Parameter Balancing in Kinetic Models

Timo Lubitz

Master Thesis

Freie Universität Berlin
Department of Mathematics and Computer Science

Advisors:
Prof. Dr. Alexander Bockmayr
Dr. Wolfram Liebermeister

Freie Universität Berlin
Habelschwerdter Allee 45
14195 Berlin
Germany

Email: lubitz@fu-berlin.de

November 01st 2009 – May 01st 2010

Abstract

Kinetic modeling is one of the major methods in systems biology. Every modeling effort is highly dependent on (i) information about the underlying metabolic pathways and (ii) the kinetic data derived from experimentalists. The kinetic data for a model, available from ever-growing web resources, is often incomplete, hard to accumulate, or simply not available at all. Incomplete kinetic data is the scope of this thesis; a problem that every kinetic modeler is faced with.
The introduced approach to cope with incomplete kinetic data is a parameter estimation within a Bayesian framwork. Its foundations are the dependencies among different types of kinetic parameters that allow the estimation of missing kinetic parameters by being provided with other, available kinetic parameters. Furthermore, the approach is extended by taking inhomogenities in the measuring circumstances (pH value and temperature) of the input data into account.
The parameter estimation is tested on one small model to check for the reliability of the results. Moreover, I have examined the results of my approach applied on larger scale models and created several test scenarios to see the impact of input data variations (e.g. different pH and temperature values, and the limitation of the input data to several parameter types).
The comparisons of the original model data to the results of my parameter estimation approach show the reliability of the latter. Nevertheless, a few problems occurred during the testing phase. I was able to either solve these problems, or at least identify and name them in order to prevent their occurence. The application of the introduced approach will soon be available for users as an extension of the Systems Biology tool semanticSBML.

Zusammenfassung

Kinetisches Modellieren ist einer der Hauptanwendungsbereiche der modernen Systembiologie. Jeder Modellierungsansatz hängt zum einen von ausführlichem Wissen über das zu modellierende metabolische Netzwerk und zum anderen von den kinetischen Daten ab, die von Experimentalisten zur Verfügung gestellt werden. Kinetische Daten, die aus Webressourcen bezogen werden können, sind jedoch oft unkomplett, schwer zu sammeln oder auch einfach nicht verfügbar. Der Themenbereich dieser Masterarbeit bezieht sich auf unkomplette kinetische Datensätze; ein Problem, mit dem sich jeder kinetische Modellierer konfrontiert sieht. Der vorgestellte Lösungsansatz ist eine Parameterschätzung innerhalb eines Bayesian Framework. Die Grundlage der Schätzung sind die Abhängigkeiten der verschiedenen Parametertypen untereinander, die es erlauben, auf der Grundlage von erhältlichen kinetischen Parametern andere, fehlende kinetische Parameter zu schätzen. Dieser Ansatz wird noch dadurch erweitert, dass Inhomogenitäten der Messumstände (pH-Wert und Temperatur) der Eingabedaten mit in Betracht gezogen werden.

Die Parameterschätzung wird an einem kleinen Modell getestet, um die Verlässlichkeit der Ergebnisse untersuchen zu können. Darüber hinaus wird die Schätzung auf großskalige Modelle angewandt, und es werden verschiedene Testszenarien, welche den Einfluss von Variationen der Eingabedaten (zum Beispiel verschiedene pH- und Temperatur-Werte oder die Einschränkung der Eingabedaten auf wenige Parametertypen) simulieren, verwirklicht.
Der Vergleich der erhältlichen, originalen Modelldaten mit den Resultaten meiner Parameterschätzung zeigen die Verlässlichkeit letzterer. Nichtsdestotrotz traten während der Testphase auch Probleme auf. Diese Probleme konnte ich entweder lösen, oder doch zumindest identifizieren und benennen, um ihr Auftreten verhindern zu können. Die Anwendung des vorgestellten Parameterschätzverfahrens wird als Erweiterung der bestehenden systembiologischen Software semanticSBML bald öffentlich erhältlich sein.

Acknowledgements

I want to thank the whole AG Theoretische Biophysik of Edda Klipp for their contributions to this thesis. Next to my first advisor Alexander Bockmayr, my honest gratitude belongs to Wolfram Liebermeister, who never got tired of my numerous questions (not even on holidays), and of course Marvin Schulz, Jannis Uhlendorf, Max Flöttmann, and Marcel Grunert for proofreading and their contributions to my work. Finally, this whole thesis is dedicated to Rainer Lubitz († 17. April 2009), who would surely have smiled over being mentioned in this work he would not have understood a word of. But without him, all these pages would be empty.

Contents

Abbreviations

Abbreviation	Meaning or Context
SRES	**S**tochastic **R**anking **E**volution **S**trategy
SBML	**S**ystems **B**iology **M**arkup **L**anguage
TSV	**T**ab **S**eparated **V**alues (file format)
XML	E**x**tensible **M**arkup **L**anguage (file format)
MIRIAM	**M**inimal **I**nformation **R**equested **I**n the **A**nnotation of biochemical **M**odels
SBO	**S**ystems **B**iology **O**ntology
CM	**C**ommon **M**odular Rate Law
DM	**D**irect Binding **M**odular Rate Law
SM	**S**imultaneous Binding **M**odular Rate Law
PM	**P**ower-Law **M**odular Rate Law
FM	**F**orce-Dependent **M**odular Rate Law
GUI	**G**raphical **U**ser **I**nterface
R	Boltzmanns gas constant ($R \approx 8.314\text{J}/(\text{molK})$)
T	Absolute temperature
PFK	Phosphofructokinase
F6P	Fructose-6-phosphate
F16P	Fructose-1,6-bisphosphate
Std	Standard deviation

List of Tables

List of Figures

1. Introduction

1.1. Kinetic Modeling in Systems Biology

Systems Biology Systems Biology is the study of organisms via integrated and interacting networks. These networks can include genes, proteins, and biochemical reactions. Systems biologists are focussed on all of the organisms components and the interactions among them as part of one system. Living systems cannot be understood completely by studying just individual parts, sophisticated computational approaches are needed for this task. The advances in information technology in combination with inexpensive computing power and comprehensive databases have significantly increased the importance and feasibility of mathematical modeling, and the simulation of complex biochemical processes (Kriete & Eils, 2006).

Kinetic Modeling The broad field of kinetic modeling in Systems Biology comprises the examination of metabolic networks by the use of mathematical models. These models represent the structure and the dynamics of a model in order to predict their behaviour under different conditions. Kinetic data from experiments are needed to make numeric simulations possible.

1.2. Kinetic Data

Problem: The Incompleteness of Kinetic Data The connection between iterative experimental testing and mathematical modeling of the interactions of cellular components is one definition of Systems Biology (Kitano, 2002). The modeling cycle comprises the first step of developing a model using experimental data, the usage of the model for a prediction of its behaviour, and finally the validation of this prediction. The prediction, either valid or not, is used for alterations on the model in order to gain improved predictions.
There exists a strong dependence of the modeler on the experimentalists. Since the computational modeler wants to produce models as close to nature as possible, a vast number of experimental results is needed. Despite an ever-growing number of biochemical data, the modeler will always be facing the problem of incomplete knowledge about the corresponding biochemical networks. On the one hand, this is due to the fact that metabolism is a broad and partially unexplored field, on the other hand, it is not easy to collect all the available experimental knowledge. Finally, of course not every measurable value has already been measured.

Why are Kinetic Data Important? The analysis of any metabolic network in combination with proven predictions can be considered one more step in solving the uncer-

tainties of biology. With growing knowledge about biochemical reactions and pathways, scientists becomes more capable of understanding biological processes. While this conclusion is very globally spoken and can have a significant impact on human life (e.g. by pharmaceutical sciences), it does originate in the small things. The bottom-up approach proposes the understanding of biology by investigations on the pathways and networks. The knowledge about pathways and networks is derived from our knowledge about the underlying reactions. Reactions themselves - as well as the involved species - are characterized via their kinetic parameters: equilibrium constants, turnover rates, Michaelis constants, velocity constants, and many more. The more we know about these parameters, the better we can step by step build up our knowledge about everything they and their corresponding reactions stand for.
The analysis of networks and efforts to cope with incomplete data are widely spread and known. Among others, the tool anNET (Zamboni, Kuemmel & Heinemann, 2008) is an example of such investigations and has been established as good working basis for a network-embedded thermodynamic analysis. AnNET uses a constraint-based approach for the analysis of fluxes and parameter estimation. Similar to this approach is the constraint-based flux-balance analysis of Holzhütter et al. (Hoppe, Hoffmann & Holzhuetter, 2007), both of which underlie several difficulities, since they rely on hardly provable optimality principles. The parameter estimation tool SBML-PET (Zi & Klipp, 2006) supports the import and export of model data in the format SBML (see Chapter 2.1), while the parameter estimation itself is performed via a stochastic ranking evolution strategy (SRES). A wide range of accessories is provided by Potters Wheel (Raue, Kreutz, Maiwald, Bachmann, Schilling, Klingmuller & Timmer, 2009) that work on the profile likelihood approach. It can detect "structural and practical non-identifiabilities" on the basis of functionally related model parameters. The output is given in form of confidence intervals corresponding to the likelihood.

An Example Network Having a look at an example network (see Figure 1.1) makes it possible to explain the data we are dealing with when analyzing biochemical networks via mathematical modeling. We can see three species involved in two reactions, which is considered to be the biochemical network, or, in kinetic modeling terms, the model. The second type of information we need are the kinetic parameters for the model (formation energies, rate constants, or concentrations of the involved species).

The obvious problem of this model and the given data is incompleteness of the kinetic parameters. Although large numbers of these parameters are stored in ever-growing web resources (e.g. Brenda (Barthelmes, Ebeling, Chang, Schomburg & Schomburg, 2007) and Sabio-RK (Wittig, Golebiewski, Kania, Krebs, Mir, Weidemann, Anstein, Saric & Rojas, 2006)), we cannot assume that we are given an equilibrium constant for every reaction of the model, a turnover rate for every species, or every single Michaelis constant we need. Furthermore, it is improbable that the given values from the literature or databases are comparable at all, since they might have been measured under completely different circumstances (pH, temperature, etc.), or might not at all be appropriate for a specific model (Chassagnole, Rais, Quentin, Fell & Mazat, 2001).

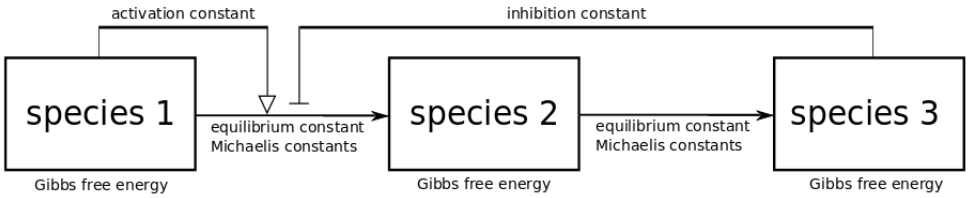

Figure 1.1.: An example network: The model itself (species and reactions) can be provided in SBML, the kinetic parameters $G^{(0)}$ (Gibbs free energy of formation), k^M (Michaelis constant), k^{eq} (equilibrium constant), k^I (inhibitory constant), and k^A (activation constant) can be taken from SBtab kinetic parameter files (see Chapter 2.3). These can also include more parameter types than shown in the figure (e.g. turnover rates, velocity constants, enzyme concentrations, or reaction affinities).

1.3. Data Representation

How are the Data Represented? The parameter estimation approach introduced in this thesis is using two different kinds of data: the biochemical network model and the corresponding kinetic parameter sets. While biochemical models can be provided in the markup language SBML, kinetic parameter sets for the models (including kinetic data like Gibbs free energies, turnover rates, inhibition constants, and many more) are often available in the form of spreadsheets of differing syntaxes. An effort to establish a generalized form for this data is the SBtab format (which is currently under development, see Chapter 2.3); the kinetic parameter tables used for my approach are provided in this format.
The Systems Biology tool semanticSBML (Krause, Uhlendorf, Lubitz, Schulz, Klipp & Liebermeister, 2010) will be the interface for the automatization of parameter balancing. SemanticSBML comprises possibilities to annotate, modifiy, and merge SBML models. It will be extended to integrate quantititative data from SBtab files into corresponding SBML models with the possibility of automated parameter balancing. Another Systems Biology tool, SBMLfill (Liebermeister, Uhlendorf & Klipp, 2010), realizes the integration of the data into the model.

1.4. Parameter Balancing

What is my Approach to Solve the Problem? Assuming an SBML model and a corresponding incomplete kinetic data set in SBtab format are given, the goal is to obtain as much information on the whole network as possible, based on

1. the model,

2. the kinetic parameters,

3. and information taken from literature or web resources.

My approach to retrieve this information is a parameter estimation within a Bayesian statistical framework (Liebermeister & Klipp, 2006). The thermodynamic and kinetic data is integrated, and a distribution of parameter values describing thermodynamically feasible model parameters is obtained. I receive an incomplete parameter set for a model and by taking this set and any available data from literature and web resources into account, the missing parameters can be estimated.
The key to these estimations are the constraints between certain kinetic parameters: biochemical reactions and their corresponding kinetic parameters need to be seen as a whole in their underlying network. If information on several parameters in the network is available, further information on the missing parameters can be derived from this knowledge. There are many laws and relations among the parameter types, which will be called dependencies. Examples for these dependencies are the Haldane relationship (relating the equilibrium constants of the reactions with the turnover rates and the Michaelis constants of the metabolites) or another relationship linking the Gibbs free energies of formation with the equilibrium constants. Based on the dependency derived from this relationship, we can calculate the equilibrium constant of a reaction with only a small effort, as soon as the Gibbs free energies of the reactions metabolites are available.
When taking these and even more relations among the parameter types into account, a set of connected parameters can be obtained that offers the possibility to be estimated by one another, if certain parameters are missing. As further knowledge, a prior distribution of values generated from web resources will be needed. Having the model, a set of kinetic parameters and the prior values, I am able to perform parameter estimates within a Bayesian framework. The underlying workflow considering the input/output data is visualized in the flowchart of Figure 1.2.

In addition to the former works (Liebermeister & Klipp, 2006; Borger, 2008) I add the possibility of a temperature and pH value regression that allows the user to take the measuring circumstances of the input data into account for the balancing process. It is important to note that the measuring circumstances can have significant influence on the magnitude of several parameter type values. The Gibbs free energies as well as the equilibrium constants are directly dependent on the temperature and pH value they have been measured in. Furthermore, a consideration of reaction affinities and Gibbs free energies will be a part of my work. A graphical user interface will be guiding the user through the process of parameter balancing and the integration of the results into the model file.

What do I Expect to Achieve by my Approach? The extensive testing of this parameter balancing approach on several SBML models with corresponding kinetic data sets will show, whether my approach can generate reliable data that is a contribution to the scientific community. If the results are turning out to be appropriate, an approach to conquer the problems of incomplete kinetic data sets for biochemical models by statistical estimations can be offered.

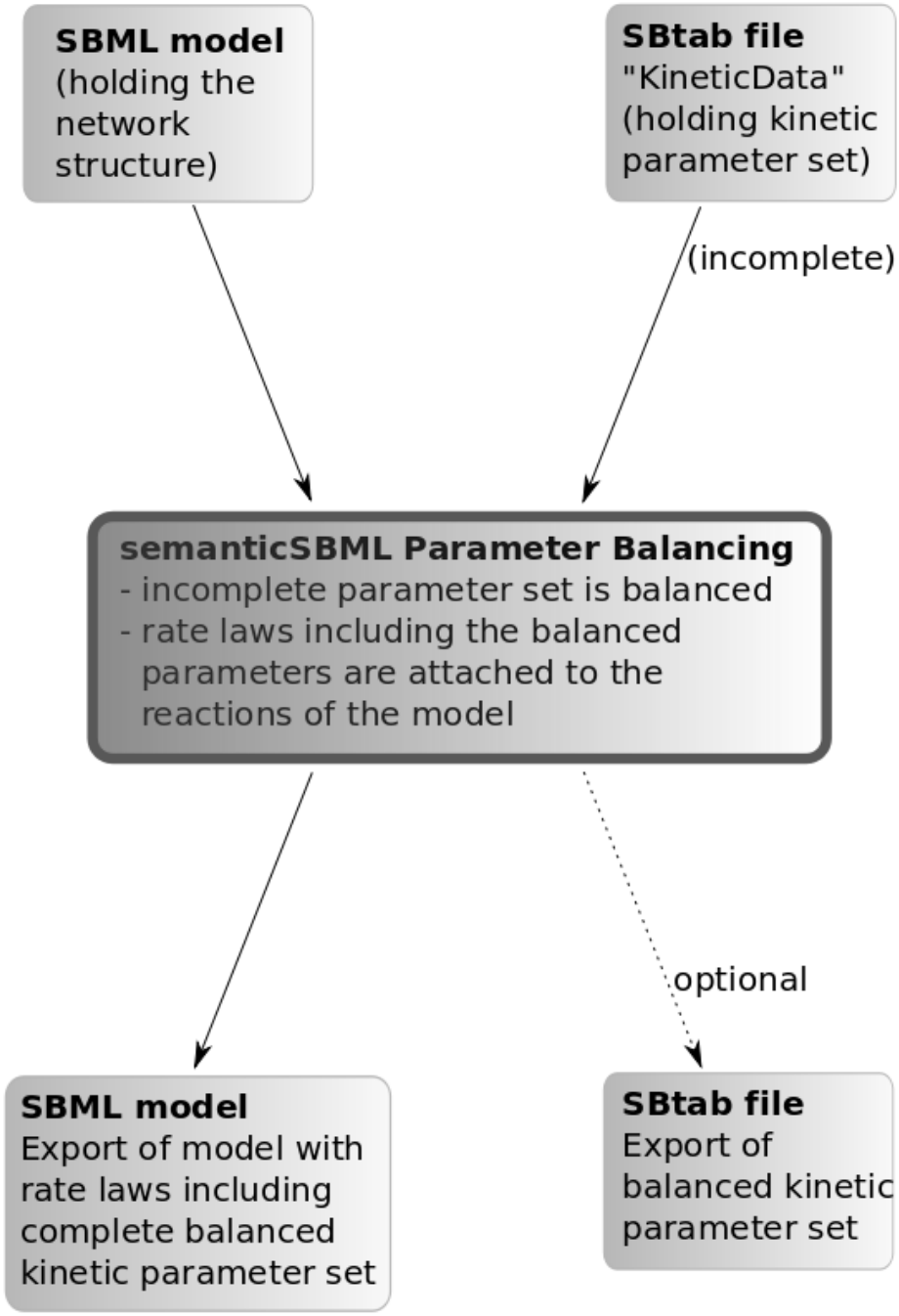

Figure 1.2.: Chart of the workflow: semanticSBML is taking an SBML model file and an SBtab file with an incomplete corresponding kinetic parameter set as input. It balances the kinetic parameters in order to complete them. The user can choose a rate law to be attached to the reactions of the model. The balancing results will be the kinetic parameters of these attached rate laws. Optionally, the user can export the balanced kinetic parameter set as an SBtab file.

