per class instead of one decision boundary for both classes.

Classification by RVMs

When evaluated on three small example network with simple pathway structures, RVMs outperformed decision trees in terms of classification accuracy and the sparseness of the trained models. For example, RVM-based classification of sta-

bility for the linear pathway revealed 5-10 RVs, whereas decision trees consisted of hundredes of rulesets in order to accurately describe the differences between stable and unstable models. A property that made RVMs particularly interesting candidates for the evaluation of SKM experiments was that they outperformed decision trees in terms of classification accuracy even when training was based on much smaller sample sizes. This was a promising advantage for the analysis of systems of a larger scale, in which Monte Carlo experiments can be time consuming because of the runtime requirements that go along with the evaluation of large Jacobian matrices.

When applied to classify models of the TCA cycle, however, it turned out that RVMs were outperformed by decision trees both in terms of runtime and classification results (Section 6.4.1). A possible explanation could be that the TCA cycle posed a more challenging classification problem due to the larger number of sampled elasticities and the more complex pathway structure. The observed high training errors indicated a high model bias, which resulted from difficulties to detect informative RVs during training. Additionally, the observed test errors distinctly exceeded the training errors, showing that the derived decision functions were prone to overfitting, and consequently not well generalizable to unknown test points. Because of these observations, RVMs could not be considered further for analysing the SK-models of the TCA cycle. It could be possible, however, that larger training samples would lead to improvements in classification performance and that the availability of faster computers and algorithms could enable RVM based classification of large SK-models in the future. For example, there are approaches to improve the runtime of RVMs (Yang *et al*, 2010) but a discussion of this topic goes beyond the scope of this thesis.

Classification by decision trees

Using decision trees to classify sampled elasticities for stable and unstable models produced good results in terms of model bias and overfitting. This observation was consistent for the small example pathways (Section 5.4.1) and for the model of mitochondrial metabolism (Section 6.4.2). Prediction of oscillations generally tended to be more challenging than stability prediction.

In contrast to RVM output, decision tree rulesets could be filtered for reliability by evaluating them on an independent test data set and inspecting the resulting Laplace values (Section 3.2.3). In doing so, identification of typical patterns for a

specific system property could focus on the best performing rulesets. For example, the most reliable rulesets were evaluated statistically to detect the most abundant elasticities together with tendencies towards low or high values in a particular class. Selected rulesets were also evaluated individually to investigate interesting scenarios in which instabilities could arise.

For the small example pathways, it was demonstrated how these rulesets were translated back into thresholds for kinetic parameters. Such thresholds can potentially be used to simulate trajectories of perturbation responses. In contrast, the complex nature of the rate laws in the TCA cycle model rendered it difficult to derive kinetic parameter thresholds in the same straight-forward manner. Instead, an alternative approach for the investigation of trajectories in unstable models was introduced (Section 6.4.5) based on the investigation of model eigenvectors fulfilling the conditions in a simple unstable ruleset. Each eigenvector belonging to a positive eigenvalue of the Jacobian matrix then indicated the time-dependent development of the trajectories in the linearised system. In doing so, it was possible to get a general impression of the system's behaviour around an unstable steady state. One has to keep in mind, however, that we can only analyse the instabilities in the linearised system, which could differ from those in the true system. Furthermore, it is not possible to infer the longterm behaviour of the trajectories when moving away from an unstable state. This would require simulations by a kinetic model, with specification of the rate laws and their kinetic parameters.

When investigating patterns associated with certain dynamic steady state properties, one has to keep in mind that not all randomly sampled models are covered by rulesets with high Laplace values. Consequently, there might be unstable models for which reliable criteria could not be found by the decision tree algorithm. Decision tree classification is intended to look at those scenarios that are easy to detect and to describe by simple patterns, instead of providing a comprehensive decision function able to classify any new data point.

7.1.3. Converting elasticities to FCCs

Principles of deriving FCCs from elasticities A particularly interesting property of elasticities is the ability to convert them into FCCs (Westerhoff and Kell, 1987; Wang *et al*, 2004). This implies that, for each set of sampled elasticities in an SKM experiment, we can obtain a detailed picture of the corresponding proportions of control exhibited by each enzyme on each steady state flux. The re-

lationship between sampled elasticities and corresponding flux control coefficients has been extensively studied in the related field of MCA. In this thesis, FCCs were derived from the sampled elasticities and their usability for describing easily interpretable conditions for instabilities or oscillations was evaluated.

Do FCCs provide sufficient information to predict stability of a steady state? Decision tree training revealed that the quality of the detected FCC patterns differed strongly depending on the complexity of the pathway structure. Classification performance for two simple pathway structures (linear and branched pathway) exceeded that of elasticity based training in terms of training and test errors. However, for the circular pathway structures of a small cyclic pathway and of the TCA cycle, elasticities turned out to be superior in terms of their information content for classification and orchestrated pattern detection.

The differences in classification between elasticities and FCCs seemed surprising at first, because the MCA framework explains how both features can be interconverted without loss of information. However, an additional processing step was performed before classification that summarized FCCs explaining the forward- and reverse-effects. Therefore, a loss of information was taken into account when calculating FCCs from elasticities. Using the full FCCs on the other hand, would not only make the interpretation of the resulting rulesets difficult, but also inflate the feature space.

7.1.4. Elucidating stabilizing mechanisms in a metabolic system

As stated in the introduction, local stability can be understood as the robustness of a steady state to small perturbations that can occur easily in a cellular environment. From this we can conclude that it would be impossible to observe unstable steady states in a metabolic system, because trajectories would diverge from it due to natural perturbations. This implies that it would be disadvantageous for a metabolic system if the operating state of a pathway best suited to fulfil a certain functional role is unstable and cannot be reached. Stability analyses in this work therefore focused on the detection of the stabilizing mechanisms present in the investigated systems, as well as scenarios in which possible mutations could alter kinetic properties in ways that could be hazardous for the stability of functional working states.

Lessons learnt from the small example pathways

Pairwise comparisons of the elasticity distributions between stable and unstable models of the small example pathways emphasized the roles of the positive feedback term and the efflux out of the system. The elasticities associated with the efflux reactions were typically located at lower values for unstable models than for stable models, indicating that fast responses of these reactions were required in order to maintain stability. This was confirmed by unstable decision tree and RVM patterns which generally described slow perturbation responses of efflux reactions but fast propagation of perturbations by the feedback term. This agrees with earlier results obtained by analysing the Calvin Benson Cycle (CBC) in plant metabolism, where allosteric regulators were shown to strongly influence stability (Girbig *et al*, 2012a).

Stability analysis of mitochondrial metabolism

A detailed SK-model of mitochondrial energy metabolism was presented and analysed in this work. It covered the TCA cycle, the respiratory chain, ATP synthesis, as well as ATP exchange with the cytosol. In total, it comprised 24 metabolites, 20 reactions and 71 sampled elasticities. In order to observe possible shifts in control during increased workload on the mitochondrion, two steady states were analysed that varied in the amount of cytosolic ATP consumption. Monte Carlo sampling of SK-models revealed strong tendencies towards stability in both steady states, and the occurrence of oscillatory trajectories in $35 - 40\%$ of all models.

Both the univariate comparisons and the decision tree rulesets showed that the most important elasticities for reference state stability were associated with reactions upstream of the TCA cycle. We detected no single elasticity that was required in every stabilizing or destabilizing ruleset. Instead, the investigated system properties could both be invoked by diverse combinations of elasticities associated with PDH, AKGDH, CITS and IDH. This coincides with observations that enzymes of the TCA cycle and respiratory chain are typically modulated together by external stimuli like Ca^{2+} uptake and that they share the control of flux through the system (Koopman *et al*, 2012). In order to systematically assess the impact of these stimuli on the system, the model presented in this work could be extended by the corresponding regulatory interactions and analysed with respect to the emerging changes in stabilizing and destabilizing patterns.