2. Materials and Methods

The materials and methods section is mainly focussing on the formats of the input data (SBML and SBtab), the tools that are used for the manipulation of the data, and the parameter balancing process in detail.

2.1. SBML - a Format for the Mathematical Representation of Biochemical Models

The Systems Biology Markup Language (SBML) (Hucka, Finney, Sauro, Bolouri, Doyle, Kitano & others, 2003) is an XML based, computer-readable format for representing computational models of biological systems. It has been developed since 2000 for all kinds of Systems Biology software tools. In the growing amount of available file formats, SBML is intended to be a generalization for the mathematical representation of biochemical models. Meanwhile, it is accepted by large parts of the Systems Biology community and due to its open accessibility it is constantly improved by software developers and users.
SBML is able to store many kinds of information needed for dynamic modeling and for performing simulations on the models. It works on a list based structure, holding lists of reactions, species, compartments, parameters, etc. The list elements can be of many different types, some of the most important ones are

Compartment holds the information, where a species is located.

Reaction describes how certain entities (reactants) are transformed into other entities (products). They have assigned kinetic rate expressions describing their speed.

Species are the single metabolites and other molecules in a certain compartment.

Parameters are numerical values defined globally or locally for a certain reaction.

The SBML representation of a reaction contains several lists of elements, which are mostly optional. First, it holds lists of reactants and products, and second, a list of modifiers. The underlying kinetic law for the reaction can be found as well as a list of parameters for the law. The corresponding SBML code to such a reaction is shown in Table 2.1.

By using SBML as standard format, the usability of a software is significantly increased, since more and more software developers and users revert to it. There exists a growing number of available SBML models and the agreement of using generalized standard formats like SBML is improving the collaboration of scientists everywhere.

```
<?xml version="1.0" encoding="UTF-8"?>
  <sbml xmlns="http://www.sbml.org/sbml/level2/version3"...>
    <model name="EnzymaticReaction">
    ...
      <listOfReactions>
        <reaction id="veq">
          <listOfReactants>
            <speciesReference species="ES"/>
          </listOfReactants>
          <listOfProducts>
            <speciesReference species="E"/>
            <speciesReference species="S"/>
          </listOfProducts>
          <kineticLaw>
            <math xmlns="http://www.w3.org/1998/Math/MathML">
              <apply>
                <times/>
                <ci>cytosol</ci>
                <ci>kcat</ci>
                <ci>ES</ci>
              </apply>
            </math>
            <listOfParameters>
              <parameter id="kcat" value="0.1" units="p/s"/>
            </listOfParameters>
          </kineticLaw>
        </reaction>
      </listOfReactions>
    </model>
  </sbml>
```

Table 2.1.: The XML based SBML code (Level 2, Version 3) for a reaction holds the obligatory lists of reactants and products next to informations on the kinetic law and a list of parameters.

2.1.1. MIRIAM Annotations

Annotations are a way to describe the elements of a model by unique identifiers. These identifiers are linked to vast biochemical web resources, such as KEGG (Kanehisa & Goto, 2000) or ChEBI (Degtyarenko, de Matos, Ennis, Hastings, Zbinden, McNaught, Alcantara, Darsow, Guedj & Ashburner, 2008). A lot of information can be gained for every annotated element via the database identifier (e.g. the name, structure, or molecular weight). They are usually taken from the SBML source code. MIRIAM ("**M**inimal **I**nformation **R**equested **I**n the **A**nnotation of **B**iochemical **M**odels") is a concept that defines what appears in the annotation and how it is stored (Le Novère,

Finney, Hucka, Bhalla, Campagne, Collado-Vides, Crampin, Halstead, Klipp, Mendes & others, 2005). One model element can have several MIRIAM Annotations from different web resources.

2.1.2. SBO Terms

The Systems Biology Ontology (Laibe & Le Novère, 2007) is a set of controlled vocabularies and ontologies tailored specifically for the kinds of problems that Systems Biology is confronted with, especially in the context of computational modelling. Just like MIRIAM annotations, SBO terms can be found in the SBML source code, and they characterize model elements in either a very general (e.g. as "entity") or in a more specific way (e.g. as "receptor"). Just like by using MIRIAM annotations, a user can link detailed information to a model element by assigning it a specific SBO identifier. In contrast to MIRIAM, SBO terms focus rather on *what* is described than *how* it is described.

2.1.3. libSBML

libSBML (Bornstein, Keating, Jouraku & Hucka, 2008) is an open-source library for the manipulation of SBML models. It can be imported into an application and be used for manipulation and validation of models written in SBML. It is written in C and C++, but provides language bindings for Python and several other programming languages.

2.2. SBMLfill - a Tool for the Integration of Standardized Kinetic Rate Laws Into SBML Models

SBMLfill (Liebermeister, Uhlendorf & Klipp, 2010) is an SBML based web application. The tool (see Figure 2.1 for the graphical user interface) allows the user to attach standardized kinetic rate laws to the reactions of an SBML model. The model can be uploaded together with a corresponding parameter table file in the SBtab format (see Chapter 2.3). If such an SBtab file is available and valid, its content can be inserted into the rate laws that will be attached to the reactions of the model. All the rate laws are based on the same parameter types. In case the user does not have a parameter table, it is possible to attach standardized kinetic rate laws to the reactions of the model with all the parameters set to a default value of 1.

When inserting the rate law into the SBML model the user has to make several choices. The default type of enzyme activation can be set to either complete, partial, or specific, the default type of enzyme inhibition can be set to the same options. Furthermore, it is feasible to comprise enzymes in rate constants and to overwrite the existing rate laws of the SBML model. Since the kinetic constants in the model need to satisfy the Wegscheider conditions as well as the Haldane relationships, it would be of advantage to apply a thermodynamically safe parametrisation using independent parameter sets (Liebermeister & Klipp, 2006). The choice of the rate law itself is the most crucial of the users choices. Provided by the tool are five types of rate laws, each applicable under different conditions (see Table 2.2). All of the rate laws are depending on the same

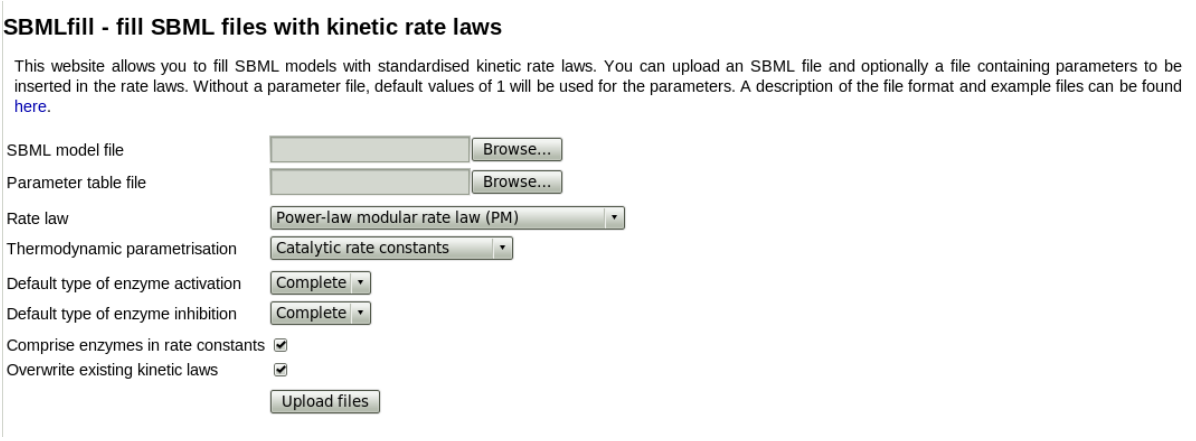

Figure 2.1.: Graphical user interface of SBMLfill: A user can upload an SBML model and a corresponding SBtab parameter file. A rate law can be chosen as well as the integration criteria. The content of the parameter file will then be inserted into the model.

parameter types (shown in Table 2.7). Inserting the kinetic rate laws into the SBML file demands the availability of these parameter types including a designated parameter value. The mathematical background concerning these rate laws can be found in the corresponding publication (Liebermeister, Uhlendorf & Klipp, 2010).

Common Modular Rate Law (CM) Generalised form of the Michaelis-Menten kinetics. It is based on a random-order enzyme mechanism and resembles the convenience kinetics (Liebermeister & Klipp, 2006) with only a slight difference for molecularities.

Direct Binding Modular Rate Law (DM) Is comparable to the CM-rate law, but contains just the substrate and product terms of the highest order. By doing so, the reaction rate is generally higher than that of the CM-rate law at the same parameter values (especially at low concentrations).

Simultaneous Binding Modular Rate Law (SM) The reaction rates are lower, especially at high concentrations.

Power-Law Modular Rate Law (PM) A mass action kinetic.

Force-Dependent Modular Rate Law (FM) serves for thermodynamic studies.

Table 2.2.: The five rate laws of SBMLfill that are offered to be attached to the reactions of a given SBML model.

2.3. SBtab - an Effort to Standardize Table Formats

SBtab is a standardized table format for Systems Biology data, which is currently under development (Liebermeister & Klipp,). Each type of SBtab, stored in the file format .tsv (tab separated values), is characterized by a set of conventions and holds different types of information. A list of possible SBtab data types is given in Table 2.3.

- Reactions
- Enzymes
- Compounds (Species)
- Genes
- Chemical reactions
- Biochemical quantities (like parameters and corresponding values)

Table 2.3.: Different types of SBtab files: the format can hold tabular information on reactions, enzymes, compounds, genes, reactions, biochemical quantities, and more.

Each of these SBtab tables has certain obligatory columns and, furthermore, several optional columns. As an example, the "Enzyme SBtab" needs to provide the columns "Enzyme", "CatalyzedReaction", and "Gene". The optional columns are "CatalyzedReactionID", "KineticLawID", "KineticLawName", "EnzymeRegulation", "GeneID", and "GeneCombination". All the columns ending on "ID" are referring to database identifiers, like for instance those of the KEGG or ChEBI web resources. An SBtab for "Enzyme", stored in .tsv-format and fulfilling the upper criteria, is shown in Table 2.4.

Enzyme	**CatalyzedReaction**	**Gene**	**GeneId**
spermine oxidase	oxygen oxidoreductase	HSA:54498(SMOX)	EC:1.5.3.16
ATP citrate synthase	citrate cycle (TCA cycle)	HSA:47(ACLY)	EC:2.3.3.8
hexokinase	Glycolysis	HSA:3098(HK1)	EC:2.7.1.1

Table 2.4.: Example for an "Enzyme" SBtab file: The required columns (Enzyme, CatalysedReaction, and Gene) are provided, moreover the optional column GeneID.

As the "Enzyme SBtab" stands as a simple example, this thesis is mainly focussing on a specific type of SBtab, the "KineticData" SBtab. This SBtab type provides a table of different kinetic parameters referring to biochemical reaction kinetics. The obligatory columns of this SBtab are "QuantityType" (the parameter type), "Reaction", "Compound", and "Value"; a "KineticData" SBtab is shown in Table 2.5.

QuantityType	Reaction	Compound	Value	Unit
inhibitory constant	vATP	P	1	mM
concentration	-	Glyc	0.02	mM
concentration	-	ADP	0.1	mM
equilibrium constant	vGly	ATP	0.1	mM
Michaelis constant	vATP	P	1	-

Table 2.5.: Example for a "KineticData" SBtab file: This file holds information on kinetic data for a corresponding model. Next to the required columns (QuantityType, Reaction, Compound, and Value), this SBtab also holds an optional column "Unit" for the unit of the content in the "Value"-column.

The content of a "KineticData" SBtab is the kinetic parameter set of a biochemical model. Inserting this kinetic data into an existing SBML model is a task that can be performed using SBMLfill (see Chapter 2.2) (provided the insertion criteria are matched). Similar to SBML, SBtab is a generalized standard format. But other than SBML, the SBtab format is not yet introduced and published for the scientific community.

2.4. semanticSBML

SemanticSBML (Krause, Uhlendorf, Lubitz, Schulz, Klipp & Liebermeister, 2010) is a Systems Biology tool for the annotation, modification, and merging of SBML models. While the tool is already freely available on the internet (www.semanticsbml.org) we are constantly improving and extending it.
The tool was written in Python and the graphical user interface (GUI) developed under the Python GUI extension PyQt (see Figure 2.2).

After having opened an SBML model, the user can perform the following actions on it:

Annotate A new tab will be opened that allows the user to add, remove, or modify the annotations of the model. SBO terms can be added or removed, and further information on the model elements can be obtained.

View The view option opens a new window showing the selected SBML model in a graph view (generated with GraphViz (Ellson, Gansner, Koutsofios, North & Woodhull, 2001)).

Check Several semantic checks on the validity of the model are performed, and in case of invalidity errors are returned:

1. Annotation errors: annotations are missing or cannot be interpreted.
2. Duplicate elements errors: model elements are annotated with identical database identifiers.
3. Overlapping compartments errors: compartments are physically overlapping.

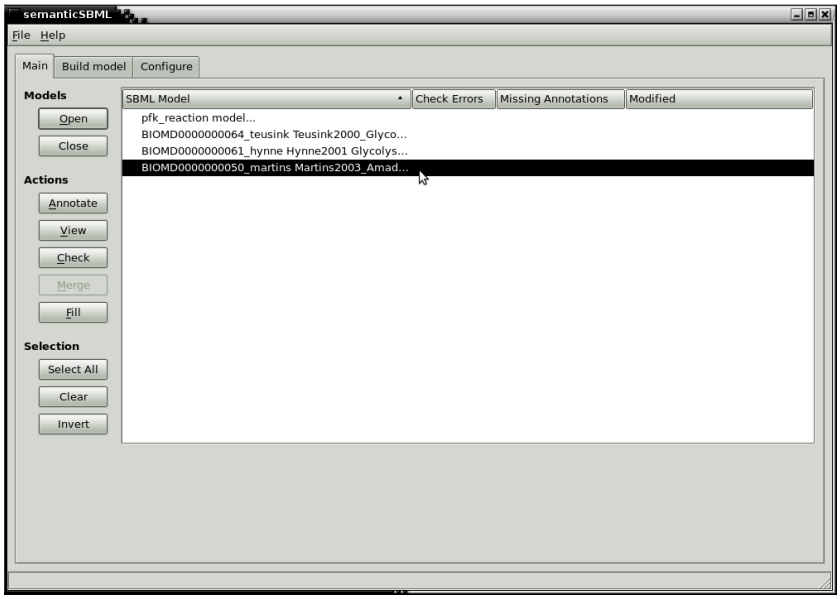

Figure 2.2.: The starting screen of semanticSBML: The user can open/close SBML models and perform several actions on them.

4. Atom number balance errors: the atom numbers in a chemical reaction are not conserved.

Merge Enables the user to merge two or more SBML models. Duplicate elements are removed, semantic errors can be solved automatically or manually. The difficulty of this operation is the recognition of similarity in model elements. The elements can be annotated with identifiers from different databases, yet still be the same in a biological context. Finding these similarities requires an internal database including the identifiers from a list of several web resources. This internal database is provided by semanticSBML and is build up at the first start.

2.5. Internal Database MetNetDB

An important part of estimation within Bayesian frameworks is the availability of a prior distribution of values. Thus, it is crucial to have access to large sets of kinetic parameters. This task is performed using our internal database for kinetic parameters, MetNetDB (Borger, 2008) (not to be confused with the Metabolic Network Exchange Database, MetNetDB, from the National Science Foundation Arabidopsis). It holds data that is collected from many different sources. Next to a large amount of published kinetic data taken from single publications, MetNetDB offers the access to data from the online databases KMedDB (http://sysbio.molgen.mpg.de/KMedDB) and Brenda (Schomburg, Chang, Hofmann, Ebeling, Ehrentreich & Schomburg, 2002). Overall the MetNetDB contains slightly over 100000 parameter values that are divided into the parameter types shown in Table 2.6.

It is an advantage to be provided with such a huge kinetic database, since the distribu-

Name	Abbreviation	Amount
Gibbs free energies of formation	$\mathbf{G^{(0)}}$	10629
Michaelis constants	$\mathbf{k^M}$	62740
Inhibitory constants	$\mathbf{k^I}$	12827
Species concentrations	$\mathbf{c}$	755
Equilibrium constants	$\mathbf{k^{eq}}$	2088
Turnover rates (forward/backward)	$\mathbf{k^{cat}}$	12083

Table 2.6.: Table of kinetic parameters taken from MetNetDB.

tion of prior values is crucial for the estimation. Nevertheless, the MetNetDB contains no values for activation constants, which is a small drawback for the used approach.

2.6. Parameter Estimation

Parameter estimation is a basic part of kinetic modelling. In this work I am implementing, extending, and applying an estimation approach within a Bayesian framework (Liebermeister & Klipp, 2006). Some of the parameter values can be measured in experiments and integrated into the mathematical calculations in the "SBtab KineticData"-format (see Chapter 2.3) or likewise. Having obtained a set of parameters for an SBML model, this set will not necessarily be complete. In order to achieve a complete parameter set, the given parameters can be used for an estimation of the missing ones. This estimation is based on linear relationships between the independent model parameters and the resulting dependent ones.
Let θ denote a vector of the logarithmic system parameters and let x be a vector holding various derived logarithmic parameters. x can be calculated by the linear relation

$$x(\theta) \quad = \quad R_\theta^x \theta, \tag{2.1}$$

where R_θ^x is a dependence matrix derived from the network structure. It represents all the relationships between the parameters and realizes the estimation of missing parameters via available parameters. The parameter types are

The system parameters should mainly be provided by the user's input file, while the dependent parameters can be estimated by the dependencies on the system parameters.

2.6.1. Bayes Estimation

The introduced parameter estimation will take place within a Bayesian framework. Bayes estimation (Gelman, 2004) can be considered a probabilistic approach that uses input data (in this case the experimental measurements of kinetic parameters) and expectations about the model parameters (prior distribution of values) for the estimation of a posterior distribution. This distribution contains the information how plausible a certain parameter set appears in correspondance to the prior and experimental data. Shown in Equation 2.1 is a linear relationship that makes it easy to use the experimental

$\mathbf{G}^{(0)}$	Gibbs free energies of formation
$\mathbf{k}^{\mathbf{M}}$	Michaelis constants for the reaction metabolites
$\mathbf{k}^{\mathbf{I}}$	Inhibitory constants for the reaction metabolites
$\mathbf{k}^{\mathbf{A}}$	Activation constants for the reaction metabolites
$\mathbf{E}$	Enzyme concentrations (one enzyme per reaction)
$\mathbf{c}$	Metabolite concentrations
$\mathbf{k}^{\mathbf{V}}$	Velocity constants for the reactions (geometric mean rate constant)
$\mathbf{k}^{\mathbf{eq}}$	Equilibrium constants for the reactions
$\mathbf{k}^{\mathbf{cat}}$	Turnover rates (forward and backward) for the reactions
$\mathbf{v}^{\mathbf{max}}$	Maximal velocities (forward and backward) for the reactions
G	Gibbs free energies
$\mathbf{A}$	Reaction affinities

Table 2.7.: The list of the parameter types to be estimated.

measurements for the parameter balancing. The posterior distribution depends on the prior distribution and the likelihood function. The prior reads

$$\theta \;=\; \mathcal{N}(\overline{\theta}_{(0)}, C_{(0)}), \tag{2.2}$$

with a probability density $p(\theta)$, mean vector $\overline{\theta}_{(0)}$, and a diagonal covariance matrix $C_{(0)}$. Our likelihood function $p(x^*|\theta)$ represents a simple model of the measurement process: the experimental values x^* are assumed to equal the values predicted by the model (plus uncorrelated additive Gaussian noise), hence

$$x^* \;=\; \mathcal{N}(x(\theta), C_x), \tag{2.3}$$

with a diagonal covariance matrix $C_x = \mathrm{diag}(\sigma)^2$. For the computation of the posterior, instead of $p(\theta|x^*)$, consider the function

$$F(\theta) \;=\; (\theta - \overline{\theta}_{(0)})^{\mathrm{T}} C_{(0)}^{-1} (\theta - \overline{\theta}_{(0)}) + (x^* - x(\theta))^{\mathrm{T}} C_x^{-1} (x^* - x(\theta)), \tag{2.4}$$

where $F(\theta)$ is assumed to be a quadratic function. As $x(\theta)$ is linear, the two terms are quadratic in θ and the corresponding posterior is Gaussian.

The posterior probability density reads $p(\theta|x^*) \sim p(x^*|\theta)p(\theta)$. $\mathcal{N}(\overline{\theta}_{(1)}, C_{(1)})$ is the multivariate Gaussian distribution with mean and covariance matrix

$$C_{(1)} \;=\; \left(C_{(0)}^{-1} + (R^{(0)})^{\mathrm{T}} C_x^{-1} R^{(0)}\right)^{-1} \tag{2.5}$$

$$\overline{\theta}_{(1)} \;=\; C_{(1)} \cdot \left((R^{(0)})^{\mathrm{T}} C_x^{(-1)} x^* + C_{(0)}^{-1} \overline{\theta}_{(0)}\right) \tag{2.6}$$

These formulae are obtained by equating 2.4 to a single quadratic function

$$(\theta - \overline{\theta}_{(0)})^{\mathrm{T}} C_{(0)}^{-1} (\theta - \overline{\theta}_{(0)}) + (x^* - x(\theta))^{\mathrm{T}} C_x^{-1} (x^* - x(\theta)) = (\theta - \overline{\theta}_{(1)})^{\mathrm{T}} C_{(1)}^{-1} (\theta - \overline{\theta}_{(1)}), \quad (2.7)$$

and solving for $\overline{\theta}_{(1)}$ and $C_{(1)}$.

$\overline{\theta}_{(1)}$ is an estimation for the given parameter values x^* and can be extended by using the complete dependence matrix R_θ^x that carries the parameter dependencies not only for the given values, but for every kinetic parameter value of the model. The construction of these necessary vectors and matrices will be elucidated in the next section.

2.6.2. Preparation of Vectors and Matrices

The parametrisation of an entire biochemical network via vectors and matrices is convenient. This usually includes as a basis the stoichiometric matrix N and the regulation matrix W (for details see example below).
An overview of the needed vectors and matrices can be found in Table 2.8.

x^* Vector holding the parameter values that are provided by the user. They are extracted from the SBtab parameter file and assumed to be incomplete, i.e. not every kinetic parameter in the model has an assigned value. If there is more than one value available for a parameter, this vector will hold the same amount of values.

C_x The diagonal covariance matrix holding the covariances for the values in x^*.

$\overline{\theta}_{(0)}$ Vector holding an average value from the prior distribution for every independent parameter in the model that shall be estimated. This prior distribution is obtained by the MetNetDB.

$C_{(0)}$ The diagonal covariance matrix holding the covariances for the values in $\overline{\theta}_{(0)}$.

R_θ^x Dependence matrix for the estimation of the parameters. It contains dependence rows for every parameter to be estimated.

$R^{(0)}$ Dependence matrix that corresponds to R_θ^x. Unlike R_θ^x, it holds only the rows of the parameters that are in fact provided in the given parameter set and lacks the parameter rows that are not provided. It can be thought of as an incomplete R_θ^x matrix.

Table 2.8.: Overview of the vectors and matrices needed for parameter estimation.

The calculation of the parameter values in Table 2.8 will be described in the following.

Thermodynamic Dependence Between Parameters

When it comes to estimating missing values, the dependencies between parameter types is crucial. This chapter enlists the dependencies of the parameters that are realized in

the dependence matrix R^x_θ, and by that are integrated into the process of parameter balancing. Background information on the shown dependencies are provided in the corresponding publication (Liebermeister, Uhlendorf & Klipp, 2010).
Derived from the second law of thermodynamics, the Gibbs free energies of formation in a metabolic system determine the equilibrium constants of the reactions. The natural logarithm of an equilibrium constant for a reaction l denotes

$$\ln k_l^{\mathrm{eq}} \quad = \quad -\sum_i n_{il} G_i^{(0)}/RT, \tag{2.8}$$

where n_{il} is the stoichiometric coefficient of metabolite i in reaction l and $G_i^{(0)}$ is the Gibbs free energy of formation of metabolite i. R is Boltzmann's gas constant ($R \approx 8.314\mathrm{J/(molK)}$), T is the absolute temperature.
The Gibbs free energy for substrate binding $\Delta G_{li}^{(0)}$ links the Gibbs free energies of metabolites l and i with the corresponding Michaelis constants k^{M}:

$$\Delta G_{li}^{(0)} \quad = \quad RT \ln k_{li}^{\mathrm{M}}. \tag{2.9}$$

Apart from upper equations, the Haldane relationship will play an important role. It can be expressed in logarithmic form as

$$\ln k_l^{\mathrm{eq}} \quad = \quad \ln k_{+l}^{\mathrm{cat}} - \ln k_{-l}^{\mathrm{cat}} + \sum_i n_{il} \ln k_{il}^{\mathrm{M}}, \tag{2.10}$$

and it does realize the dependencies between the equilibrium constants k^{eq}, the turnover rates $k_\pm^{\mathrm{cat}}$, and the Michaelis constants k^{M}. The calculation of the forward and backward turnover rates $k_\pm^{\mathrm{cat}}$ of a reaction l can be achieved by

$$\ln k_{\pm l}^{\mathrm{cat}} \quad = \quad \ln k_l^{\mathrm{V}} \pm \frac{1}{2} \sum_i n_{il}((G_i^{(0)})/RT + \ln k_{li}^{\mathrm{M}}). \tag{2.11}$$

A look at the maximal velocities $v_\pm^{\max}$ shows their influence on the enzyme concentration E, the velocity constants k^{V}, the Gibbs free energies of formation $G^{(0)}$, and the Michaelis constants k^{M}.

$$\ln v_{\pm l}^{\max} \quad = \quad \ln E_l + \ln k_l^{\mathrm{V}} \pm \frac{1}{2} \sum_i n_{il}((G_i^{(0)})/RT + \ln k_{li}^{\mathrm{M}}) \tag{2.12}$$

For this parameter balancing approach I am newly introducing the estimation of concentration dependent Gibbs free energies G and the reaction affinities A. The calculation of these values is denoted as

$$G_i \quad = \quad G_i^{(0)} + \ln c_i \cdot RT, \tag{2.13}$$

with c_i as the concentration of metabolite i (the concentrations formulated in the unit of the standard concentration 1 mM) and

$$\ln A_l = -\sum_i n_{il} G_i^{(0)} - \sum_i n_{il} \ln c_i \cdot RT. \tag{2.14}$$

These equations are implying that parameters in the whole network are coupled and dependent on each other. This fact is crucial when trying to estimate certain parameter types via other parameter types. An overview of the dependencies is shown in Table 2.9.

	$\mathbf{G}^{(0)}$	$\mathbf{k}^{\mathrm{V}}$	$\mathbf{k}^{\mathrm{M}}$	$\mathbf{E}$	$\mathbf{c}$
$\mathbf{k}^{\mathrm{eq}}$	✓	-	-	-	-
$\mathbf{k}_{\pm}^{\mathrm{cat}}$	✓	✓	✓	-	-
$\mathbf{v}_{\pm}^{\mathrm{max}}$	✓	✓	✓	✓	-
$\mathbf{G}$	✓	-	-	-	✓
$\mathbf{A}$	✓	-	-	-	✓

Table 2.9.: Parameter type dependencies, according to the model parameters (x-axis) and the dependent parameters (y-axis). A checkmark denotes the dependence among the parameter types.

Provided Parameter Set and Prior Distribution

The user provides an SBtab file that contains numeric values for the kinetic parameters of the model. The logarithms of these parameter values are stored in a large vector (except for the $G^{(0)}$ values, where the actual value is taken) x^*. The size of this vector is dependent on the provided parameter set. It can contain multiple values for one single parameter (measured under the same or under different conditions) or the model parameters can be completely lacking any measured value. Corresponding to the vector x^* I can compute the diagonal covariance matrix C_x.
Next to x^* and C_x a prior distribution is needed for the parameter estimation. For my parameter set θ the prior distribution is defined as a multivariate Gaussian distribution $\mathcal{N}$:

$$\theta = \mathcal{N}(\overline{\theta}_{(0)}, C_{(0)}) \tag{2.15}$$

The formula provides the prior mean vector $\overline{\theta}_{(0)}$, holding a mean value for every model parameter I want to estimate. These mean values are taken from the MetNetDB. Analogously to x^* and C_x, $C_{(0)}$ describes the corresponding prior diagonal covariance matrix to $\overline{\theta}_{(0)}$.
The last thing needed for the parameter estimation are the two related matrices R_θ^x and $R^{(0)}$.

Construction of Dependence Matrices

The dependence matrix R_θ^x represents all the dependencies between the parameter types (see Table 2.9) and thus realizes the possibility of estimating missing parameters via other, given parameters. The construction of R_θ^x is dependent on the given metabolic network and can be illustrated by a small example network (Liebermeister & Klipp, 2006) (see Figure 2.3).

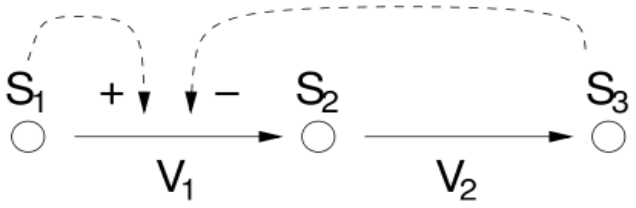

Figure 2.3.: Small example network with 3 species and two irreversible reactions: species S1 is converted to species S2, which is then converted to species S3. Moreover, S1 activates reaction V1, species S3 inhibits V1.

The corresponding stoichiometric matrix N (can be obtained from the SBML source code) and the regulation matrix W are shown in Figure 2.4.

$$N = \begin{pmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{pmatrix}, \quad W = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}$$

Figure 2.4.: Corresponding stoichiometric matrix N and regulation matrix W. The columns of N are referring to the reactions of the network, the rows are the species. If a position of N holds a 1, the species is produced in the reaction, a -1 denotes the consumption of the species. The regulation matrix W works analogously, except for the columns referring to species and the rows referring to reactions. A value of -1 in W denotes an inhibitory influence, 1 is an activation.

Assuming that all system parameters, equilibrium constants, turnover rates, and maximal velocities can be measured, the matrix R_θ^x reads:

	$G_1^{(0)}$	$G_2^{(0)}$	$G_3^{(0)}$	k_1^{V}	k_2^{V}	k_{11}^{M}	k_{12}^{M}	k_{22}^{M}	k_{23}^{M}	k_{11}^{A}	k_{13}^{I}	E_1	E_2	c_1	c_2	c_3
$G_1^{(0)}$	1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
$G_2^{(0)}$	·	1	·	·	·	·	·	·	·	·	·	·	·	·	·	·
$G_3^{(0)}$	·	·	1	·	·	·	·	·	·	·	·	·	·	·	·	·
k_1^{V}	·	·	·	1	·	·	·	·	·	·	·	·	·	·	·	·
k_2^{V}	·	·	·	·	1	·	·	·	·	·	·	·	·	·	·	·
k_{11}^{M}	·	·	·	·	·	1	·	·	·	·	·	·	·	·	·	·
k_{12}^{M}	·	·	·	·	·	·	1	·	·	·	·	·	·	·	·	·
k_{22}^{M}	·	·	·	·	·	·	·	1	·	·	·	·	·	·	·	·
k_{23}^{M}	·	·	·	·	·	·	·	·	1	·	·	·	·	·	·	·
k_{11}^{A}	·	·	·	·	·	·	·	·	·	1	·	·	·	·	·	·
k_{13}^{I}	·	·	·	·	·	·	·	·	·	·	1	·	·	·	·	·
E_1	·	·	·	·	·	·	·	·	·	·	·	1	·	·	·	·
E_2	·	·	·	·	·	·	·	·	·	·	·	·	1	·	·	·
c_1	·	·	·	·	·	·	·	·	·	·	·	·	·	1	·	·
c_2	·	·	·	·	·	·	·	·	·	·	·	·	·	·	1	·
c_3	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	1
k_1^{eq}	1	−1	·	·	·	·	·	·	·	·	·	·	·	·	·	·
k_2^{eq}	·	1	−1	·	·	·	·	·	·	·	·	·	·	·	·	·
k_{+1}^{cat}	$-\frac{1}{2}$	$\frac{1}{2}$	·	1	·	$\frac{1}{2}$	$-\frac{1}{2}$	·	·	·	·	·	·	·	·	·
k_{+2}^{cat}	·	$-\frac{1}{2}$	$\frac{1}{2}$	·	1	·	·	$\frac{1}{2}$	$-\frac{1}{2}$	·	·	·	·	·	·	·
k_{-1}^{cat}	$\frac{1}{2}$	$-\frac{1}{2}$	·	1	·	$-\frac{1}{2}$	$\frac{1}{2}$	·	·	·	·	·	·	·	·	·
k_{-2}^{cat}	·	$\frac{1}{2}$	$-\frac{1}{2}$	·	1	·	·	$-\frac{1}{2}$	$\frac{1}{2}$	·	·	·	·	·	·	·
$v_{+1}^{\max}$	$-\frac{1}{2}$	$\frac{1}{2}$	·	1	·	$\frac{1}{2}$	$-\frac{1}{2}$	·	·	·	·	1	·	·	·	·
$v_{+2}^{\max}$	·	$-\frac{1}{2}$	$\frac{1}{2}$	·	1	·	·	$\frac{1}{2}$	$-\frac{1}{2}$	·	·	·	1	·	·	·
$v_{-1}^{\max}$	$\frac{1}{2}$	$-\frac{1}{2}$	·	1	·	$-\frac{1}{2}$	$\frac{1}{2}$	·	·	·	·	1	·	·	·	·
$v_{-2}^{\max}$	·	$\frac{1}{2}$	$-\frac{1}{2}$	·	1	·	·	$-\frac{1}{2}$	$\frac{1}{2}$	·	·	·	1	·	·	·
G_1	1	·	·	·	·	·	·	·	·	·	·	·	·	1	·	·
G_2	·	1	·	·	·	·	·	·	·	·	·	·	·	·	1	·
G_3	·	·	1	·	·	·	·	·	·	·	·	·	·	·	·	1
A_1	1	−1	·	·	·	·	·	·	·	·	·	·	·	1	−1	·
A_2	·	1	−1	·	·	·	·	·	·	·	·	·	·	·	1	−1