Ruleset analysis additionally showed that the high ATP demand under gamma oscillations lead to a shift in control that included reactions involved in ATP

synthesis. Additionally, the statistical analysis of the ruleset conditions revealed that a slow response of AKGDH to perturbations was a serious threat for stability. Eigenvector analysis confirmed that indeed the simplest way to induce instabilities was to slow down the responses of AKGDH to perturbations.

Overall, the decision tree algorithm discovered more rulesets reliably causing instability than ensuring stability (see Table 6.4 in Chapter 6). This showed that the reasons for instabilities in this system could be diverse and comprise different combinations of elasticities. Interestingly, this observation differs from the results previously obtained for the CBC (Girbig *et al*, 2012a), where the number of reliable destabilizing patterns (3 rulesets for a training size of 10,000) was found to be distinctly lower than the number of stabilizing ones (62 rulesets). However, because the overall proportion of unstable SK-models of the CBC (16 %) was lower than for the TCA cycle (< 1 %), these differences in ruleset numbers could be explained by a generally higher vulnerability of the CBC model to instabilities. As a consequence, elasticity modifications leading to instabilities in the CBC model would not need to be as fine-tuned as for the TCA cycle. However, the study by Girbig *et al* (2012a) also showed that the incorporation of additional allosteric regulators, which were not part of established kinetic models, could significantly decrease the occurrence of unstable models to $\leqslant 3\%$ in SKM experiments. It would therefore be interesting to repeat the decision tree analysis of the CBC on SK-models that take these regulators into account in order to test if they increase the complexity of the destabilizing patterns.

Limitations of decision tree-based stability analysis

We have to keep in mind that the decision tree rulesets do not explain all instabilities that can occur in mitochondrial energy metabolism for two reasons:

1. Although a large number of SK-models was created during Monte Carlo sampling, it could be possible that certain elasticity combinations leading to instabilities were not included. In fact, due to the high dimensionality of the sampling space that consisted of all possible value combinations for 71 elasticities, even a set of 20,000 SK-models only enabled a sparse coverage.

2. Focusing on the most reliable rulesets with Laplace values $\geqslant 0.95$ implied that the majority of rulesets was discarded after training (see also Table D.2 (a) in the Appendix). Although the rulesets was not mutually exclusive, so

that most elasticity combinations were represented by more than one ruleset, it is possible that there were some SK-models with elasticities not meeting any criteria of any remaining ruleset.

Therefore, the goal of the presented analysis was to find simple and reproducible mechanisms that could impair stability, not cover the whole range of theoretically possible destabilizing mechanisms.

7.1.5. Conditions for oscillatory trajectories around steady states

Analysis of the small example pathways revealed that conditions for oscillations were more complex than conditions for instabilities. Statistical evaluation of the most reliable rulesets for oscillatory and non-oscillatory trajectories in the small example pathways showed that oscillating systems favoured large values in the relevant elasticities. This explains the strong increase in oscillatory models due to elasticity filtering, because the filtering criterion in equation (3.1) favoured large elasticity values.

When analysing the TCA cycle model, the main interest lay on the detection of possible destabilizing mutations. However, in systems for which the study of oscillations is of particular interest, a similar procedure could be performed with focus on the comparisons of oscillatory versus non-oscillatory models instead of stable versus unstable ones. For example, oscillations in glycolysis have been studied intensively by Heinrich and Schuster (1996). Therefore, it would be of great interest to extend the knowledge obtained so far by analysing oscillatory decision tree rulesets.

Despite the gain of mechanistic insights that could explain possible reasons for observed oscillations in a system, oscillation prediction can also render itself helpful for predicting whether experimental values can be expected to fluctuate during measurements.

7.2. Conclusions and outlook

The fast progress in metabolomics technologies has enabled the collection of vast amounts of data about metabolites and fluxes within cellular systems in the recent years. In order to obtain a full understanding of the ways in which metabolism

is organized and controlled, however, we need to extend the established statistical approaches for analysing and comparing measured metabolite profiles (Boccard *et al*, 2010; Kholodenko *et al*, 2012; Xia *et al*, 2012; Franceschi *et al*, 2013) by taking into account the underlying network structure and dynamics. However, elucidating the dynamics of complex metabolic systems remains a big challenge in those cases in which detailed kinetic models are not available. The SKM framework (Steuer *et al*, 2006) allows the investigation of the mechanisms that control the system's behaviour around experimentally observed steady states without requiring estimates of all rate laws and kinetic parameters involved. This makes it a promising approach for the study of large metabolic networks that typically have many unknowns in terms of kinetic parameters and rate laws. It is important to note that this work does not focus on the types and strengths of changes in order to alter flux (Stitt *et al*, 2010) but instead on the intrinsic mechanisms protecting stability and ensuring that the system only significantly changes in response to targeted and coordinated signals.

It has already been demonstrated in the past how SKM can help to detect single stabilizing sites in metabolic networks (Grimbs *et al*, 2007) and how this information can be used to aid in the construction of new kinetic models (Bulik *et al*, 2009). In this thesis, I investigated how the sampling procedure can be refined in certain cases, and how multivariate pattern search can help to obtain a more detailed picture of possible reasons for instabilities of a metabolic steady state.

We have to keep in mind, however, that the applicability of the SKM approach depends on the scope and on the quality of available data. Today, obtaining accurate quantitative measurements of concentrations and fluxes is still a challenging task (Maier *et al*, 2013; Da Silva *et al*, 2013; Sellami *et al*, 2013). In the related work of Wang *et al* (2004), this problem was circumvented by estimating steady state fluxes by FBA (Orth *et al*, 2010), and by sampling the concentrations randomly. van Nes *et al* (2009) avoid the incorporation of steady state data by sampling the components of the Jacobian matrix directly. These approaches enabled to obtain insights about the general capacities of a network based on their stoichiometry and (in the case of Wang *et al* (2004)) on prior knowledge about all rate laws and kinetic parameters. In contrast, the crucial advantage of SKM is the possibility to analyse local dynamic properties of actually observed metabolic steady states. After incorporating the observed steady state data into the stoichiometric matrix, we only need to sample the elasticities from predefined distributions

in order to obtain and evaluate the corresponding Jacobian matrices of the system.

Ultimately, we can conclude that the choice of a suitable method for investigating the dynamic properties of a metabolic system in steady state relies on the availability of suitable data and on the depth of information that should be achieved. If steady state information is only partially available, it could also be possible to combine metabolite measurements with estimated steady state fluxes. For example, Maier *et al* (2013) describe a large study in which all detectable metabolites of *Mycoplasma pneumoniae* were measured and partly quantified with respect to their absolute concentrations. These quantitative measurements could be combined with FBA fluxes in order to perform a large-scale SKM analysis.

The ability to detect sets of enzymes that exhibit control on the system's dynamic properties in a coordinated manner could aid in the design of systems with desired behaviours via synthetic biology (Kruse, 2010). Another interesting field of research is the investigation of the genomic organization of the involved enzymes. It was recently postulated that genes for metabolic enzymes are organized in an economical way so that transcriptional regulation is sparse and highly selective in cost-intensive pathways (Wessely *et al*, 2011). A related question to investigate would be whether similar preferences for coordinated regulation of selected genes can also be detected for enzymes jointly involved in the control of dynamic system properties. We can conclude that, with metabolomics technologies becoming more and more refined, SKM is a promising technique to explore the dynamics of systems of larger scales in the near future.