where the upper part is an identity matrix and every row corresponds to one system parameter (dots represent zeros). The lower part is showing the dependence of the equilibrium constants, turnover rates, maximal velocities, Gibbs free energies, and the reaction affinities to the system parameters, based on $\ln \mathbf{k}^{\mathrm{eq}} = -N^{\mathrm{T}} G^{(0)}/RT$. The dependence matrix R_θ^x can be written in block matrix form

$$
\begin{array}{c}
 \\
G^{(0)} \\
k^{\mathrm{V}} \\
k^{\mathrm{M}} \\
k^{\mathrm{A}} \\
k^{\mathrm{I}} \\
E \\
c \\
k^{\mathrm{eq}} \\
k_{+}^{\mathrm{cat}} \\
k_{-}^{\mathrm{cat}} \\
v_{+}^{\max} \\
v_{-}^{\max} \\
G^{0} \\
A
\end{array}
\begin{array}{c}
\begin{array}{cccccccc}
G^{(0)} & k^{\mathrm{V}} & k^{\mathrm{M}} & k^{\mathrm{A}} & k^{\mathrm{I}} & E & c
\end{array} \\
\left(\begin{array}{cccccccc}
\mathbf{I} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \mathbf{I} & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \mathbf{I} & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \mathbf{I} & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \mathbf{I} & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \mathbf{I} & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \mathbf{I} \\
-N^{\mathrm{T}}/RT & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\frac{1}{2}N^{\mathrm{T}}/RT & \mathbf{I} & -\frac{1}{2}Z & \cdot & \cdot & \cdot & \cdot \\
-\frac{1}{2}N^{\mathrm{T}}/RT & \mathbf{I} & \frac{1}{2}Z & \cdot & \cdot & \cdot & \cdot \\
\frac{1}{2}N^{\mathrm{T}}/RT & \mathbf{I} & -\frac{1}{2}Z & \cdot & \cdot & \mathbf{I} & \cdot \\
-\frac{1}{2}N^{\mathrm{T}}/RT & \mathbf{I} & \frac{1}{2}Z & \cdot & \cdot & \mathbf{I} & \cdot \\
\mathbf{I} & \cdot & \cdot & \cdot & \cdot & \cdot & RT \cdot \mathbf{I} \\
-N^{\mathrm{T}} & \cdot & \cdot & \cdot & \cdot & \cdot & -N^{\mathrm{T}} \cdot RT
\end{array}\right)
\end{array}
$$

where N is the stoichiometric matrix and each column of the matrix Z corresponds to one of the k_{li}^{M} values (for a reaction l and metabolite i):

$$
\begin{array}{c}
 \\
v_1 \\
v_2
\end{array}
\begin{array}{c}
\begin{array}{cccc}
k_{11}^{\mathrm{M}} & k_{12}^{\mathrm{M}} & k_{22}^{\mathrm{M}} & k_{23}^{\mathrm{M}}
\end{array} \\
\left(\begin{array}{cccc}
-1 & 1 & \cdot & \cdot \\
\cdot & \cdot & -1 & 1
\end{array}\right)
\end{array}
$$

This corresponds to the stoichiometric coefficients of reaction l and zeros for every other reaction. Keeping this block matrix form in mind, it is now necessary to perform several alterations on it. The following construction of $R^{(0)}$ is depending on the given kinetic parameters. If the vector x^* holds a value for a specific parameter, then the corresponding parameter row from R_θ^x is also a part of $R^{(0)}$. Otherwise, if there is no value for a certain parameter of the model, the corresponding row from R_θ^x does not appear in $R^{(0)}$. Furthermore, if there are multiple values for one single parameter, this leads to duplications of the corresponding rows. If four values for a single parameter are given, the corresponding row from R_θ^x is appearing four times in $R^{(0)}$. If parameters for the example model in the form of a "KineticData" SBtab are provided, the parameter set can look like shown in Appendix A.1. Given the example network (see Figure 2.3) in SBML format and the parameter set (see Appendix A.1) in SBtab format, I can construct the incomplete dependence matrix $R^{(0)}$, based on the complete R_θ^x and the given parameters (see Table 2.10).

This matrix $R^{(0)}$ is the foundation of the parameter estimation, it can even be extended to take the pH values and temperatures as differing measuring circumstances into account (see Chapter 2.6.4).

	$G_1^{(0)}$	$G_2^{(0)}$	$G_3^{(0)}$	k_{11}^{M}	k_{12}^{M}	k_{22}^{M}	k_{23}^{M}	k_{11}^{A}	k_{13}^{I}	E_1	E_2	c_1	c_2	c_3
$G_1^{(0)}$	1	·	·	·	·	·	·	·	·	·	·	·	·	·
$G_1^{(0)}$	1	·	·	·	·	·	·	·	·	·	·	·	·	·
$G_1^{(0)}$	1	·	·	·	·	·	·	·	·	·	·	·	·	·
$G_3^{(0)}$	·	·	1	·	·	·	·	·	·	·	·	·	·	·
k_{11}^{M}	·	·	·	1	·	·	·	·	·	·	·	·	·	·
k_{11}^{M}	·	·	·	1	·	·	·	·	·	·	·	·	·	·
k_{12}^{M}	·	·	·	·	1	·	·	·	·	·	·	·	·	·
k_{13}^{I}	·	·	·	·	·	·	·	·	1	·	·	·	·	·
k_{13}^{I}	·	·	·	·	·	·	·	·	1	·	·	·	·	·
c_2	·	·	·	·	·	·	·	·	·	·	·	·	1	·
k_1^{eq}	1	−1	·	·	·	·	·	·	·	·	·	·	·	·
k_1^{eq}	1	−1	·	·	·	·	·	·	·	·	·	·	·	·
k_{+1}^{cat}	$-\frac{1}{2}$	$\frac{1}{2}$	·	$\frac{1}{2}$	$-\frac{1}{2}$	·	·	·	·	·	·	·	·	·

Table 2.10.: The corresponding $R^{(0)}$-matrix for my given network. Only the lines with provided parameters are present (compare to the input data in Table A.1). If multiple parameter values are given, the lines for this value are duplicated as well. If one parameter value is missing, the corresponding line does not appear in $R^{(0)}$.

2.6.3. Calculation Specifics

There are several conditions concerning the calculations that have to be mentioned. It is important to note the distributions I work with, and how my tool handles mean values of multiple entries, or reacts on missing standard deviations.

Log-normal Distributions

In order to be able to work with linear relationships between the parameter values, I am working with log-normal distributions (with the exception of the $G^{(0)}$ values). If the logarithm of the value is normal distributed, its probability distribution is a log-normal distribution (Aitchison & Brown, 1969). One value can be described as log-normal, if it is able to be the multiplicative product of positive independent random variables: Since the values are normal distributed, it is possible to perform multiplications with the matrices (see Chapter 2.6.5) and again retrieve normal distributed values as a result. The linear relationships and normal distributions are consistent, but the mean values are difficult to translate (due to the dependence on the standard deviation). A possible solution is the use of the median. Since the mean and standard deviation of the value are known, parameters μ and σ of the parameter value can be obtained as

$$\mu_{\ln} = \ln(\mu) - \frac{1}{2}\ln\left(1 + (\frac{\sigma}{\mu})^2\right), \tag{2.16}$$

and

$$\sigma_{\ln} = \sqrt{\ln\left(1 + (\frac{\sigma}{\mu})^2\right)}. \tag{2.17}$$

Computing a Mean Over Multiple Parameter Values

When the given kinetic data includes more than one value for a parameter, its mean value can be computed for the use in further calculations. If the vector of values for parameter x is x_i and the corresponding standard deviations are σ_i, the mean can be calculated as

$$\overline{x} = \frac{\sum_i \frac{x_i}{\sigma_i^2}}{\sum_i \frac{1}{\sigma_i^2}}. \tag{2.18}$$

The corresponding standard deviation to $\overline{x}$ is

$$\sigma_{\overline{x}} = \sqrt{\frac{1}{\sum_i \frac{1}{\sigma_i^2}}}. \tag{2.19}$$

.

Coefficient of Variation

If a parameter entry with a mean value is given, but no corresponding standard deviation, a default standard variation that is based on the coefficient of variation, is constructed. This is the ratio between the standard deviation and the mean:

$$c_v = \frac{\sigma}{\mu}, \tag{2.20}$$

where σ is the default standard deviation chosen arbitrarily by the user (or taken from standard prior values provided by the tool).

Pseudo Equilibrium Constants

Performed tests of the parameter balancing approach have revealed problems when it comes to the lack of certain input data. The missing of equilibrium constants has the biggest influence on many other parameters (such as the velocities, the turnover rates, and the reaction affinities). In case of a missing equilibrium constant, the tool

automatically generates a new one via the dependence on the Gibbs free energies. Since these energies are often provided with high standard deviations and their magnitude itself differs from the other parameter types (for Gibbs free energies no logarithmic value is used, see 2.6.3), the produced equilibrium constants are often inappropriately high and unstable.
To avoid the generation of missing equilibrium constants via the Gibbs free energies, I have introduced pseudo equilibrium constants. If the equilibrium constant of a reaction in a model is missing in the input data, the program automatically generates a pseudo constant that corresponds to the prior values for equilibrium constants. The prior mean generated from the MetNetDB denotes 2980.96, with a standard deviation of 8886110.0. These values originate in the assumption of the logarithm of the underlying value to be balanced at 0 (equilibrium). Using pseudo equilibrium constants for a newly generated equilibrium constant is a big advantage in comparison to a value generation via the Gibbs free energies, which is demonstrated in the results section. Furthermore, their use can be described as equivalent to an alteration of the prior data for Gibbs free energies of formation, which makes extreme equilibrium constant values unlikely.

2.6.4. Regression of Temperature and pH Values

A significant influence on the parameter values is exerted by the measuring circumstances. The Gibbs free energies of formation and the equilibrium constants directly depend on the temperature and the pH values they have been measured in. It would be appropriate to take these circumstances into account in order to be able to generate more reliable estimation results. Fortunately, often the kinetic parameter values from the MetNetDB are coming with a corresponding temperature and pH value.
The calculation of a Gibbs free energy of formation in dependence of the corresponding temperature denotes

$$\mathrm{G}_{\mathrm{T}}^{(0)} = \mathrm{G}_{\mathrm{T}_{(0)}}^{(0)} \cdot \Delta\mathrm{T}, \tag{2.21}$$

where T is the temperature that is delivered together with the value, $\mathrm{T}_{(0)}$ is the desired temperature for the reaction, and ΔT is the difference between the two. The calculation of a dependence on the pH value works analogously. Having in mind the generation of the dependence matrix R_θ^x, I have to come up with a method to integrate this dependence into the structure of the matrix. Until so far, the matrix rows for the parameter entries of the Gibbs free energies of formation are corresponding to the unity matrix. By expanding these specific rows in reference to Equation 2.21 the temperature and pH values can be taken into account. In addition to the unity matrix now the values for $\mathrm{dG}^{(0)}/\mathrm{dpH}$ and $\mathrm{dG}^{(0)}/\mathrm{dT}$ values are integrated as factors for the Gibbs free energies of formation and for the equilibrium constants. The resulting matrix is shown in Table 2.11.

For the calculation of the $\mathrm{dG}^{(0)}/\mathrm{dT}$, respectively $\mathrm{dG}^{(0)}/\mathrm{dpH}$ values not only the provided value is needed but also a favoured value. In my implementation this value is set to a default of 300 K for the temperature and a pH value of 7. The user is able to alter these values according to the requirements.

	$G_1^{(0)}$	$G_2^{(0)}$	$G_3^{(0)}$	$G_T^{(0)}$	$G_T^{(0)}$	$G_T^{(0)}$	$G_{pH}^{(0)}$	$G_{pH}^{(0)}$	$G_{pH}^{(0)}$	k_{11}^{M}	...
$G_1^{(0)}$	1	·	·	ΔT	·	·	ΔpH	·	·	·	...
$G_1^{(0)}$	1	·	·	ΔT	·	·	ΔpH	·	·	·	...
$G_1^{(0)}$	1	·	·	ΔT	·	·	ΔpH	·	·	·	...
$G_3^{(0)}$	·	·	1	·	·	ΔT	·	·	ΔpH	·	...
k_{11}^{M}	·	·	·	·	·	·	·	·	·	1	...
k_{11}^{M}	·	·	·	·	·	·	·	·	·	1	...
k_{12}^{M}	·	·	·	·	·	·	·	·	·	·	...
k_{13}^{I}	·	·	·	·	·	·	·	·	·	·	...
k_{13}^{I}	·	·	·	·	·	·	·	·	·	·	...
c_2	·	·	·	·	·	·	·	·	·	·	...
k_1^{eq}	1	−1	·	$1 \cdot \Delta T$	$-1 \cdot \Delta T$	·	$1 \cdot \Delta pH$	$-1 \cdot \Delta pH$	·	·	...
k_1^{eq}	1	−1	·	$1 \cdot \Delta T$	$-1 \cdot \Delta T$	·	$1 \cdot \Delta pH$	$-1 \cdot \Delta pH$	·	·	...
k_{+1}^{cat}	$-\frac{1}{2}$	$\frac{1}{2}$	·	·	·	·	·	·	·	$\frac{1}{2}$	...

Table 2.11.: The final $R^{(0)}$ matrix including the new columns for pH and temperature regression. It includes the columns $\mathrm{G}_{\mathrm{T}}^{(0)}$ and $\mathrm{G}_{\mathrm{pH}}^{(0)}$ for temperature and pH regression. According to that, the rows of the G^0 values and the equilibrium constants have to be extended corresponding to their different temperature and pH values. The $\mathrm{dG}^{(0)}/\mathrm{dpH}$ and $\mathrm{dG}^{(0)}/\mathrm{dT}$ values are representing the differences between provided value and favoured value.

Finally it is important to note that the newly integrated columns hold the $\mathrm{dG}^{(0)}/\mathrm{dpH}$ and $\mathrm{dG}^{(0)}/\mathrm{dT}$ values for the measuring circumstances of Gibbs free energies of formation and the equilibrium constants (if available), but are always set to zero for every other kind of parameter type. Furthermore, the necessity for using this regression method is that there are more than only one value for the specific parameter provided by the input file.
The only difference in this approach for Gibbs free energies of formation and equilibrium constants, is the multiplication of the $\mathrm{dG}^{(0)}/\mathrm{dpH}$ and $\mathrm{dG}^{(0)}/\mathrm{dT}$ values with the stoichiometric matrix instead of the unity matrix. For clarification please see Table 2.11.

2.6.5. Parameter Balancing

By now I have constructed the vector of given parameter values x^*, the corresponding diagonal covariance matrix C_x, as well as the prior distribution vector $\overline{\theta}_{(0)}$ and its corresponding prior diagonal covariance matrix $C_{(0)}$. Furthermore I have the complete dependence matrix R_θ^x and the related $R^{(0)}$. Being equipped with these vectors and matrices the parameter estimation (Liebermeister & Klipp, 2006) can be realized. The

posterior covariance matrix $C_{(1)}$ is calculated as

$$C_{(1)} = \left(C_0^{-1} + (R^{(0)})^{\mathrm{T}} C_x^{-1} R^{(0)}\right)^{-1}. \tag{2.22}$$

The posterior mean vector $\overline{\theta}_{(1)}$ denotes

$$\overline{\theta}_{(1)} = C_{(1)} \cdot \left((R^{(0)})^{\mathrm{T}} C_x^{(-1)} x^* + C_{(0)}^{-1} \overline{\theta}_{(0)}\right). \tag{2.23}$$

Until now I only have constructed the estimation for the given values according to the prior distributions. In order to gain the whole vector of parameter values and the complete covariance matrix I have to include the complete dependence matrix R_θ^x into my calculation:

$$x_{fit} = R_\theta^x \overline{\theta}_{(1)}, \tag{2.24}$$

and

$$C(x_{fit}) = R_\theta^x C_{(1)} (R_\theta^x)^{\mathrm{T}}. \tag{2.25}$$

The vector x_{fit} is holding one estimated value for every parameter appearing in the model, the covariance matrix $C(x_{fit})$ contains the corresponding covariances.

2.7. SBML Models Under Examination

For classifying the powers of my parameter estimation I have tested it on different models with various parameter sets. In order to adjust and configure the big number of matrices and vectors that are constructed out of the model and the model data, I have chosen a very small SBML model. The phosphofructokinase reaction will be the foundation for my testings. The other models, numerous in reactions and species, will deliver estimation results that have to be tested for their reliability. An overview of the examined models can be found in Table 2.12.

- Phosphofructokinase reaction
- Nazaret citrate cycle model (Nazaret, Heiske, Thurley & Mazat, 2009)
- Teusink glycolysis model (Teusink, Passarge, Reijenga, Esgalhado, van der Weijden, Schepper, Walsh, Bakker, van Dam, Westerhoff & others, 2000)
- Chassagnole threonine model (Chassagnole, Rais, Quentin, Fell & Mazat, 2001)

Table 2.12.: The SBML models used in this thesis.

3. semanticSBML Extension

This chapter gives on overview of the implementation results. I have extended the semanticSBML GUI for the parameter balancing approach with numerous options and possibilities. A detailed description of these functions including screenshots and important hints can be found in the appendix A.2, while this chapter only holds an extract concerning its most important informations.