Appendices

A. List of Abbreviations

A.1. General abbreviations

BER	Balanced error rate
CBC	Calvin-Benson cycle
FBA	Flux balance analysis
FCC	Flux control coefficient
MCA	Metabolic control analysis
ODE	Ordinary differential equation
SKM	Structural kinetic modelling
SK-model	Structural kinetic model
TCA cycle	Tricarboxylic acid cycle (also called citric acid cycle)

A.2. Enzyme names

ACON	Aconitase
AKGDH	α-Ketoglutarate dehydrogenase
ATP/ADP-anti	ATP-ADP antiport
ATP-syn	ATP synthase
ATP-use	Cytosolic ATP consumption
CI	Complex I
CIII	Complex III
CIV	Complex IV
CITS	Citrate synthase
FU	Fumarase
H-leak	Proton leak
IDH	Isocitrate dehydrogenase
K/H-anti	K^+-proton antiport
MDH	Malate dehydrogenase
Na/H-anti	Na^+-proton antiport
PDH	Pyruvate dehydrogenase
P/H-sym	Phosphate-proton symport
PYR/H-sym	Pyruvate-proton symport
SDH	Succinate dehydrogenase (also called complex II)
SCS	Succinyl-CoA synthetase

A.3. Metabolite names

ACCOA	Acetyl-CoA
AKG	α-Ketoglutarate
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
CIT	Citrate
COA	Coenzyme A
$CytC^{(ox)/(red)}$	Cytochrome C (oxidised / reduced)
FUM	Fumarate
H	Proton ion
ICIT	Isocitrate
MAL	Malate
NAD	Nicotinamide adenine dinucleotide (oxidised)
NADH	Nicotinamide adenine dinucleotide (reduced)
OAA	Oxalacetate
P	Inorganic phosphate
PYR	Pyruvate
Q / QH2	Ubiquinone/ ubiquinol
SUC	Succinate
SUCCOA	Succinyl-CoA

B. Notation

m	Number of metabolites in the network
n	Number of reactions in the network
S_i	Name of the i^{th} metabolite in the network ($i = 1, 2, \ldots m$)
r_j	Name of the j^{th} reaction in the network ($j = 1, 2, \ldots n$)
$\mathbf{N} \in \mathbb{R}^{m \times n}$	Stoichiometric matrix of the metabolic network
$\mathbf{J} \in \mathbb{R}^{m \times m}$	Jacobian matrix of the metabolic network
$[S_i]$	Concentration of metabolite S_i (in units of $\frac{\text{mol}}{\text{l}}$, if not stated otherwise)
$v_j := v_j([\mathbf{S}])$	Rate of reaction r_j (in units of $\frac{\text{mol}}{\text{l}\cdot\text{sec}}$, if not stated otherwise), typically dependent on metabolite concentrations $[\mathbf{S}]$ and kinetic parameters $\mathbf{k}$
$f_i = f_i([\mathbf{S}](t))$	Time-dependent changes in metabolite S_i
$[S_i]^*$	Steady state concentration of metabolite S_i
$v_j^* = v_j([\mathbf{S}]^*)$	Steady state rate of reaction r_j
$x_i = \frac{[S_i]}{[S_i]^*}$	Concentration of metabolite S_i normalized to the steady state
$\mu_j = \frac{v_j}{v_j^*}$	Rate of reaction r_j normalized to the steady state
$V_{max}^{v_j}$	Maximum velocity parameter of the Michaelis Menten equation for reaction r_j
$K_M^{v_j}$	Michaelis constant of the Michaelis Menten equation for reaction r_j. Interpretation: substrate concentration of at which v_j has reached half of its maximum velocity
F	Steady state flux through the metabolic system

e_j	Concentration of the enzyme catalysing reaction r_j
$C_{e_j}^{F}$	Flux control coefficient quantifying the control of reaction r_j on flux F
$\varepsilon_{\mathrm{S_i}}^{\mathrm{v_j}}$	Elasticity coefficient quantifying the influence of the normalized concentration $[S_i]$ on the normalized reaction rate v_j in the steady state
$\mathbf{X} = \{\chi_i\}_{i=1}^{q}$	Input data set for the binary classification problem ($\chi_i \in \mathbb{R}^p$)
$\mathbf{T} = \{t_i\}_{i=1}^{q}$	Binary class labels associated with the input data serving as targets for the classification problem ($t_i \in \{0, 1\}$)

C. Additional results (example pathways)

C.1. Additional results of decision tree classification

C.1.1. Balanced error rates for increasing sample size

The following tables show the balanced error rates (mean ± standard deviation) for increasing training data size, determined by training five individual decision trees on independent training sets of each size. All values have been rounded to three decimal places.

Balanced error rates of the linear pathway

	Elasticities		Flux control coefficients	
TrainingSize	Stability	Oscillations	Stability	Oscillations
50	0.008 +- 0.011	0.06 +- 0.032	0 +- 0	0.088 +- 0.046
100	0.02 +- 0.012	0.052 +- 0.052	0 +- 0	0.102 +- 0.053
200	0.015 +- 0.007	0.06 +- 0.048	0 +- 0	0.144 +- 0.039
500	0.018 +- 0.008	0.049 +- 0.019	0 +- 0	0.15 +- 0.006
1000	0.014 +- 0.003	0.037 +- 0.007	0 +- 0	0.166 +- 0.023
2000	0.015 +- 0.002	0.052 +- 0.006	0 +- 0	0.161 +- 0.01
5000	0.012 +- 0.001	0.045 +- 0.005	0 +- 0	0.161 +- 0.012
10000	0.013 +- 0.001	0.044 +- 0.003	0 +- 0	0.164 +- 0.004
15000	0.011 +- 0.001	0.042 +- 0.001	0 +- 0	0.164 +- 0.004
20000	0.011 +- 0.001	0.039 +- 0.001	0 +- 0	0.166 +- 0.004

(a) Classification errors determined on the training data.

	Elasticities		Flux control coefficients	
TrainingSize	Stability	Oscillations	Stability	Oscillations
50	0.127 +- 0.011	0.301 +- 0.038	0.018 +- 0.016	0.277 +- 0.033
100	0.111 +- 0.023	0.254 +- 0.046	0.015 +- 0.006	0.274 +- 0.028
200	0.088 +- 0.014	0.218 +- 0.01	0.004 +- 0.003	0.239 +- 0.016
500	0.066 +- 0.009	0.186 +- 0.01	0.002 +- 0.001	0.207 +- 0.01
1000	0.053 +- 0.006	0.168 +- 0.006	0.001 +- 0.001	0.207 +- 0.008
2000	0.047 +- 0.004	0.143 +- 0.001	0 +- 0	0.193 +- 0.004
5000	0.039 +- 0.002	0.122 +- 0.002	0 +- 0	0.185 +- 0.004
10000	0.034 +- 0.001	0.106 +- 0.002	0 +- 0	0.181 +- 0.001
15000	0.031 +- 0.001	0.097 +- 0.002	0 +- 0	0.18 +- 0.002
20000	0.029 +- 0.001	0.094 +- 0.001	0 +- 0	0.179 +- 0.001

(b) Classification errors determined on the test data.

Table C.1.: Balanced error rates of decision tree classification in the linear pathway.