I have implemented a user-friendly interface for working on parameter estimation and data integration by extending the existing Systems Biology tool semanticSBML (see Chapter 2.4). Apart from the available functions of SBMLfill, I added the possibility to access, modify, and integrate kinetic data into an SBML model. The automated process of parameter balancing is part of this implementation as well.
The insertion of model parameters including parameter balancing via the semanticSBML user interface can be performed in five steps:

1. Open an SBML model and use the Fill option in semanticSBML (explained in Appendix A.2.1): This new option offers the possibility to choose a kinetic rate law that is attached to the reactions of the model. Several insertion options, similar to those of SBMLfill (see Chapter 2.2), are available.

2. Open an SBtab parameter file that corresponds to the model (explained in Appendix A.2.2): The user can open an SBtab file, perform a validation test on its semantics, and has the possibility to do alterations on the file's content.

3. Automatically generate a new SBtab that matches all SBML insertion criteria (explained in Appendix A.2.3): In order to match the strict insertion criteria, a new SBtab file is generated automatically. It consists of the users SBtab file and entries for those parameters that are required, but not provided by the users SBtab file.

4. Optional: insert default values (explained in Appendix A.2.4): default mean values from the prior distribution are attached to the newly generated parameter entries in the second SBtab file.

5. Start parameter balancing for the parameter set (explained in Appendix A.2.5): parameter balancing is performed. Several options are available (setting a favoured pH value and temperature, excluding certain parameter types from the balancing process).

6. Insert the kinetic rate laws and the balanced parameter set into the SBML model (explained in Appendix A.2.6): the user can attach the favoured kinetic rate laws to the model. The balancing process has produced the parameter values for the kinetic rate laws.

The implementation is working stable for any SBML model, while the SBtab files are underlying strict criteria. The extensive manual to the GUI is located in the Appendix A.2.

4. Parameter Estimation on Models

This chapter contains the testing of my parameter balancing approach on several SBML models. In order to configure and adjust all the complex calculations for the matrices and vectors that need to be generated for every model, I chose to test the program's abilities on a very small model. The whole model can be analyzed and its corresponding vectors and matrices checked by hand, to assure their correctness. Since they are build automatically for any kind of input model, this process has to be checked in detail to prevent possible implementation errors. I will perform the subsequent parameter balancing processes for larger scale models, using visualization methods provided by Highcharts (www.highcharts.com). The upcoming analyzes comprise:

Phosphofructokinase reaction This reaction is a small model that will hold for extensive testings of the parameter balancing process and its several options:

1. Collecting all the data vectors and matrices needed for parameter balancing (see Chapter 4.1.1).
2. Performing a simple parameter balancing (see Chapter 4.1.2).
3. Checking the balancing result for thermodynamic validity (see Chapter 4.1.3).
4. Estimations of the missing equilibrium constant (see Chapter 4.2).
5. Parameter balancing with temperature and pH value regression (see Chapter 4.2.1).
6. Parameter balancing with limited input concerning the parameter types (see Chapter 4.2.2).

Citric acid cycle For the model of the citric acid cycle by Nazaret et al. (Nazaret, Heiske, Thurley & Mazat, 2009) I am performing

1. a parameter balancing with temperature and pH value regression (see Chapter 4.3),
2. and a repairing of thermodynamic invalidities of the input data (see Chapter 4.3.1)

Glycolysis model Parameter balancing for the whole glycolysis model by Teusink et al. (Teusink, Passarge, Reijenga, Esgalhado, van der Weijden, Schepper, Walsh, Bakker, van Dam, Westerhoff & others, 2000) (see Chapter 4.4).

Threonine model Parameter balancing for the whole threonine model by Chassagnole et al. (Chassagnole, Rais, Quentin, Fell & Mazat, 2001) (see Chapter 4.5)

A table that holds information on the models can be found in Appendix A.3.1: the amount of reactions and species, the available data on MetNetDB, the original model data, and comparisons of the latter with the balanced results.

4.1. Configuration and Adjust: the Phosphofructokinase Reaction

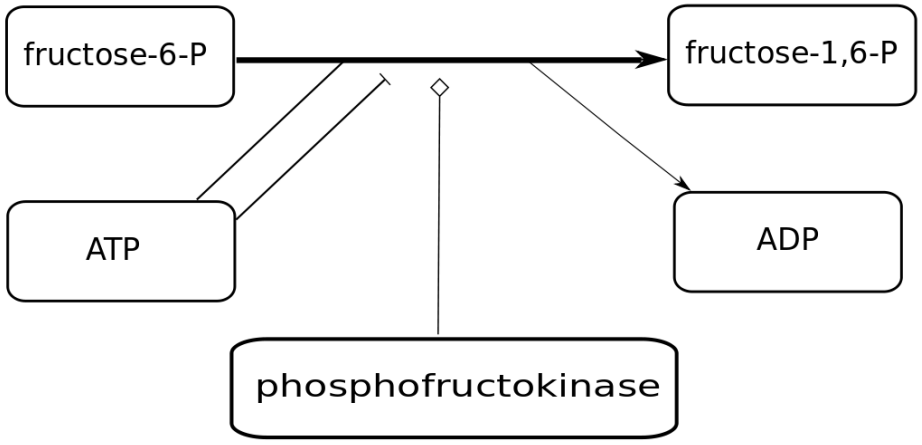

Figure 4.1.: The phosphofructokinase reaction: fructose-6-phosphate becomes fructose-1,6-bisphosphate under the conversion of ATP to ADP. The enzyme phosphofructokinase catalyses the reaction, and ATP is inhibiting the reaction.

For the configuration, adjust, and detailed analysis of the parameter balancing process and its options, I have chosen the small model of the phosphofructokinase reaction (PFK) (see Figure 4.1), which is part of the glycolysis: it describes the conversion of the metabolite fructose-6-phosphate into fructose-1,6-bisphosphate by adding a phosphate group. The enzyme phosphofructokinase catalyses the reaction, and ATP is consumed and converted to ADP. The analysis of such a small model is actually no waste of time, since even the very extensive Sabio-RK database (Wittig, Golebiewski, Kania, Krebs, Mir, Weidemann, Anstein, Saric & Rojas, 2006) does not hold a complete kinetic parameter set for it. Available are only one concentration, one Michaelis constant, the reaction affinity, and the maximal velocity. In contrast, the parameter set that is obtained be the parameter balancing, can be found in Table A.4 in the Appendix.
My objective is to gain a complete kinetic parameter set, hence the program analyses the input model (PFK reaction) via its SBML code. The code delivers the information on how many species, reactions, and/or modifiers there are, and which kinetic parameters will be needed to describe the model in a whole. A list of every kinetic parameter needed for the PFK reaction is shown in the Appendix Table A.4.
The process of parameter balancing requires the generation of several matrices and vectors (listed and explained in Chapter 2.6). In the following, I describe how these

matrices and vectors are built for the example of the PFK reaction. They can be divided in the model data x^* and C_x that is provided by the user, the prior distribution $\overline{\theta}_{(0)}$ and C_0 derived from MetNetDB, and finally the dependence matrices R_θ^x and $R^{(0)}$.

4.1.1. Collecting the Required Data for the Parameter Balancing of PFK Reaction

Model Values for PFK Reaction: x^* and C_x The model values are without exception obtained by the data that is provided by the user in the SBtab file format. This model data is assumed to be incomplete, since several kinetic parameters cannot be measured, have not been measured, or simply are not available for the user. In order to get first balancing results, I am submitting an incomplete SBtab file to the program, which holds values for the concentrations of the species, the inhibitory constant, the Michaelis constants, and the standard chemical potentials. The values are originating mainly from Brenda (Barthelmes, Ebeling, Chang, Schomburg & Schomburg, 2007) and KMedDB (http://sysbio.molgen.mpg.de/KMedDB), but also in several cases from yeastGFP (Chen, Zhao, Gordon & Murphy, 2007) and some single publications (Beyer et al. (2004), MolCellProt, Greenbaum et al. (2002), Bioinformatics). An extract of this initial incomplete kinetic parameter set is shown in the Appendix A.4. The incomplete parameter set is representing the model data and by that will be my foundation in constructing the model vector x^* that consists of the logarithmic values of the given parameter values (please note that the Gibbs free energies of formation are not calculated as logarithms). The prior vector x^* for the PFK reaction consists of

- 20 Gibbs free energies of formation,
- 19 concentrations,
- 14 Michaelis constants,
- and 8 inhibitory constants (Please note that the inhibitory constant entries carrying only a reaction, but no compound, are automatically removed, since they cannot be identified clearly. Furthermore, the inhibitory constants that appear in the SBtab holding the **wrong** compound (ADP instead of ATP) are removed as well (seldomly wrong assignments might occur)),

and so adds up to a vector of 61 entries. These entries have been measured in 24 different organisms. The largest fraction holds 6 entries (Rattus sp.), followed by Rattus norvegicus and Saccharomyces cerevisiae (both 4 entries). Corresponding to the entries of the x^* vector, the diagonal covariance matrix C_x is of the dimension 61x61 and holds on its diagonal the covariances for each of the given values. The covariances are taken from the given standard deviations in the SBtab file, for which we denote

$$\mathrm{Cov_x} = \sigma_\mathrm{x}^2, \tag{4.1}$$

where $\mathrm{Cov_x}$ is the covariance of the measured parameter value x and σ_x is the standard deviation. As it is shown in Table A.4, not every parameter value comes with a standard deviation. Often this column is set to '-', because there is no standard deviation

available. In this case we use the default standard deviation for the specific parameter type. This default value is taken from a list of averages for each parameter type, taken from the data of the MetNetDB. This data is also the foundation for the next step, the generation of the prior distribution.

MetNetDB data for prior distribution of PFK Reaction: $\overline{\theta}_{(0)}$ and C_0 The reasonable choice of a prior distribution is crucial for the success of the parameter balancing approach: everytime a parameter value is missing in the input data, this value has to be generated from the knowledge of the other parameters and the prior distribution. A prior distribution can be generated by collecting extensive information on parameter values from web resources. The prior distribution that I have constructed for the parameter balancing is referring to the data available in the MetNetDB (see Chapter 2.5). It comprises 101.122 parameter values. The calculation of average values for each parameter type is shown in Table 4.1.

Parameter type	Average of values	Standard deviation	Amount of values
Gibbs free energy of formation	-435.15 kJ/mol	642.99	10629
Michaelis constant	15.5 mM	392.1	62740
Inhibitory constant	11.9 mM	173	12827
Concentration	1.42 mM	4.5	755
Equilibrium constant	2980.96	8886110.0	2088
Turnover rates	1898 1/s	26757	12083

Table 4.1.: The prior distribution derived from MetNetDB. Used for the prior vector $\overline{\theta}_{(0)}$ and the prior covariance matrix C_0 are the logarithmic values of the parameter type averages. The generation of these values is explained in Chapter 2.6.3.

The prior distribution consists of average values from the MetNetDB and the corresponding standard deviations. k^{V} values (geometric mean rate constant) are derived from the turnover rates k^{cat}. Pseudo equilibrium constants (see Chapter 4.2) are generated by the values for the equilibrium constants. Since Table A.4 shows which parameters are needed for the model, the corresponding values from the prior distribution can be obtained. The values that cannot be obtained by the MetNetDB (maximal velocitites, reaction affinities), are excluded from the prior distribution, and the estimation of these values will be realized via the dependence matrices.
Given this data I construct a prior vector $\overline{\theta}_{(0)}$ of length 15. It consists of the logarithmic values for the following parameter types (the mean of the logarithm in brackets):

- 4 Gibbs free energies of formation for the 4 species
- 1 catalytic rate constant geometric mean for the reaction (0.66)
- 4 Michaelis constants for the species involved in the reaction (-0.86)
- 1 inhibitory constant for the species ATP (-1.95)

- 4 species concentrations (-0.92)
- 1 enzyme concentration (-0.92)

Corresponding to this prior vector of length 15 the prior covariance matrix $C_{(0)}$ of the dimension 15x15 is constructed, holding the covariances of the values in vector $\overline{\theta}_{(0)}$.

The Dependence Matrices for PFK Reaction: R_θ^x and $R^{(0)}$ Finally, the dependence matrices have to be built. The matrix R_θ^x, representing the dependencies among the parameter types and build up like shown in Chapter 2.6.2, has the dimension of 26x15. Its y-axis (length = 15 columns) corresponds to the system parameter values that can also be found in vector $\overline{\theta}_{(0)}$. Meanwhile, the x-axis (length = 26 rows) corresponds to the same values as the y-axis plus the dependent parameter types: one row for the equilibrium constant, two rows for the turnover rates forward and backward, two rows for the maximal velocities forward and backward, five rows for the G values and one row for the reaction affinity, which adds up to the amount of 26 rows. How this matrix looks like is described in Chapter 2.6.2 in detail.
The dependence matrix $R^{(0)}$, which holds all the rows of R_θ^x for which we have measured values in our model data, has the dimension of 61x15. Its 15 columns are obviously referring to the same entries of the system parameters that can be found in the complete R_θ^x matrix. Each of the values that is given in my model data (length of vector x^*: 61) represents one of the 61 rows. By now I have constructed the model data, the prior values, and the two dependence matrices. Given these vectors and matrices I can perform the parameter balancing according to Chapter 2.6.5.

4.1.2. Parameter Balancing for the PFK Reaction

Given the required matrices and vectors, the parameter balancing can be performed. The balancing results for the PFK reaction referring to our given model data and the prior distribution are shown in Table 4.2. The several influences on the values can be seen easily: while the automatically generated parameter entries that did not provide an input value are similar or equal to their corresponding prior values generated from MetNetDB, the other parameter values are mutually influenced by the prior values, their provided input values, and the standard deviation.
The results show serious problems concerning the magnitude of the dependent parameter types: the equilibrium constant, the turnover rates, the maximal velocities, and the reaction affinity are unusually high. Since the equilibrium constant for the appearing reaction was not provided by the input data, it had to be constructed automatically. By that, it is strongly depending on the given Gibbs free energies and their prior standard deviations (the prior standard deviation is used for them, because they were not provided with one in the input data). I assume that the prior standard deviation of the Gibbs free energies of formation is responsible for the high values of the equilibrium constants, and the other dependent parameter types. To fix this problem, an option will be added to the GUI that creates "pseudo equilibrium constants" in case of missing constants, in order to keep the values in a sensible range, instead of relying on the values of the Gibbs free energies. This is equivalent to the choice of a better prior standard deviation for the Gibbs free energies.

QuantityType	Reaction	Compound	Unit	Value	Std	μ(ln)	σ(ln)
Gibbs free energy of formation	-	F16P	kJ/mol	-1695.02	10.35	-	-
Gibbs free energy of formation	-	ATP	kJ/mol	-1643.38	10.35	-	-
Gibbs free energy of formation	-	F6P	kJ/mol	-1101.83	10.35	-	-
Gibbs free energy of formation	-	ADP	kJ/mol	-1061.63	10.35	-	-
geometric mean rate constant	PFK-reaction	-	1/s	426.76	1277.38	4.91	1.52
Michaelis constant	PFK-reaction	F6P	mM	2.18	7.46	-0.49	1.59
Michaelis constant	PFK-reaction	F16P	mM	2.18	7.46	-0.49	1.59
Michaelis constant	PFK-reaction	ATP	mM	0.02	0.002	-3.97	0.08
Michaelis constant	PFK-reaction	ADP	mM	1.09	1.94	-0.63	1.19
inhibitory constant	PFK-reaction	ATP	mM	0.21	0.01	-1.58	0.05
concentration of enzyme	PFK-reaction	PFK	mM	0.93	1.79	-0.85	1.25
concentration	PFK-reaction	F6P	mM	0.93	1.79	-0.85	1.25
concentration	PFK-reaction	F16P	mM	0.93	1.79	-0.85	1.25
concentration	PFK-reaction	ATP	mM	2.02	0.43	0.68	0.21
concentration	PFK-reaction	ADP	mM	1.22	0.21	0.19	0.17
equilibrium constant	PFK-reaction	-		9.0e+16	8.3e+31	4.59	8.30
turnover rate forward	PFK-reaction	-	1/s	2.8e+06	1.1e+11	4.28	4.6
turnover rate backward	PFK-reaction	-	1/s	9.8e+06	3.9e+11	5.53	4.6
maximal velocity forward	PFK-reaction	-	1/s	2.6e+06	2.2e+11	3.43	4.77
maximal velocity backward	PFK-reaction	-	1/s	9.2e+06	7.8e+11	4.68	4.77
Gibbs free energy	PFK-reaction	ATP	mM	-1641.68	10.37	-	-
Gibbs free energy	PFK-reaction	F6P	mM	-1103.96	10.81	-	-
Gibbs free energy	PFK-reaction	ADP	mM	-1061.17	10.36	-	-
Gibbs free energy	PFK-reaction	F16P	mM	-1697.15	10.81	-	-
reaction affinity	PFK-reaction	-	mM	7.3e+17	3.3e+33	5.09	8.49

Table 4.2.: Overview of the balancing result for the PFK reaction. The newly generated equilibrium constant and other dependent values are unrealistically high. A solution to this problem is offered in Chapter 4.2.

4.1.3. Check of the Thermodynamic Dependencies of the Balancing Results for PFK Reaction

It is of big importance to check, whether the results of the parameter balancing are thermodynamically feasible in reference to the relationship of the equilibrium constants and the Gibbs free energies of formation (see Chapter 2.6.2). It denotes

$$\begin{aligned} \ln k_l^{\text{eq}} &= -\sum_i n_{il} G_i^{(0)}/RT \\ \ln(9.0e+16) &= -(-1061.63\text{kJ/mol} - 1695.03\text{kJ/mol} \\ &\quad +1101.83\text{kJ/mol} + 1643.38\text{kJ/mol})/\text{RT} \\ 4.59 &= 11.45\text{kJ/mol/RT}\checkmark, \end{aligned}$$

where R is Boltzmann's gas constant ($R \approx 8.314\text{J/(molK)}$), and T is the absolute temperature. The relationship is confirmed and it holds with the balanced values. The further dependencies that are described in the corresponding chapter have to be valid by construction, since they are realized by the dependence matrix R_θ^x. Since I am testing the validity of the approach, the check for these dependencies can be found in the Appendix A.4.2.

4.2. Estimation of the Missing Equilibrium Constant for the PFK Reaction by Using a Pseudo Constant

As visible in the output Table 4.2, the dependent system parameters are unnormally high in magnitude. I am introducing the approach of using pseudo equilibrium constants to cope with this problem: There exists a direct connection between the Gibbs free energies of formation of the substrates and products of a reaction and the equilibrium constant, which is derived from the second law of thermodynamics and can be seen in Equation 2.8. To lessen the dependence of the equilibrium constants on the Gibbs free energies of formation, pseudo constants are used. Such a pseudo constant holds the mean value 2980.96 and the standard deviation 8886110.0. These are values derived from knowledge about our database (for further information see Chapter 4.2), and their use should show a significant improvement in comparison with the equilibrium constants being unexceptionally dependent on the Gibbs free energies. In order to test the functionality of this approach, the parameter balancing process on the PFK reaction model is performed another time, now with a provided equilibrium constant. The results are shown in the Appendix Table A.5 and they refer only to those parameter types that we have encountered problems with in the first balancing. The results of the dependent parameters have significantly improved, the equilibrium constant now has the value 1837.77. Also the other dependent parameters have taken on more reasonable values.
It is still important to note that these ocurring problems are only due to the fact of completely missing equilibrium constants. If a user provides values for the equilibrium

constants, the balancing results seem very appropriate according to their numeric values. An example for a parameter balancing for the PFK reaction including a provided equilibrium constant can be found in the Appendix A.4.4.

4.2.1. Balancing With pH Value and Temperature Regression

An important part of my work is the regression of different pH values and temperatures according to Chapter 2.6.4. Such a regression can be performed for the Gibbs free energies of formation and the equilibrium constants, but it is necessary that the user's SBtab input file is providing the pH values and temperatures for the parameters. The input data for this analysis is mainly taken from the Brenda web resource (Schomburg, Chang, Hofmann, Ebeling, Ehrentreich & Schomburg, 2002), the value for the equilibrium constants originate from Nissler et al. (Nissler, Otto, Schellenberger & Hofmann, 1983). The important parts of the input data are shown in the Appendix A.4.5. Apart from the need for provided temperature and pH values, the user has to choose favoured values for these parameters. In this example, I set the desired values to a temperature of 300 K and a pH value of 7. The balanced parameters referring to the input dataset are given in Table 4.3. According to these circumstances and under the influence of the Gibbs free energies, the equilibrium constant for the PFK reaction evolves from a mean input value of 0.013 to a value of 0.0004, with a corresponding standard deviation of 0.003.

QuantityType	Reaction	Compound	Value	Std	Unit
Gibbs free energy of formation	-	F16P	-1686.06	9.05	kJ/mol
Gibbs free energy of formation	-	ATP	-1652.35	9.05	kJ/mol
Gibbs free energy of formation	-	F6P	-1110.81	9.05	kJ/mol
Gibbs free energy of formation	-	ADP	-1052.66	9.05	kJ/mol
...					
equilibrium constant	PFK-reaction	-	**0.0004**	**0.0003**	-
...					

Table 4.3.: Parameter estimation including pH value and temperature regression. The desired pH value is 7, the desired temperature is 300 K. The standard biochemical potentials have all been set to a pH value of 5 and to a temperature of 360 K.

If the same input data is taken and the favoured circumstances are varied to 310 K and a pH value of 8, the balancing results (shown in Appendix A.4.6) show the equilibrium constant at a value of 0.0002, and the corresponding standard deviation at 0.001. As visible in Appendix A.8, the Gibbs free energies are differing only slightly from the formerly balanced values. The equilibrium constant that is provided with different pH values and temperatures (taken from (Nissler, Otto, Schellenberger & Hofmann, 1983)) is sensitive referring to the desired temperature, and behaves differently at different target values.

4.2.2. Balancing with Limited Input

The input data that is provided by the user for the underlying model is most likely incomplete. How are the values behaving when the majority of the parameter types is omitted? Given the same input table as for the first balancing (see Table A.4), all the Michaelis constants, inhibitory constants, and concentrations are removed. Instead, I append an equilibrium constant (mean value = 0.53, std = 0.17) to the input data set. In fact, the input data is reduced to only equilibrium constants and Gibbs free energies of formation. The balanced result is shown in the Appendix A.4.7. The choice of these two parameter types is due to the dependence Table 2.9, which shows, they have the biggest influence on the other kinetic parameter types. Furthermore, the results show that the balanced Gibbs free energies of formation are not differing from the formerly balanced ones. The newly generated values for Michaelis constants, concentrations, turnover rates, and the geometric mean rate constant are very close to their prior counterpart.
Being provided with the Gibbs free energies of formation and the values for the equilibrium constant, it is possible to estimate the missing parameters. The dependend parameters are generated in dependence to the Gibbs free energies, while the system parameters are resembling the corresponding mean values of the prior distribution.

The PFK reaction has served me as a very detailed example for the analysis of the parameter balancing process. I have tested the simple balancing process, the impact of pH value and temperature regression, as well as the use of pseudo constants, and the limitation of the input data to several parameter types (in this case Gibbs free energies of formation and equilibrium constants). In the following chapters, I will examine larger scale models.

4.3. pH Value and Temperature Regression on the Citric Acid Cycle

The model of the citric acid cycle (Nazaret, Heiske, Thurley & Mazat, 2009) will provide an example of the pH value and temperature regression under strongly differing conditions. Details on the amount of species, reactions, parameters, and available data are shown in Appendix A.3.1. The MetNetDB provides data on two of the equilibrium constants of the model (21 values for reaction v3, 11 values for reaction v6). Furthermore, it holds 60 Gibbs free energies of formation, which is important to denote, since the balancing results of equilibrium constants and Gibbs free energies of formation are so closely related. The first balancing is performed without a regression. After that I set the input data to different fictitious temperatures and pH values to show the differing results: the values evolve due to the differences in measuring circumstance. The favoured pH value is set to 7, the favoured temperature to 300 K. Figure 4.2 shows the results of the different balancings.

The balancing results get bigger, the more the temperature and pH value differ from the desired values (temperature: 300 K, pH: 7). The values become very large due to the standard deviations of the Gibbs free energies of formation.

Figure 4.2.: Temperature and pH value regression in the citric acid cycle: The blue bars are representing the equilibrium constant of reaction 3, the red bars represent reaction 6. On the left hand side of the black bar, the input data and the balanced result without any regression is shown. The right hand side of the black bar shows the balanced results under the circumstance of different input temperatures and pH values: the further the pH values and temperatures are moving away from the desired values, the higher the balanced results become.

4.3.1. Repairing Thermodynamical Dependencies via Parameter Balancing

The kinetic parameter set for a model that is provided by the user does not automatically hold valid for the thermodynamic dependencies that are shown in Chapter 2.6.2. I want to check, whether a kinetic parameter input set for the citric acid cycle that does **not** hold the relationship of equilibrium constants and Gibbs free energies of formation (see Equation 2.8), can be balanced by the parameter balancing approach, so that it does. The focus lies on the reaction v6, a 2-oxoglutarate aminotransferase, that converts alpha-ketoglutarat (KG) to oxaloacetate (OAA). The inconsistent parameter set, derived from NIST and Alberty, is shown in Table 4.4

QuantityType	Reaction	Compound	Value	Std	Unit
Gibbs free energy of formation	-	OAA	-1691.82	8.97	kJ/mol
Gibbs free energy of formation	-	KG	-1646.59	8.97	kJ/mol
...					
equilibrium constant	v6	-	0.17	1.21	
...					

Table 4.4.: A thermodynamically inconsistent kinetic parameter set due to the relationship of Gibbs free energies and equilibrium constants (see Equation 2.6.2).

The relationship of Gibbs free energies and the equilibrium constants for this parameter set can be expressed as

$$\begin{aligned} \ln k_l^{\text{eq}} &= -\sum_i n_{il} G_i^{(0)}/RT \\ \ln(0.17) &= -(-733.403\text{kJ/mol} + 667.03\text{kJ/mol})/\text{RT} \\ -3.72 &\neq 21.004. \end{aligned} \tag{4.2}$$

After balancing this kinetic parameter set, the parameter values are as shown in Table 4.5.

QuantityType	Reaction	Compound	Value	Std	Unit
Gibbs free energy of formation	-	OAA	-330.656	3.94	kJ/mol
Gibbs free energy of formation	-	KG	-383.06	4.53	kJ/mol
...					
equilibrium constant	PFK-reaction	-	2.2e+09	2.97e+09	
...					

Table 4.5.: The balanced output for the kinetic parameter set holds the relationship of Gibbs free energies and equilibrium constants.

$$\begin{aligned} \ln k_l^{\mathrm{eq}} &= -\sum_i n_{il} G_i^{(0)}/RT \\ \ln(99.88) &= -(-1693.32 + 1645.09 + 1103.54 - 1059.92)/RT \\ 1.847 &= 1.847\checkmark \end{aligned} \tag{4.3}$$

The parameter balancing has made the parameter set valid concerning the discussed dependence. The Gibbs free energies of formation as well as the equilibrium constant have changed their values significantly, which is probably due to the standard deviation of the energies.

4.4. Teusink Glycolysis Model

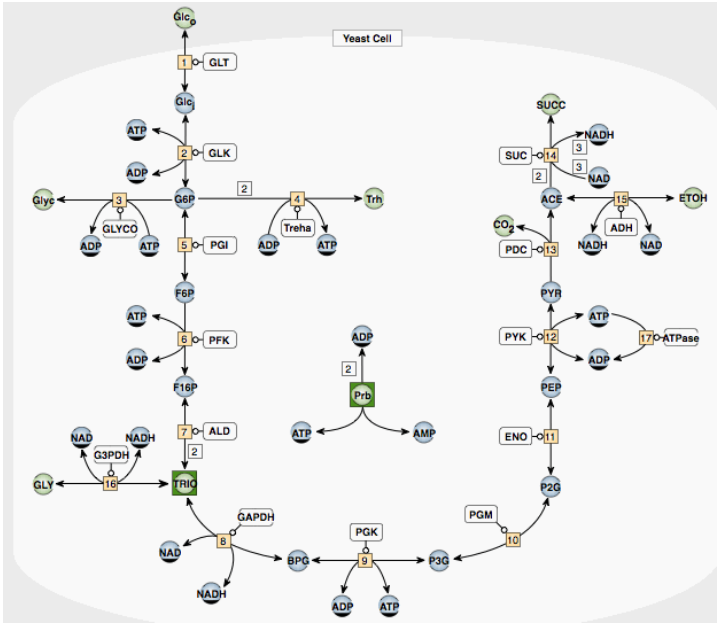

Figure 4.3.: The Teusink glyolysis model (Teusink, Passarge, Reijenga, Esgalhado, van der Weijden, Schepper, Walsh, Bakker, van Dam, Westerhoff & others, 2000). Its visualization is derived from JWS online (Snoep & Olivier, 2002).

The next model under examination is the glycolysis model of Teusink (Teusink, Passarge, Reijenga, Esgalhado, van der Weijden, Schepper, Walsh, Bakker, van Dam, Westerhoff & others, 2000) (see Figure 4.3), which also contains the PFK reaction already analyzed. Again, information on the model and the available model data is shown in Appendix A.3.1. The averaged input values in comparison to their balanced results are shown in the Appendices A.10, A.11 and A.12.

As visible in these figures, the parameter values are grouped in reference to their parameter types. An average of the types is calculated and shown as histograms. Due to several outlier values in the balanced output data, the average of the parameter groups is not a good foundation for comparisons. Instead, the median of the value distributions will be used. The values for the averages and the corresponding median are available for detailed comparisons in the Appendix A.3.1. Their medians are showing reasonable balancing results: newly generated parameter types are settling to magnitudes that are similar to the original model data (see Chapter 4.6).
Since the dependent model parameter types (equilibrium constants, turnover rates, maximal velocities, reaction affinities) are of high magnitude, they are represented as the averages of their logarithms and the corresponding median. Again, these values are unnormally high due to the standard deviations of the Gibbs free energies of formation, which will be proved and coped with in Chapter 4.5.

4.5. Chassagnole Threonine Model

Finally, also the information on the threonine model of Chassagnole (Chassagnole, Rais, Quentin, Fell & Mazat, 2001) (see Figure 4.4) are given in Appendix A.3.1. The comparison of the averaged input data and the averaged balanced data is visualized in the Appendices A.13, A.14, and A.15.

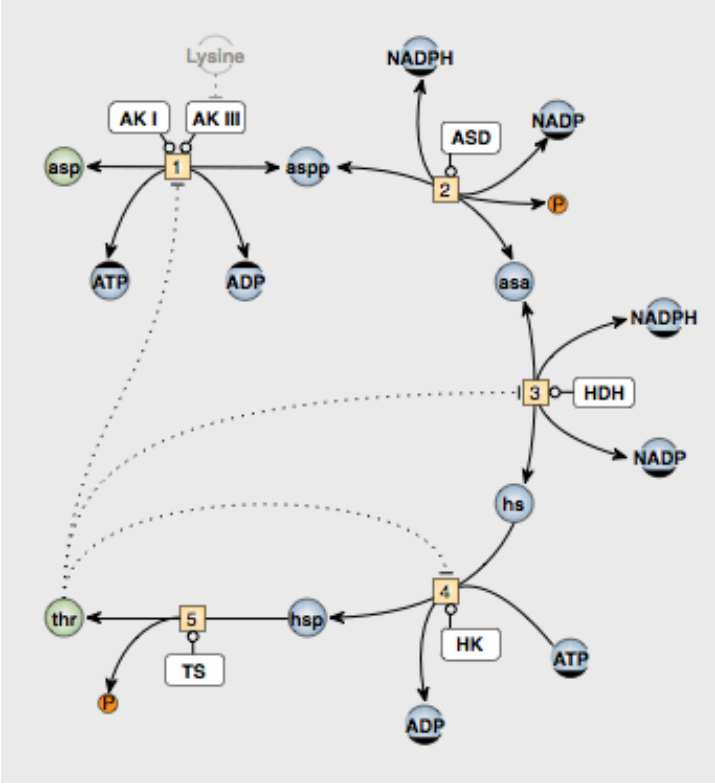

Figure 4.4.: The Chassagnole threonine model (Snoep & Olivier, 2002). Its visualization is derived from JWS online (Snoep & Olivier, 2002).

The balanced data is very similar to the one for the Teusink glycolysis model. Again, the dependent model parameters are reaching unnormally high values, and are visualized in the averages of their logarithms. A far better comparison can be achieved by focussing on the medians. Nevertheless, I want to examine the presence of high magnitudes of the dependent parameter types. As formerly stated, it is likely that these values are unusually high due to the standard deviations of the Gibbs free energies. Since the dependent parameters are easy to influence by these standard deviations, they can reach the observed value levels. For the examination of this assumption, I am setting the standard deviations of the Gibbs free energies of formation to a small value (0.1) and balance the input data set again. As Figure 4.5 shows, the high values of the dependent parameters have settled down to more appropriate values, so the assumption is proven (the figure shows the actual average values of the parameter types, the average of their logarithms is no longer needed for visualization).

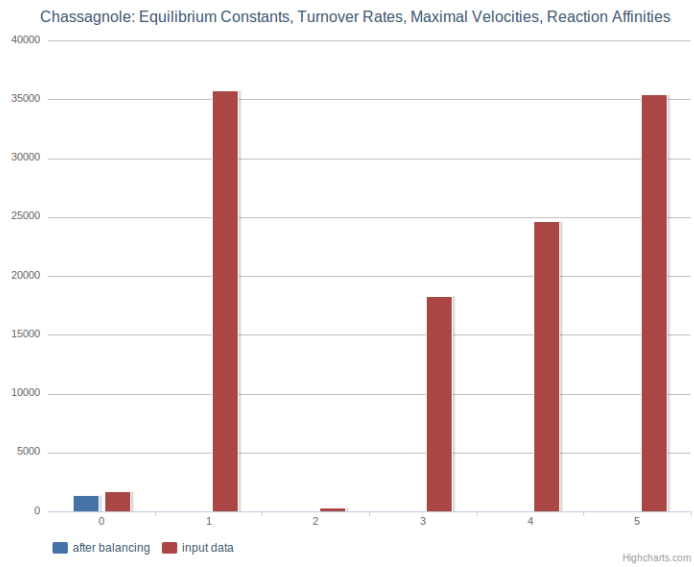

Figure 4.5.: Chassagnole: Shown are the actual average values of average equilibrium constants, 1 = Average turnover rate forward, 2 = Average turnover rate backward, 3 = Average maximal velocity forward, 4 = Average maximal velocity backward, 5 = Average reaction affinity. Still, the turnover rate forward and the reaction affinity are quite high, but compared to the former results and keeping in mind that this figure shows actual values and no longer the means of the logarithmic ones, show the improvement that occured by setting small standard deviations instead of using the default values (which is no solution to the problem, but it demasks the origin of my former problematic results).

4.6. Comparing Balancing Results to Literature Results

The balancing results generated can only be proven as valuable by comparisons to original values from the literature. For comparing my balanced parameters with literature data, I have used the web resource of Sabio-RK (Wittig, Golebiewski, Kania, Krebs, Mir, Weidemann, Anstein, Saric & Rojas, 2006) and the original model values from the underlying publications of the models. Sometimes, not only the consideration of the median is interesting, but also the average value of the input, output, and original values. The following chapters are comparing these numbers, and the itemized view of the comparisons can also be found in Appendix A.6.1.