Balanced error rates of the branched pathway

	Elasticities		Flux control coefficients	
TrainingSize	Stability	Oscillations	Stability	Oscillations
50	0.02 +- 0.02	0.024 +- 0.017	0.008 +- 0.011	0.024 +- 0.022
100	0.026 +- 0.019	0.042 +- 0.019	0 +- 0	0.024 +- 0.005
200	0.013 +- 0.006	0.053 +- 0.015	0.003 +- 0.003	0.039 +- 0.011
500	0.009 +- 0.003	0.039 +- 0.005	0.001 +- 0.001	0.034 +- 0.005
1000	0.011 +- 0.002	0.042 +- 0.003	0.002 +- 0.001	0.042 +- 0.009
2000	0.013 +- 0.003	0.049 +- 0.006	0.002 +- 0.001	0.057 +- 0.007
5000	0.01 +- 0.001	0.042 +- 0.003	0.002 +- 0.001	0.057 +- 0.003
10000	0.01 +- 0	0.04 +- 0.002	0.003 +- 0	0.057 +- 0.003
15000	0.009 +- 0.001	0.036 +- 0.001	0.002 +- 0.001	0.058 +- 0.003
20000	0.009 +- 0.001	0.034 +- 0.002	0.002 +- 0	0.059 +- 0.002

(a) Classification errors determined on the training data.

	Elasticities		Flux control coefficients	
TrainingSize	Stability	Oscillations	Stability	Oscillations
50	0.152 +- 0.063	0.235 +- 0.038	0.049 +- 0.027	0.139 +- 0.031
100	0.109 +- 0.018	0.213 +- 0.03	0.021 +- 0.009	0.116 +- 0.011
200	0.078 +- 0.008	0.186 +- 0.007	0.018 +- 0.009	0.116 +- 0.01
500	0.066 +- 0.008	0.154 +- 0.009	0.007 +- 0.002	0.093 +- 0.002
1000	0.051 +- 0.003	0.14 +- 0.003	0.006 +- 0.001	0.092 +- 0.007
2000	0.044 +- 0.002	0.129 +- 0.003	0.006 +- 0.001	0.084 +- 0.005
5000	0.034 +- 0.002	0.109 +- 0.005	0.004 +- 0	0.078 +- 0.002
10000	0.029 +- 0.001	0.099 +- 0.002	0.004 +- 0	0.073 +- 0.001
15000	0.027 +- 0.001	0.092 +- 0.002	0.004 +- 0	0.072 +- 0.001
20000	0.024 +- 0.001	0.087 +- 0.002	0.004 +- 0	0.071 +- 0.001

(b) Classification errors determined on the test data.

Table C.2.: Balanced error rates of decision tree classification in the branched pathway.

Balanced error rates of the cyclic pathway

	Elasticities		Flux control coefficients	
TrainingSize	Stability	Oscillations	Stability	Oscillations
50	0.016 +- 0.026	0.028 +- 0.027	0.052 +- 0.023	0.104 +- 0.071
100	0.022 +- 0.004	0.038 +- 0.019	0.044 +- 0.021	0.128 +- 0.057
200	0.011 +- 0.004	0.036 +- 0.011	0.075 +- 0.021	0.167 +- 0.051
500	0.012 +- 0.005	0.026 +- 0.005	0.078 +- 0.024	0.172 +- 0.033
1000	0.009 +- 0.004	0.028 +- 0.006	0.082 +- 0.019	0.19 +- 0.023
2000	0.01 +- 0.003	0.031 +- 0.006	0.098 +- 0.008	0.198 +- 0.011
5000	0.009 +- 0.002	0.028 +- 0.005	0.103 +- 0.004	0.2 +- 0.006
10000	0.007 +- 0.002	0.025 +- 0.003	0.1 +- 0.006	0.199 +- 0.003
15000	0.007 +- 0.001	0.024 +- 0.002	0.101 +- 0.004	0.193 +- 0.005
20000	0.007 +- 0	0.025 +- 0.002	0.1 +- 0.002	0.198 +- 0.003

(a) Classification errors determined on the training data.

	Elasticities		Flux control coefficients	
TrainingSize	Stability	Oscillations	Stability	Oscillations
50	0.113 +- 0.016	0.203 +- 0.024	0.23 +- 0.048	0.299 +- 0.042
100	0.11 +- 0.041	0.19 +- 0.025	0.177 +- 0.021	0.262 +- 0.013
200	0.067 +- 0.011	0.147 +- 0.013	0.17 +- 0.017	0.249 +- 0.011
500	0.055 +- 0.005	0.12 +- 0.003	0.146 +- 0.009	0.241 +- 0.013
1000	0.045 +- 0.005	0.108 +- 0.007	0.132 +- 0.004	0.226 +- 0.004
2000	0.038 +- 0.004	0.094 +- 0.004	0.129 +- 0.006	0.218 +- 0.006
5000	0.028 +- 0.002	0.078 +- 0.001	0.122 +- 0.003	0.215 +- 0.003
10000	0.023 +- 0.001	0.068 +- 0.001	0.116 +- 0.002	0.209 +- 0.003
15000	0.021 +- 0	0.062 +- 0.001	0.115 +- 0.002	0.208 +- 0.001
20000	0.02 +- 0.001	0.059 +- 0.001	0.113 +- 0.001	0.209 +- 0.001

(b) Classification errors determined on the test data.

Table C.3.: Balanced error rates of decision tree classification in the cyclic pathway.

C.1.2. Ruleset numbers and sizes

The following tables show the numbers and sizes of the decision tree rulesets obtained for selected training data size, determined by training five individual decision trees on independent training sets of each size. All numbers are given in the format x/y, where x and y are the numbers of rulesets with Laplace values $\geqslant 0.95$, or $\geqslant 0.5$, respectively. Observed minimum and maximum ruleset sizes are given in brackets. Decision trees have been trained either using elasticities (left) or flux control coefficients (right) as features.

Ruleset statistics derived for the linear pathway

	Elasticities		Flux Control Coefficients	
TrainingSize	Stability	Instability	Stability	Instability
200	17/23 (1, 3)	5/17 (3, 4)	5/5 (1, 1)	5/5 (1, 1)
500	33/41 (1, 3)	20/26 (2, 4)	5/5 (1, 1)	5/5 (1, 1)
1000	47/53 (1, 4)	20/30 (2, 5)	5/5 (1, 1)	5/5 (1, 1)
5000	153/157 (1, 5)	38/46 (3, 6)	5/5 (1, 1)	5/5 (1, 1)
10000	219/224 (1, 6)	67/78 (2, 6)	5/5 (1, 1)	5/5 (1, 1)
20000	453/455 (1, 6)	68/79 (3, 7)	5/5 (1, 1)	5/5 (1, 1)

(a) Classification of stable versus unstable states.

	Elasticities		Flux Control Coefficients	
TrainingSize	No oscillations	Oscillations	No oscillations	Oscillations
200	0/36 (0, 0)	0/10 (0, 0)	3/17 (1, 2)	0/12 (0, 0)
500	5/63 (1, 4)	0/30 (0, 0)	9/30 (1, 2)	0/11 (0, 0)
1000	42/116 (1, 5)	0/49 (0, 0)	11/33 (1, 3)	0/16 (0, 0)
5000	184/301 (1, 7)	15/130 (5, 8)	38/77 (1, 3)	0/22 (0, 0)
10000	414/559 (1, 7)	59/226 (5, 9)	56/106 (1, 4)	0/21 (0, 0)
20000	794/953 (1, 7)	132/341 (5, 10)	74/132 (1, 4)	0/29 (0, 0)

(b) Classification of oscillating versus non-oscillating models.

Table C.4.: Observed numbers and sizes of the rulesets derived for the linear pathway.