4.6.1. Comparison of Results for the PFK Reaction

The results generated with the balancing approach are shown in Table 4.2. In the web resource Sabio-RK (Wittig, Golebiewski, Kania, Krebs, Mir, Weidemann, Anstein, Saric & Rojas, 2006) the following kinetic parameter values are available: the **concentration** of the species fructose-1,6-bisphosphate (start value: 0 mM, end value: 5 mM), and the **Michaelis constant** for fructose-1,6-bisphosphate (19.2 mM).
The results produced via the parameter balancing are partially similar to those of the Sabio-RK. The concentration for fructose-1,6-bisphosphate was estimated to a value of 0.93 mM (with a standard deviation of 1.79) which corresponds to the Sabio-RK value in a range of 0-5 mM. The Michaelis constant for the same species has an estimated value of 2.18 (with a standard deviation of 7.46). This is deviating from the Sabio value (19.2 mM), but due to the fact that I have had no input value for this species and the Michaelis constant was estimated with a high influence of the prior value for Michaelis constants, the estimation is decent. The itemized information on this comparison is found in Appendix A.6.1.

4.6.2. Comparison of Results for the Glycolysis Model

The available data from the original Teusink glycolysis model (Teusink, Passarge, Reijenga, Esgalhado, van der Weijden, Schepper, Walsh, Bakker, van Dam, Westerhoff & others, 2000) are shown in Appendix A.3.1. The balanced values produced by the parameter balancing approach can be seen in Appendices A.11 and A.12.
The **Michaelis constants** that have served as input for the parameter balancing, are higher in median (2.2 mM) than those declared in the data of Teusink. The balanced values have decreased in their median and come close to the Teusink median (0.3 mM). Still, the balanced values are higher than the ones we declare as reliable. This might be due to the fact that the used prior value (which was used for the generation of numerous Michaelis constants of the model) is higher (15.5 mM) than the values in this model.
The median of input values for **inhibitory constants** (1.6 mM), the median of the balanced values (1.8 mM), and the Teusink model's median (1 mM) move in a short range and prove the reliability of the data.
The values for **equilibrium constants** and **maximal velocities** are visualized in logarithmic form, so I have taken the means of the logarithms of the Teusink data to compare the two. The input values (average 6, median 0.2) are surprisingly similar to

the Teusink values (average 7.48, median 1.1). The balanced values (average 92, median -0.5) suffer from the few outliers, but the median is still in a reasonable range to the one of Teusink. The maximal velocities that without exception have been balanced without any input value, suffer from the outliers as well, but the median of the balanced results (5) are located in the close neighborhood of the Teusink values median (5.48).
The itemized information on this comparison is found in Appendix A.6.1.

4.6.3. Comparison of Results for the Threonine Model

The available data from the original Chassagnole threonine model (Chassagnole, Rais, Quentin, Fell & Mazat, 2001) are given in Appendix A.3.1. The parameter balancing for this model has produced values listed in Figures A.14 and A.15.
Different from the Teusink model, the available data for the Chassagnole model also holds **concentrations** of species. The average of these original values (2.49 mM) is much lower than the input data used for the parameter balancing (49 mM). The balancing process has lowered the average value to 21 mM and the corresponding median as well (from 2.1 mM to 1.9 mM). Thus, not only the average value has been approached to the Chassagnole data's average, also the median is very close to the one from Chassagnole. An improvement from input to output data is clearly visible.
The input **Michaelis constants** are, just like the concentrations, much higher in average than the corresponding original data. Nevertheless, the input data's median (1.5 mM) is close to the one of the original data of Chassagnole (0.22 mM), and by balancing it even gets closer (0.9 mM). Also the average value is lowered (from 5 mM to 1 mM) in magnitude and approaches the average value of Chassagnole (0.87 mM).
For the **inhibition constants**, Chassagnoles data has an average value of 1.92 mM and a median of 0.39 mM. While the average value of the input data (8.2 mM) is increased by the balancing process (to 15.2 mM), the median is still staying very low (input data: 0.5, output data: 0.3). The increase in magnitude of the average value might have arisen through the generation of new inhibition constants that were not provided by the input data (the prior mean for inhibition constants is 11.9, the corresponding standard deviation 173).
While the input values for the means of the logarithms of **equilibrium constants** is quite low (7.1) compared to the value from Chassagnole (10.62), the balanced value rises to 15.2. The median of the original data (0.09) lies between the input median (-0.6) and the balanced one (0.9).
The very low **maximal velocities** that are provided by Chassagnoles original data (mean of logarithms average -2.21 and median -2.3) cannot be foreseen by the balancing process. Without any input data for this parameter type, the calculation estimates are higher than the ones from Chassagnole (mean of logarithms average 11.6 and median 4.2). The assumption is that for the sensible estimation of the maximal velocity more input data are needed.
Apart from the estimation of maximal velocities in the Chassagnole model the results of the parameter balancing are very similar to the data derived from Teusink, Chassagnole, and Sabio-RK. In most cases, the balancing could produce reasonable results, sometimes even despite bad input data. The itemized information on this comparison is found in Appendix A.6.1.

5. Discussion

The results I was able to obtain with the parameter balancing did not match all of my expectations in the first place, but I was able to add some regulatory instances in order to prevent errors and to decrease the instability of the estimation caused by missing values. Estimates concerning the system parameters of the Gibbs free energies of formation, Michaelis constants, inhibitory constants, and species concentrations produced sensible results in comparison with the original data derived from the underlying publications of the models. Estimates of dependent parameters, such as equilibrium constans, turnover rates, maximal velocities, and reaction affinities turned out to be very dependent on the input parameters and the standard deviation of the Gibbs free energies of formation. Due to this fact, the results for these parameters often include several outliers and reach unusually high values. This problem, nevertheless, could be controlled by the introduction of pseudo equilibrium constants as an equivalent to a better prior standard deviation of Gibbs free energies. Furthermore, the comparison of the balanced values and those taken from the original model data was satisfactory, and the check for thermodynamical feasibility of the balancing results confirmed their validity. Following to these general statements, I want to go further into detail.

5.1. The Prior Distribution

The prior distribution, in this case derived from our MetNetDB, is the key to a successful estimation process. Due to missing or incomplete prior values I have encountered several problems during my work. A drawback is the complete lack of specific parameter types: our database does not contain values for activation constants or geometric mean rate constants. The result of this incompleteness is that the estimated values are even more depending on the input data. If the input data does not provide activation constants I simply do not have the ability to produce them. Values for reaction affinities are not available as well, but I am able to estimate them via their dependence concerning available entries (the Gibbs free energies and the concentrations). Of course, this has the side effect that the reaction affinities are exclusively dependent on the (possibly problematic) input data and cannot at least rely on any kind of values from the prior distribution.
The prior distribution I constructed, has turned out to be the foundation to reliable estimates concerning the data it was generated from. But the limitations of the distribution (only few equilibrium constants and high standard deviations) have in turn also lead to problems in the estimation (outliers in the distribution of dependent parameters). A direct consequence of the missing standard deviations is discussed in the following.

The Problems and Benefits of Dependencies The strong connection between Gibbs free energies and equilibrium constants has been mentioned multiple times. While I was able to perform perfectly appropriate estimates for the first, the latter confronted me with problems. Since the MetNetDB contains only few standard deviations for Gibbs energies and equilibrium constants (and the available ones are often of high magnitude), I had to use the default prior standard deviation for those two parameter types. Both of these parameter types can vary strongly depending on the underlying models and the circumstances they are measured in. This leads to a high standard deviation in the prior distribution. This circumstance in combination with the strong dependence of equilibrium constants on Gibbs free energies of formation confronted me with unusually high equilibrium constant values. This was carrying on the misestimate by influencing the turnover rates, the maximal velocities, and the reaction affinities. Tests I performed by altering the input data, showed that reasonable standard deviations for Gibbs free energies and equilibrium constants are able to prevent this scenario.
Furthermore, I was able to constrain the dependence of equilibrium constants on Gibbs free energies by introducing pseudo equilibrium constants: in case of missing equilibrium constants I am automatically appending generated constants to the data set that are of appropriate value, and far better than the alternative: the very high dependence on the energies of formation and their lack of availability of many standard deviations, which turned out to be a major obstacle to the generation of equilibrium constants.

5.2. Technical Difficulties

5.2.1. Problems Concerning SBML

Next to the problems concerning the provision of input data, I also encountered several technical difficulties. At first, I had to overcome programming difficulties concerning SBML. Unfortunately, SBML did not hold all the possibilities that I would have liked to have for my programming challenges. It was hard to gain specific knowledge about model elements: SBML does not make a difference between a species in general or an enzyme in specific. This problem would be best solved by introducing a "ListOfEnzymes" into the list structure of SBML. The only way to identify an enzyme was if the SBtab file held a "concentration of enzyme" for the species or if there had been added an SBO term to the species that might have indicated the true identity. Such a proposed "ListOfEnzymes" must not exclude its enlisted elements from the "ListOfSpecies", since often the differentiation between the two is difficult: An enzyme is always a species, but not vice versa. So the "ListOfEnzymes" might hold every species that has the potential to act as an enzyme (and is also represented in the "ListOfSpecies").
A very similar problem is the identification of species that are inhibitors or activators. In SBML, they are only marked as being a "modifying" species, which is just not enough information. This too was only evitable by attaching an SBO term to the species identifying it as an inhibitor or activator. If my tool encounters species that are modifier but cannot be identified as either inhibitor or activator, I have to treat it as a normal species, in order to prevent misestimates. Nevertheless, I have added the ability of the tool to identify the species clearly as soon as a SBO term is attached to them.

5.2.2. Insertion of Parameter Files into SBML

Another obstacle was the restrictivity of the data integration of the balanced values into the SBML file. As it becomes obvious in Chapter 3, the use of my tool is demanding to the user. This is due to the fact that very specific criteria have to be fulfilled for the input data, if the user wants to integrate it into the SBML file. In order to match all criteria I had to automatically generate a second SBtab file out of the given values in combination with new blank values. A detailed documentation for users will be very appropriate, so the complex steps that lead to a successful data integration can be understood clearly.

5.2.3. The Unit Problem

The entries for enzyme concentrations that I obtained from the MetNetDB are often inappropriate, since they are differing in their units. Values with different units must not be mixed together, so I had to sort out the entries with the unit "molecules per cell" and only used those in "mM" (millimolar).

5.3. A Conclusion: Cooperation of Experimentalists and Modelers

The problems during the testing of my parameter balancing approach mostly arose from incomplete or erraneous kinetic data. Although the fact that a close cooperation between experimentalists and modelers is absolutely crucial not only for my work, but in general, is not new, I still want to point out its significance. The results have shown that the success of parameter estimations is very dependent on the input data. Without reasonable data it is not possible to construct useful results. The exchange of information between experimentalists and analysts should have the highest priority and can be supported by using standard data formats like SBML for models and SBtab for the provision of measured kinetic data.

5.4. Summary and Future

5.4.1. Summary

The early encounter of problems with missing equilibrium constants and high standard deviations have made me introduce the use of pseudo equilibrium constans. This solution was very useful in retrospective, since the results for the dependent parameters included better and more appropriate estimates than before. The use of pseudo equilibrium constants is now set as a default in my developed tool.
Furthermore, I have extended the former balancing approach (Liebermeister & Klipp, 2006) by adding a functionable temperature and pH value regression approach. The

tests of this function showed the formerly estimated results: if the input data is provided with temperature and pH values highly differing from the desired ones, this fact is taken into account and reflected in the balancing results. Furthermore, the amount of parameter types to be estimated has increased by the complementation of reaction affinities and Gibbs free energies in a dependence to provided concentration values.
The comparison of the balancing results to data taken from the original model publications revealed the current powers as well as small weaknesses of the parameter balancing approach. It will be of great importance to improve the tool and dispose the shortcomings.
I have implemented an extension of the graphical user interface of semanticSBML (Krause, Uhlendorf, Lubitz, Schulz, Klipp & Liebermeister, 2010) that enables the user to perform the balancing process, including several useful options. Finally, the integration of the resulting kinetic data into the corresponding SBML file can be performed.

5.4.2. Outlook

Since the whole process of data provision, parameter balancing, and data integration is not intuitive and I do not want the users of my tool to be completely depending on an installation of the semanticSBML GUI, I plan to implement a web application. This allows me to make the tool more user-friendly and easier to understand, which is crucial for the complex process of parameter balancing. An important novelty will be the specification of organisms as input data. While MetNetDB exports the values as a mix of different organisms, my web application is going to offer the user a choice of favoured organisms and either collect only this organisms values, or collect all values and set the standard deviation of unrelated organisms to a higher default value. An improvement of the prior distribution concerning the availability of more standard variations is absolutely eligible, and I plan to implement it. With these analysis improvements, an even more appropriate parameter balancing approach can be achieved.

A. Appendix

A.1. Example Input File for Small Network

QuantityType	Reaction	Compound	Value	pH	Temperature
Gibbs free energy of formation	Reaction 1	Species 1	-100	4	290
Gibbs free energy of formation	Reaction 1	Species 1	-400	5	289
Gibbs free energy of formation	-	Species 1	-220	4	290
Gibbs free energy of formation	Reaction 2	Species 3	-310	5	295
equilibrium constant	Reaction 1	-	1.11	3	301
equilibrium constant	Reaction 1	-	2.58	3	305
inhibitory constant	Reaction 1	Species 3	4.11	4	299
inhibitory constant	Reaction 1	Species 3	1.98	5	294
Michaelis constant	Reaction 1	Species 1	2.84	6	283
Michaelis constant	Reaction 1	Species 1	2.24	5	294
Michaelis constant	Reaction 1	Species 2	8.26	6	311
concentration	-	Species 2	10.22	4	301
turnover rate forward	Reaction 2	-	4.76	3	300

Table A.1.: The example SBtab file shows some duplicate parameter values for several parameter types (e.g. two values for the equilibrium constant of Reaction 1), other kinetic parameters are not provided at all (e.g. Gibbs free energy of formation for Species 2).

A.2. Manual for SemanticSBML Parameter Balancing

I have implemented a user-friendly interface for working on parameter estimation and data integration by extending the existing Systems Biology tool semanticSBML (see Chapter 2.4). Apart from the available functions of SBMLfill I added the possibility to access, modify, and integrate kinetic data into an SBML model. The automated process of parameter balancing is part of this implemenetation as well. The start screen of semanticSBML now offers a new possibility: Fill - which opens a new tab (see Figure A.1).

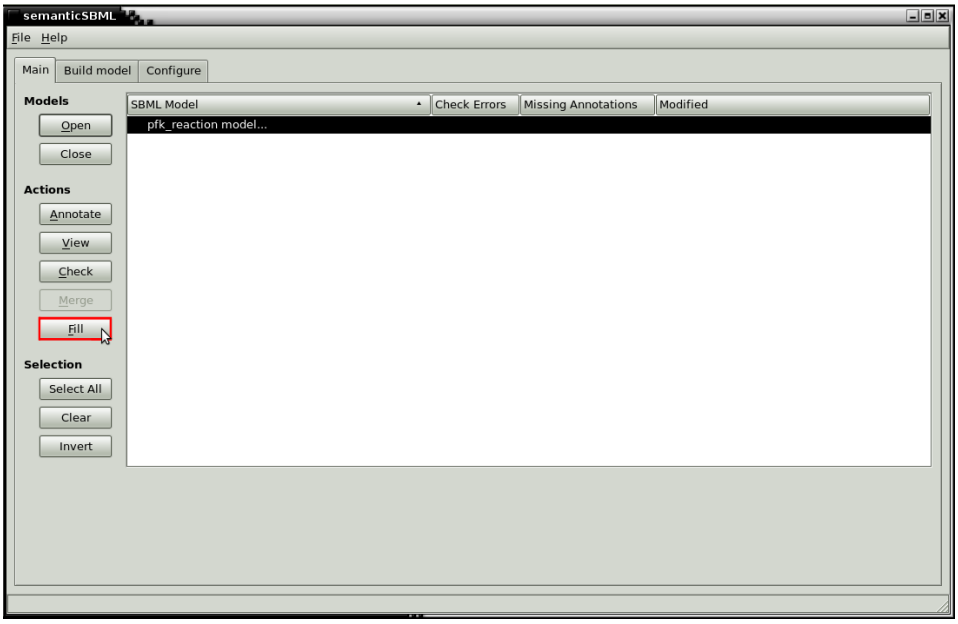

Figure A.1.: The start screen of semanticSBML is extended by a new item: "Fill" opens a new tab for the task of handling model parameters.

The insertion of model parameters including parameter balancing via the semanticSBML user interface can be performed in five steps:

1. Open an SBML model and use the Fill option in semanticSBML (explained in Chapter A.2.1)
2. Open an SBtab parameter file that corresponds to the model (explained in Chapter A.2.2)
3. Automatically generate a new SBtab that matches all SBML insertion criteria (explained in Chapter A.2.3)
4. Start parameter balancing for the parameter set (explained in Chapter A.2.5)
5. Insert the kinetic rate laws and the balanced parameter set into the SBML model (explained in Chapter A.2.6)

In the following, these steps will be exemplified in detail.

A.2.1. semanticSBML Fill - Overview and Insertion

By clicking the newly introduced Fill button the user is guided to a new tab that allows the upload, modification, and integration of model parameters (see Figure A.2).