Ruleset statistics derived for the branched pathway

	Elasticities		Flux Control Coefficients	
TrainingSize	Stability	Instability	Stability	Instability
200	20/26 (1, 2)	7/14 (2, 4)	4/5 (2, 2)	5/9 (1, 1)
500	28/35 (1, 5)	7/16 (2, 5)	5/5 (2, 2)	5/11 (1, 1)
1000	47/53 (1, 5)	32/37 (2, 6)	5/5 (2, 2)	5/10 (1, 1)
5000	125/130 (1, 5)	45/53 (3, 7)	4/5 (2, 2)	6/17 (1, 2)
10000	226/226 (1, 6)	43/48 (3, 7)	3/5 (2, 2)	5/19 (1, 1)
20000	395/395 (1, 6)	73/79 (2, 7)	2/7 (2, 2)	14/36 (1, 3)

(a) Classification of stable versus unstable states.

	Elasticities		Flux Control Coefficients	
TrainingSize	No oscillations	Oscillations	No oscillations	Oscillations
200	3/25 (2, 2)	0/24 (0, 0)	2/13 (1, 2)	4/15 (1, 2)
500	18/46 (1, 4)	1/37 (4, 4)	10/29 (1, 3)	8/18 (1, 4)
1000	26/72 (1, 4)	0/35 (0, 0)	16/33 (1, 4)	18/29 (1, 4)
5000	155/228 (1, 7)	31/157 (3, 7)	40/50 (1, 7)	33/50 (1, 4)
10000	269/367 (1, 8)	82/249 (3, 9)	62/73 (1, 6)	56/70 (1, 7)
20000	525/683 (1, 8)	270/529 (3, 9)	68/78 (2, 7)	78/96 (1, 7)

(b) Classification of oscillating versus non-oscillating models.

Table C.5.: Observed numbers and sizes of the rulesets derived for the branched pathway.

Ruleset statistics derived for the cyclic pathway

	Elasticities		Flux Control Coefficients	
TrainingSize	Stability	Instability	Stability	Instability
200	19/23 (1, 2)	7/14 (3, 4)	2/14 (3, 3)	1/19 (1, 1)
500	33/34 (1, 3)	5/13 (2, 4)	3/14 (2, 5)	3/25 (1, 1)
1000	43/46 (1, 4)	25/30 (2, 5)	4/20 (2, 4)	3/42 (1, 2)
5000	129/131 (1, 5)	38/44 (2, 5)	13/31 (3, 5)	3/36 (1, 1)
10000	197/200 (1, 5)	64/72 (3, 7)	30/64 (2, 6)	5/70 (1, 2)
20000	332/332 (1, 6)	57/64 (3, 6)	39/85 (2, 6)	7/78 (1, 3)

(a) Classification of stable versus unstable states.

	Elasticities		Flux Control Coefficients	
TrainingSize	No oscillations	Oscillations	No oscillations	Oscillations
200	11/28 (1, 2)	0/16 (0, 0)	0/10 (0, 0)	0/11 (0, 0)
500	34/62 (1, 4)	1/25 (4, 4)	0/12 (0, 0)	0/17 (0, 0)
1000	51/91 (1, 5)	1/24 (3, 3)	0/15 (0, 0)	0/16 (0, 0)
5000	218/272 (1, 7)	1/37 (5, 5)	1/17 (2, 2)	0/20 (0, 0)
10000	382/431 (1, 7)	13/89 (4, 8)	3/29 (3, 4)	0/25 (0, 0)
20000	665/725 (1, 7)	54/133 (4, 10)	4/48 (2, 5)	0/42 (0, 0)

(b) Classification of oscillating versus non-oscillating models.

Table C.6.: Observed numbers and sizes of the rulesets derived for the cyclic pathway.

C.2. Additional results of RVM classification

C.2.1. Balanced error rates for increasing sample size

The following tables show the balanced error rates (mean ± standard deviation) for increasing training data size, determined by training five individual RVMs on independent training sets of each size. All values have been rounded to three decimal places.

	Elasticities		Flux control coefficients	
TrainingSize	Stability	Oscillations	Stability	Oscillations
50	0.008 +- 0.011	0.06 +- 0.042	0 +- 0	0.136 +- 0.036
100	0.008 +- 0.004	0.06 +- 0.046	0 +- 0	0.162 +- 0.031
200	0.006 +- 0.005	0.018 +- 0.016	0 +- 0	0.188 +- 0.025
500	0.015 +- 0.004	0.02 +- 0.006	0.001 +- 0.002	0.17 +- 0.007
1000	0.002 +- 0.001	0.015 +- 0.004	0 +- 0.001	0.16 +- 0.006

(a) Classification errors determined on the training data.

	Elasticities		Flux control coefficients	
TrainingSize	Stability	Oscillations	Stability	Oscillations
50	0.054 +- 0.017	0.186 +- 0.03	0 +- 0	0.251 +- 0.031
100	0.044 +- 0.007	0.15 +- 0.02	0 +- 0	0.217 +- 0.009
200	0.029 +- 0.004	0.11 +- 0.017	0 +- 0	0.201 +- 0.017
500	0.02 +- 0.001	0.066 +- 0.003	0 +- 0	0.189 +- 0.002
1000	0.014 +- 0.002	0.049 +- 0.003	0 +- 0	0.182 +- 0.003

(b) Classification errors determined on the test data.

Table C.7.: Balanced error rates of RVM classification of linear pathway models.

	Elasticities		Flux control coefficients	
TrainingSize	Stability	Oscillations	Stability	Oscillations
50	0 +- 0	0.028 +- 0.03	0.016 +- 0.026	0.036 +- 0.033
100	0 +- 0	0.032 +- 0.024	0.004 +- 0.009	0.066 +- 0.029
200	0.002 +- 0.003	0.054 +- 0.019	0.001 +- 0.002	0.061 +- 0.012
500	0.003 +- 0.005	0.016 +- 0.013	0.001 +- 0.002	0.062 +- 0.008
1000	0.001 +- 0.001	0.011 +- 0.004	0 +- 0	0.061 +- 0.006

(a) Classification errors determined on the training data.

	Elasticities		Flux control coefficients	
TrainingSize	Stability	Oscillations	Stability	Oscillations
50	0.031 +- 0.003	0.143 +- 0.052	0.028 +- 0.011	0.094 +- 0.009
100	0.027 +- 0.006	0.106 +- 0.02	0.017 +- 0.009	0.086 +- 0.004
200	0.019 +- 0.003	0.09 +- 0.006	0.009 +- 0.004	0.08 +- 0.005
500	0.011 +- 0.003	0.073 +- 0.005	0.006 +- 0.001	0.073 +- 0.003
1000	0.008 +- 0.001	0.045 +- 0.001	0.006 +- 0.002	0.069 +- 0.001

(b) Classification errors determined on the test data.

Table C.8.: Balanced error rates of RVM classification of branched pathway models.

	Elasticities		Flux control coefficients	
TrainingSize	Stability	Oscillations	Stability	Oscillations
50	0 +- 0	0.028 +- 0.033	0.5 +- 0	0.188 +- 0.05
100	0 +- 0	0.038 +- 0.033	0.5 +- 0	0.24 +- 0.032
200	0.001 +- 0.002	0.032 +- 0.01	0.274 +- 0.207	0.21 +- 0.017
500	0.004 +- 0.003	0.017 +- 0.005	0.136 +- 0.02	0.226 +- 0.012
1000	0.004 +- 0.002	0.009 +- 0.007	0.126 +- 0.01	0.212 +- 0.018

(a) Classification errors determined on the training data size.

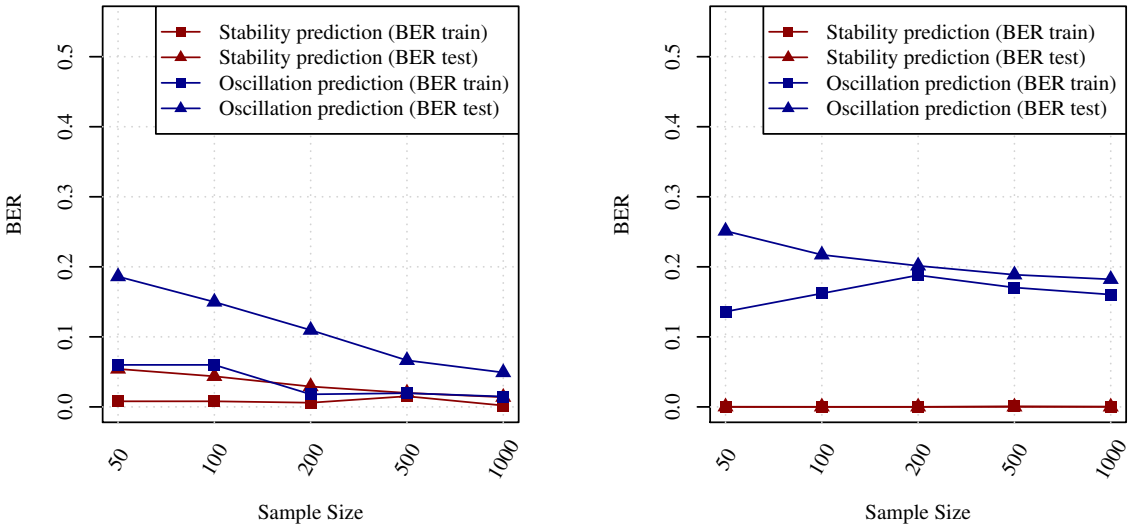
	Elasticities		Flux control coefficients	
TrainingSize	Stability	Oscillations	Stability	Oscillations
50	0.052 +- 0.014	0.139 +- 0.021	0.498 +- 0.001	0.252 +- 0.012
100	0.041 +- 0.008	0.092 +- 0.027	0.498 +- 0.001	0.23 +- 0.005
200	0.018 +- 0.006	0.078 +- 0.01	0.289 +- 0.193	0.223 +- 0.005
500	0.011 +- 0.001	0.057 +- 0.008	0.142 +- 0.003	0.22 +- 0.007
1000	0.008 +- 0	0.043 +- 0.003	0.134 +- 0.007	0.215 +- 0.003

(b) Classification errors determined on the test data.

Table C.9.: Balanced error rates of RVM classification of cyclic pathway models.

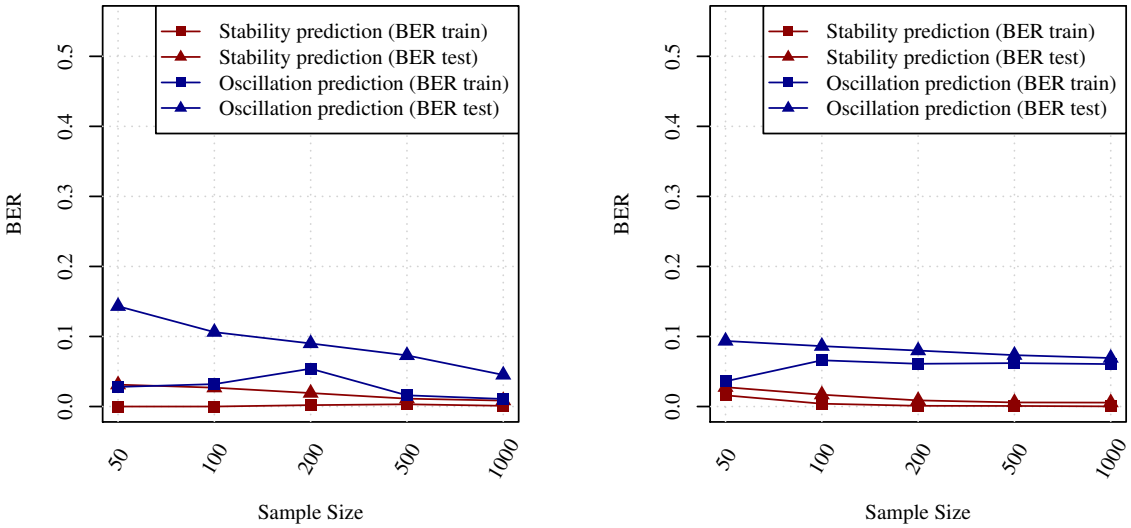
C.2.2. Learning curves

This Section illustrates the learning curves with BER values listed in Tables C.7 - C.9.



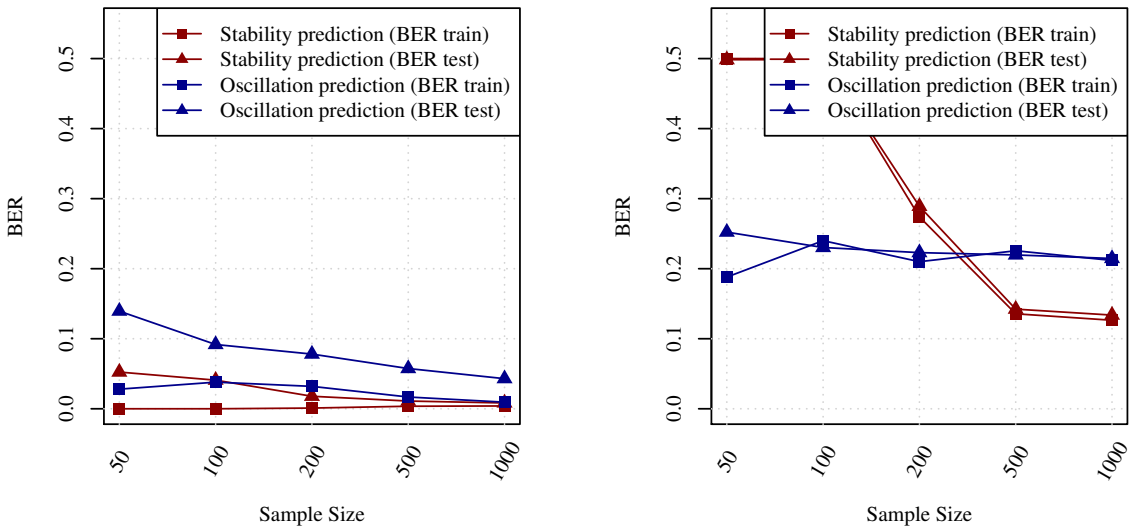
(a) Classification by elasticities. (b) Classification by FCCs.

Figure C.1.: Classification of linear pathway models



(a) Classification by elasticities. (b) Classification by FCCs.

Figure C.2.: Classification of branched pathway models



(a) Classification by elasticities.

(b) Classification by FCCs.

Figure C.3.: Classification of cyclic pathway models

D. Additional results (TCA cycle)

D.1. Additional results of decision tree classification

D.1.1. Balanced error rates for increasing sample size

The following tables show the balanced error rates (mean ± standard deviation) for selected training data sizes, determined by training five individual decision trees on independent training sets of each size.

	Elasticities		Flux control coefficients	
TrainingSize	Stability	Oscillations	Stability	Oscillations
2000	0.039 +- 0.006	0.124 +- 0.015	0.264 +- 0.018	0.307 +- 0.021
4000	0.033 +- 0.003	0.121 +- 0.006	0.246 +- 0.009	0.292 +- 0.01
8000	0.032 +- 0.005	0.12 +- 0.008	0.244 +- 0.007	0.292 +- 0.006
12000	0.029 +- 0.002	0.118 +- 0.008	0.246 +- 0.009	0.284 +- 0.012
16000	0.029 +- 0.002	0.11 +- 0.004	0.244 +- 0.006	0.289 +- 0.008
20000	0.028 +- 0.003	0.112 +- 0.006	0.242 +- 0.003	0.282 +- 0.004

(a) Classification errors determined on the training data.

	Elasticities		Flux control coefficients	
TrainingSize	Stability	Oscillations	Stability	Oscillations
2000	0.092 +- 0.005	0.221 +- 0.011	0.302 +- 0.012	0.363 +- 0.013
4000	0.081 +- 0.003	0.209 +- 0.003	0.28 +- 0.003	0.341 +- 0.008
8000	0.072 +- 0.002	0.199 +- 0.003	0.271 +- 0.003	0.327 +- 0.003
12000	0.065 +- 0.002	0.19 +- 0.003	0.269 +- 0.003	0.318 +- 0.005
16000	0.062 +- 0.001	0.186 +- 0.002	0.264 +- 0.003	0.316 +- 0.004
20000	0.058 +- 0.002	0.181 +- 0.002	0.261 +- 0.002	0.314 +- 0.001

(b) Classification errors determined on the test data.

Table D.1.: Balanced error rates of decision tree classification in the TCA cycle model.

D.1.2. Ruleset numbers and sizes

The following tables show the numbers and sizes of the decision tree rulesets obtained by classification on the reference state for selected training sizes. They were determined by training five individual decision trees on independent training sets of each size. All numbers are given in the format x/y, where x and y are the numbers of rulesets with Laplace values $\geqslant 0.95$, or $\geqslant 0.5$, respectively. Observed minimum and maximum ruleset sizes are given in brackets. Decision trees have been trained either using elasticities (left) or flux control coefficients (right) as features.

	Elasticities		Flux Control Coefficients	
TrainingSize	Stability	Instability	Stability	Instability
2000	10/71 (4, 6)	22/62 (2, 5)	0/23 (0, 0)	0/20 (0, 0)
4000	29/117 (3, 7)	27/87 (2, 6)	0/37 (0, 0)	0/35 (0, 0)
8000	59/188 (3, 9)	54/145 (2, 6)	0/40 (0, 0)	0/46 (0, 0)
12000	68/222 (3, 8)	114/250 (2, 9)	0/51 (0, 0)	0/33 (0, 0)
16000	105/277 (3, 9)	88/234 (2, 8)	0/64 (0, 0)	0/43 (0, 0)
20000	102/315 (3, 10)	120/314 (2, 10)	0/58 (0, 0)	0/57 (0, 0)

(a) Classification of stable versus unstable states.

	Elasticities		Flux Control Coefficients	
TrainingSize	No oscillations	Oscillations	No oscillations	Oscillations
2000	6/90 (2, 3)	0/43 (0, 0)	0/23 (0, 0)	0/11 (0, 0)
4000	11/171 (2, 5)	0/44 (0, 0)	0/45 (0, 0)	0/10 (0, 0)
8000	26/284 (2, 5)	0/74 (0, 0)	0/64 (0, 0)	0/6 (0, 0)
12000	63/405 (2, 8)	0/78 (0, 0)	0/77 (0, 0)	0/23 (0, 0)
16000	79/548 (2, 8)	0/99 (0, 0)	0/79 (0, 0)	0/17 (0, 0)
20000	83/610 (2, 10)	1/142 (8, 8)	0/116 (0, 0)	0/32 (0, 0)

(b) Classification of oscillating versus non-oscillating models.

Table D.2.: Numbers and observed sizes of the rulesets derived for reference state of the TCA cycle model.

Bibliography

Berndt N, Bulik S, Holzhütter HG (2012) Kinetic Modeling of the Mitochondrial Energy Metabolism of Neuronal Cells: The Impact of Reduced α-Ketoglutarate Dehydrogenase Activities on ATP Production and Generation of Reactive Oxygen Species. *International Journal of Cell Biology* **2012**: 1–11

Bisswanger H (2008) *Enzyme kinetics: Principles and methods*. Wiley-VCH, 2 edition

Boccard J, Veuthey JL, Rudaz S (2010) Knowledge discovery in metabolomics: An overview of MS data handling. *Journal of Separation Science* **33**: 290–304

Bornstein BJ, Keating SM, Jouraku A, Hucka M (2008) LibSBML: an API library for SBML. *Bioinformatics* **24**: 880–881

Brittain JS, Brown P (2014) Oscillations and the basal ganglia: Motor control and beyond. *NeuroImage* **85**: 637–647

Bruggeman FJ, Westerhoff HV (2007) The nature of systems biology. *Trends in Microbiology* **15**: 45–50

Bulik S, Grimbs S, Huthmacher C, Selbig J, Holzhütter HG (2009) Kinetic hybrid models composed of mechanistic and simplified enzymatic rate laws - a promising method for speeding up the kinetic modelling of complex metabolic networks. *FEBS Journal* **276**: 410–424

Cornish-Bowden A (2004) *Fundamentals of enzyme kinetics*. London: Portland Press, 3 edition

Cornish-Bowden A (2006) Putting the Systems Back into Systems Biology. *Perspectives in Biology and Medicine* **49**: 475–489

Da Silva L, Godejohann M, Martin FPJ, Collino S, Bürkle A, Moreno-Villanueva M, Bernhardt J, Toussaint O, Grubeck-Loebenstein B, Gonos ES, Sikora E,

Grune T, Breusing N, Franceschi C, Hervonen A, Spraul M, Moco S (2013) High-Resolution Quantitative Metabolome Analysis of Urine by Automated Flow Injection NMR. *Analytical Chemistry* **85**: 5801–5809

Darling DA (1957) The Kolmogorov-Smirnov, Cramer-von Mises Tests. *Annals of Mathematical Statistics* **28**: 823–838

Doyle J, Csete M (2005) Motifs, Control, and Stability. *PLoS Biology* **3**: e392

Dunn WB, Broadhurst DI, Atherton HJ, Goodacre R, Griffin JL (2010) Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chemical Society Reviews* **40**: 387–426

Fell DA (1997) *Understanding the control of metabolism*. London and Miami and Brookfield and VT: Portland Press and Distributed by Ashgate Pub. Co. in North America, 1 edition

Fell DA, Sauro HM (1985) Metabolic control and its analysis. Additional relationships between elasticities and control coefficients. *European Journal of Biochemistry FEBS* **148**: 555–561

Fell J, Axmacher N (2011) The role of phase synchronization in memory processes. *Nature Reviews Neuroscience* **12**: 105–118

Folger O, Jerby L, Frezza C, Gottlieb E, Ruppin E, Shlomi T (2011) Predicting selective drug targets in cancer through metabolic networks. *Molecular Systems Biology* **7**

Franceschi P, Giordan M, Wehrens R (2013) Multiple comparisons in mass-spectrometry-based -omics technologies. *TrAC Trends in Analytical Chemistry* **50**: 11–21

Girbig D, Grimbs S, Selbig J (2012a) Systematic Analysis of Stability Patterns in Plant Primary Metabolism. *PLoS ONE* **7**: e34686

Girbig D, Selbig J, Grimbs S (2012b) A MATLAB toolbox for structural kinetic modeling. *Bioinformatics* **28**: 2546–2547

Gopinath KGS, Pal S, Samui P, Sarkar BK (2013) Support Vector Machine and Relevance Vector Machine for Prediction of Alumina and Pore Volume Fraction in Bioceramics. *International Journal of Applied Ceramic Technology* **10**: E240–E246

Grimbs S, Selbig J, Bulik S, Holzhütter HG, Steuer R (2007) The stability and robustness of metabolic states: identifying stabilizing sites in metabolic networks. *Molecular Systems Biology* **3**: 146

Gross T, Feudel U (2006) Generalized models as a universal approach to the analysis of nonlinear dynamical systems. *Physical Review E Statistical Nonlinear and Soft Matter Physics* **73**: 016205

Hanslmayr S, Staudigl T (2014) How brain oscillations form memories - A processing based perspective on oscillatory subsequent memory effects. *NeuroImage* **85**: 648–655

Hastie T, Tibshirani R, Friedman JH (2009) *The elements of statistical learning: data mining, inference, and prediction*. Springer Series in Statistics

Heinrich R, Schuster S (1996) *The regulation of cellular systems*. New York: Chapman & Hall

Jamshidi N, Palsson BO (2008) Formulating genome-scale kinetic models in the post-genome era. *Molecular Systems Biology* **4**: 171

Kacser H, Porteous J (1987) Control of metabolism: What do we have to measure? *Trends in Biochemical Sciences* **12**: 5–14

Kholodenko B, Yaffe MB, Kolch W (2012) Computational Approaches for Analyzing Information Flow in Biological Networks. *Science Signaling* **5**: re1

Koopman WJH, Distelmaier F, Am Smeitink J, Willems PH (2012) OXPHOS mutations and neurodegeneration. *The EMBO Journal* **32**: 9–29

Kruse K (2010) *Towards the identification of (regulatory) motifs/modules in metabolic networks*. Master's thesis, Universität Potsdam, Potsdam

Li X, Wu F, Beard DA (2012) Identification of the kinetic mechanism of succinyl-CoA synthetase. *Bioscience Reports* **33**: 145–163

Maier T, Marcos J, Wodke JAH, Paetzold B, Liebeke M, Gutiérrez-Gallego R, Serrano L (2013) Large-scale metabolome analysis and quantitative integration with genomics and proteomics data in Mycoplasma pneumoniae. *Molecular BioSystems* **9**: 1743–1755

Makarov YV, Dong ZY (2001) Eigenvalues and Eigenfunctions, In *Wiley Encyclopedia of Electrical and Electronics Engineering*, Webster JG (ed), Hoboken and NJ and USA: John Wiley & Sons, Inc, ISBN 047134608X

MATLAB (2010) *version 7.11.0.584 (R2010b)*. Natick, Massachusetts: The MathWorks Inc.

Nägele T, Weckwerth W (2012) Mathematical Modeling of Plant Metabolism-From Reconstruction to Prediction. *Metabolites* **2**: 553–566

Nelson DL, Cox MM (2004) *Lehninger Principles of Biochemistry:4th ed*. New York: W.H.FREEMAN & CO LTD

Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? *Nature Biotechnology* **28**: 245–248

Pfau T, Christian N, Ebenhöh O (2011) Systems approaches to modelling pathways and networks. *Briefings in Functional Genomics and Proteomics*

Quinlan JR (1998) *C4.5: Programs for Machine Learning*. San Mateo and Calif: Kaufmann

Quinlan JR (2013) Data Mining Tools See5 and C5.0. Url: http://www.rulequest.com/see5-info.html (Last accessed: Aug-22-2013)

R Core Team (2012) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0

Reznik E, Segrè D (2010) On the stability of metabolic cycles. *Journal of Theoretical Biology* **266**: 536–549

Rodríguez A, Infante D (2009) Network models in the study of metabolism. *Electronic Journal of Biotechnology* **12**: 1–19

Rokach L, Maimon O (2005) Decision Trees, In *Data Mining and Knowledge Discovery Handbook*, Maimon O, Rokach L (eds), New York: Springer-Verlag, ISBN 0-387-24435-2, pp. 165–192

Rokach L, Maimon O (2008) *Data mining with decision trees: Theory and applications*. volume 69 of *Series in machine perception and artificial intelligence*. Singapore: World Scientific

Schmidt H (2013) Systems Biology Toolbox 2 (SBTOOLBOX2) for MATLAB. Url: http://www.sbtoolbox2.org (Last accessed: Jun-11-2013)

Schmidt H, Jirstrand M (2006) Systems Biology Toolbox for MATLAB: a computational platform for research in systems biology. *Bioinformatics* **22**: 514–515

Schölkopf B, Burges CJ, Smola AJ (1999) *Advances in kernel methods: support vector learning*. The MIT press

Sellami HK, Napolitano A, Masullo M, Smiti S, Piacente S, Pizza C (2013) Influence of growing conditions on metabolite profile of Ammi visnaga umbels with special reference to bioactive furanochromones and pyranocoumarins. *Phytochemistry* **95**: 197–206

Singer W (2013) Cortical dynamics revisited. *Trends in Cognitive Sciences* **17**: 616–626

Steuer R (2007) Computational approaches to the topology, stability and dynamics of metabolic networks. *Phytochemistry* **68**: 2139–2151

Steuer R, Gross T, Selbig J, Blasius B (2006) Structural kinetic modeling of metabolic networks. *Proceedings of the National Academy of Sciences* **103**: 11868–11873

Steuer R, Nesi AN, Fernie AR, Gross T, Blasius B, Selbig J (2007) From structure to dynamics of metabolic pathways: application to the plant mitochondrial TCA cycle. *Bioinformatics* **23**: 1378–1385

Stitt M, Sulpice R, Keurentjes JJB (2010) Metabolic Networks: How to Identify Key Components in the Regulation of Metabolism and Growth. *Plant Physiology* **152**: 428–444

Thiele I, Palsson BØ (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols* **5**: 93–121

Tipping ME (2001) Sparse Bayesian Learning and the Relevance Vector Machine. *Journal of Machine Learning Research* **1**: 211–244

Tipping ME (2013) SPARSEBAYES Matlab Toolbox. Url: http://www.vectoranomaly.com (Last accessed: Dec-20-2013)

Tipping ME, Faul AC, *et al* (2003) Fast marginal likelihood maximisation for sparse Bayesian models. In *Proceedings of the ninth international workshop on artificial intelligence and statistics*, volume 1

van Nes P, Bellomo D, Reinders M, Ridder Dd (2009) Stability from Structure: Metabolic Networks Are Unlike Other Biological Networks. *EURASIP Journal on Bioinformatics and Systems Biology* **2009**: 1–15

Wang L, Birol I, Hatzimanikatis V (2004) Metabolic Control Analysis under Uncertainty: Framework Development and Case Studies. *Biophysical Journal* **87**: 3750–3763

Weiss RJ, Ellis DPW (2006) Estimating single-channel source separation masks: Relevance vector machine classifiers vs. pitch-based masking. In *ISCA Tutorial and Research Workshop on Statistical and Perceptual Audition*

Wessely F, Bartl M, Guthke R, Li P, Schuster S, Kaleta C (2011) Optimal regulatory strategies for metabolic pathways in Escherichia coli depending on protein costs. *Molecular Systems Biology* **7**

Westerhoff HV, Kell DB (1987) Matrix Method for Determining Steps Most Rate-Limiting to Metabolic Fluxes in Biotechnological Processes. *Biotechnology and Bioengineering* **30**: 101–107

Wolfram Research I (2012) *Mathematica, Version 9.0*. Champaign, Illinois: Wolfram Research, Inc.

Xia J, Mandal R, Sinelnikov IV, Broadhurst D, Wishart DS (2012) MetaboAnalyst 2.0–a comprehensive server for metabolomic data analysis. *Nucleic Acids Research* **40**: W127–W133

Yang D, Liang G, Jenkins D, Peterson GD, Li H (2010) High performance relevance vector machine on GPUs. In *Symposium on Application Accelerators in High Performance Computing*

Yuan JS, Galbraith DW, Dai SY, Griffin P, Stewart CN (2008) Plant systems biology comes of age. *Trends in Plant Science* **13**: 165–171

Zhang A, Sun H, Wang P, Han Y, Wang X (2011) Modern analytical techniques in metabolomics analysis. *The Analyst* **137**: 293–300