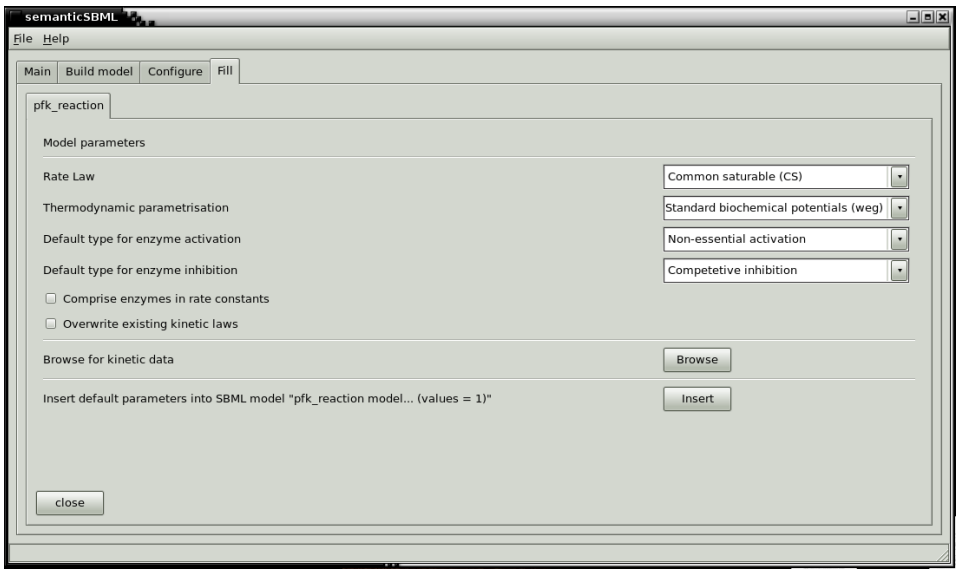

Figure A.2.: Control menu of model parameter handling: the headbase of the new Fill tab.

The possibility to insert a kinetic rate law into the currently loaded SBML model is accompanied by a set of choices the user has to make. The choices are similar to those of SBMLfill:

Rate Law The user can choose the rate law to be inserted into the SBML model (common modular rate law, simultaneous binding modular rate law, direct binding modular rate law, power-law modular rate law, and force-dependent modular rate law). For further details on the rate laws see Chapter 2.2.

Thermodynamic Parametrisation The user chooses between Gibbs free energies of formation, equilibrium constants, and catalytic rate constants for the inclusion into the kinetics of the model. For further details see Chapter 2.2.

Default type for enzyme inhibtion Complete, partial, or specific. In the complete and partial regulation mechanisms, regulators bind independently of the reactant and influence the conversion step. In specific inhibition, we assume that inhibitor binding prevents the binding of any other reactants.

Default type for enzyme activation Complete, partial, or specific. Specific activation works analogously to inhibition: the activator binding is essential for the binding of any other reactant, so there is just one more non-activated state, which again contributes a denominator term.

Comprise enzyme concentrations in rate constants Yes or no.

Overwrite existing kinetic laws Yes or no.

To this point the SBMLfill interface has been taken as role model and been integrated into the semanticSBML user interface. I added the following options, which are beyond the possibilities that SBMLfill offers. The user now can to open a kinetic parameter set file in SBtab format that corresponds to the opened model. If there is no such file at hand, the user still has the possibility to insert a kinetic rate law into the model with the given options and all parameter values set to a default value of 1. Otherwise, if the user opens a parameter set file from the hard drive, a new subtab is opened, showing the opened SBtab parameter file (see Figure A.3).

A.2.2. semanticSBML Fill - SBtab File Provided by User

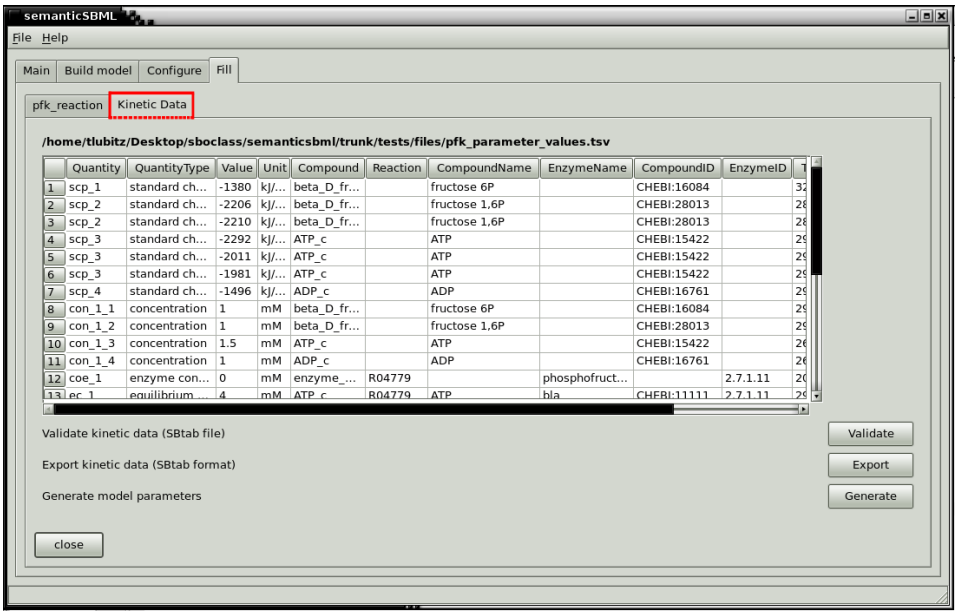

Figure A.3.: An opened SBtab file holding a parameter set for the currently loaded SBML model.

The SBtab file is holding an either complete or incomplete parameter set for the SBML model. It is crucial that the SBtab is of the type KineticData and by that fulfills the criteria of this type: the availability of the columns "QuantityType", "Reaction", "Compound", and "Value". Alternatively to the "Value" column, the "Mean" column can hold a value for the parameter. Another import column is "Std", which holds the standard deviation to the value. If this column criterion is not matched, the following editing, balancing, and modeling steps cannot be performed. The incompleteness of parameter values will not be considered a problem, as we will see soon.
The opened SBtab is shown in a table view that is editable: all values and names can be altered. This subtab also offers some actions that can be performed on the uploaded SBtab:

Validate The content of the uploaded SBtab file will undergo a general validity check. This check will fail if certain obvious criteria are not matched: the *Value* column needs to contain numerical values as well as the *Std* (standard deviation) and *Mean* column.

Export The SBtab file can be exported onto the hard drive in the current state. If any alterations have been done in the table view, these alterations will be saved.

Generate Generates model parameters. This button will open two new subtabs, one holding a new SBtab file that is fitted to be inserted into the SBML model. The necessity of having two SBtab files will be elucidated next.

A.2.3. semanticSBML Fill - New SBtab File With all Required Model Parameters

If the aim of the user is the insertion of the content of the SBtab file into the SBML model, one will activate the "Generate" button and open two new subtabs by that:

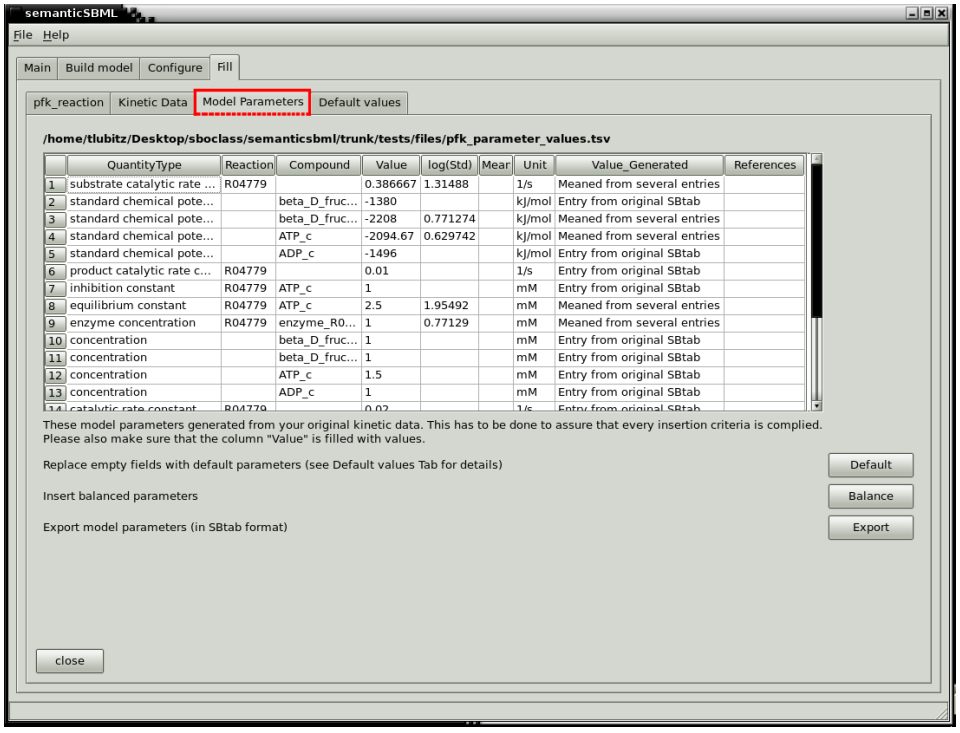

Figure A.4.: A new SBtab file that is generated from the first one and possibly extended in order to match the SBML insertion criteria.

A new subtab *Model Parameters* - holding a newly generated SBtab file - is opened and offers new options for the user. While the first SBtab file that is opened from the users hard drive does actually not underlie any format criteria (except for the presence of the four columns mentioned earlier), the insertion of the parameter file into the SBML model is restrictive. It demands the following information:

- 1 substrate catalytic rate constant for every reaction.
- 1 product catalytic rate constant for every reaction.
- 1 Michaelis constant for every species appearing in the reactions (if one species appears in two reactions, also two Michaelis constants will be needed).
- 1 velocity constant per reaction (geometric mean rate constant).
- 1 equilibrium constant per reaction.
- 1 Gibbs free energy of formation per species.
- 1 activation/inhibition constant per reaction.

If any of these parameters are not provided, the insertion into the SBML model cannot take place and will result in an error. Since it is quite improbable that the user will be providing a SBtab file holding all the required information, the program generates this new SBtab file that consists of the given values from the former SBtab file and moreover automatically adds all the possibly missing (but required) parameters in form of blank entries. The appearance of the new SBtab file is fixed (while of course the values and names can still be edited). Its columns are:

- QuantityType - the parameter type
- Reaction - SBML name of the reaction
- Compound - SBML name of the species
- Value - the value of the parameter
- Std - the standard deviation of the parameter
- Unit - the unit for the value measurement
- Value Generated - declares how the parameter has found its way to the new SBtab file. The options are:
  1. Entry from original SBtab
  2. Averaged from several entries - the original SBtab file held more than one value for this parameter, so the values have been meaned (see Chapter 2.6.3).
  3. Entry generated automatically - if the original SBtab did not provide this parameter although it is obligatory, it was generated automatically
- Reference - if the value is taken from the former SBtab file and has a given reference, this reference is available in this column

At this point it is important to note that any parameter taken from the users SBtab file that is not clearly assignable to an element of the corresponding SBML model will be ignored by the program. This means, if the user uploads a SBtab file holding a velocity constant for a specific reaction which is no part of the SBML model, this velocity constant will not appear in the new SBtab file. This action is necessary since parameter values for the wrong model might dislocate the whole parameter balancing step. Now that it is provided that all needed parameters are available (and obsolete ones are removed), suitable values for the automatically generated ones have to be found. The user can either set all missing parameter values to default values (see Chapter A.2.4) or start the parameter balancing (see Chapter A.2.5).

A.2.4. semanticSBML Fill - Setting Default Values

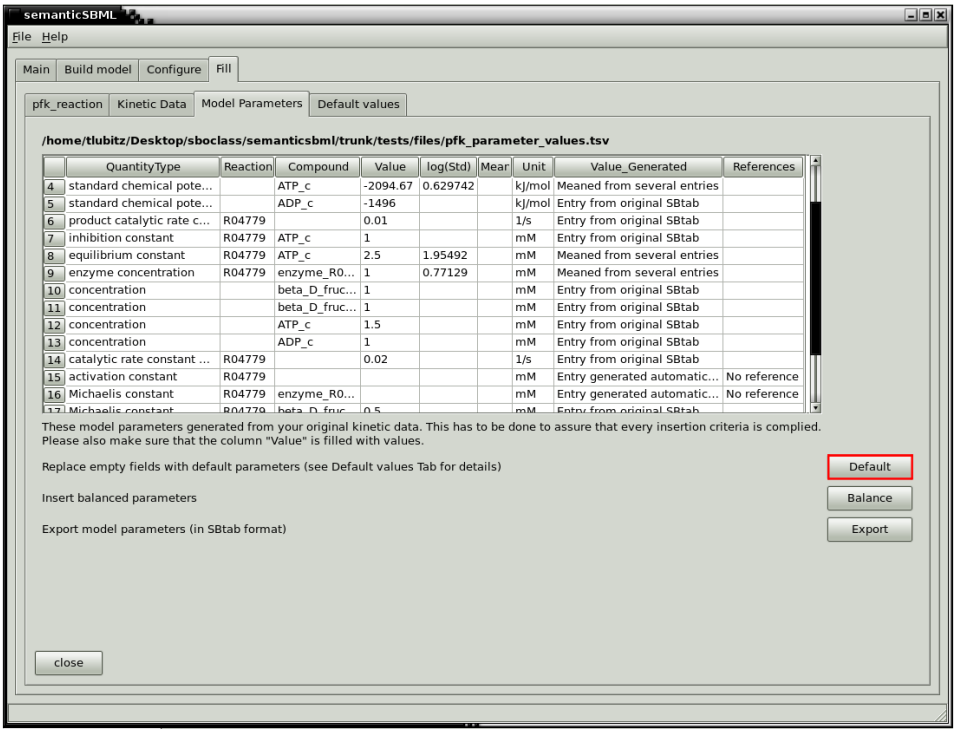

Figure A.5.: The user can set all missing parameter values to default values by hitting the “Default” button.

The “Default” button automatically inserts default values for every missing parameter value, which can be edited by the user (see Figure A.5).

The subtab “Default values” (see Figure A.6) holds a table with default values for all types of required parameters. Here, default values for the mean and for the standard deviation std(logvalue) of any parameter type are listed. They are referring to literature and our internal database MetNetDB, which provides a wide variety of parameter values, including those of KMedDB, Brenda (Barthelmes, Ebeling, Chang, Schomburg & Schomburg, 2007), and several more. Any of the default values can be edited if the user wants to do so. The rest of the columns and parameter choices in this subtab are

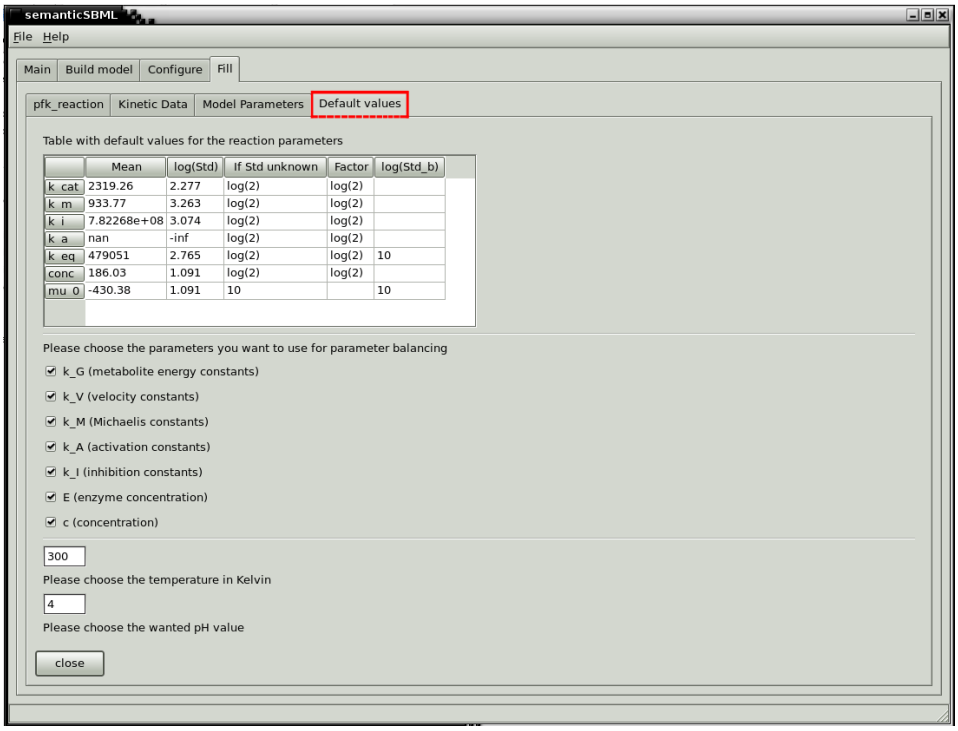

Figure A.6.: The subtab "Default values" offers the possibility to edit the default values and choose the parameters used for parameter balancing (see section 3.1.5).

not important for the setting of default values, but for the parameter balancing (see Chapter A.2.5).

A.2.5. semanticSBML Fill - Parameter Balancing

The parameter balancing is performed by hitting the "Balance" button in the subtab holding the new SBtab (see Figure A.7).

Just like the options for the default values, the options for parameter balancing can be edited in the "Default values" subtab (see Figure A.8).

Parameter entries of the SBtab file can provide a mean value, but in some cases they might lack a standard deviation. Since the standard deviation is crucial for the parameter balancing, it has to be substituted. The value that is set to the parameter with the missing standard deviation can be found in the table view of the "Default values" subtab under the column *If Std unknown*.
Furthermore, the "Default values" subtab offers the possibility to choose the parameter types used for balancing. As a default, all the types are selected and this scenario corresponds to the normal build-up of the vectors and matrices explained in Chapter 2.6.2. If any of the parameter types is unchosen in this subtab, this leads to a consequent omitting of this parameter type in all vectors and matrices. The type will not be taken into account for the parameter balancing.
The user can moreover choose a temperature (in Kelvin) and a pH value that is preferred for the reactions from the model. This incorporates a temperature and pH value

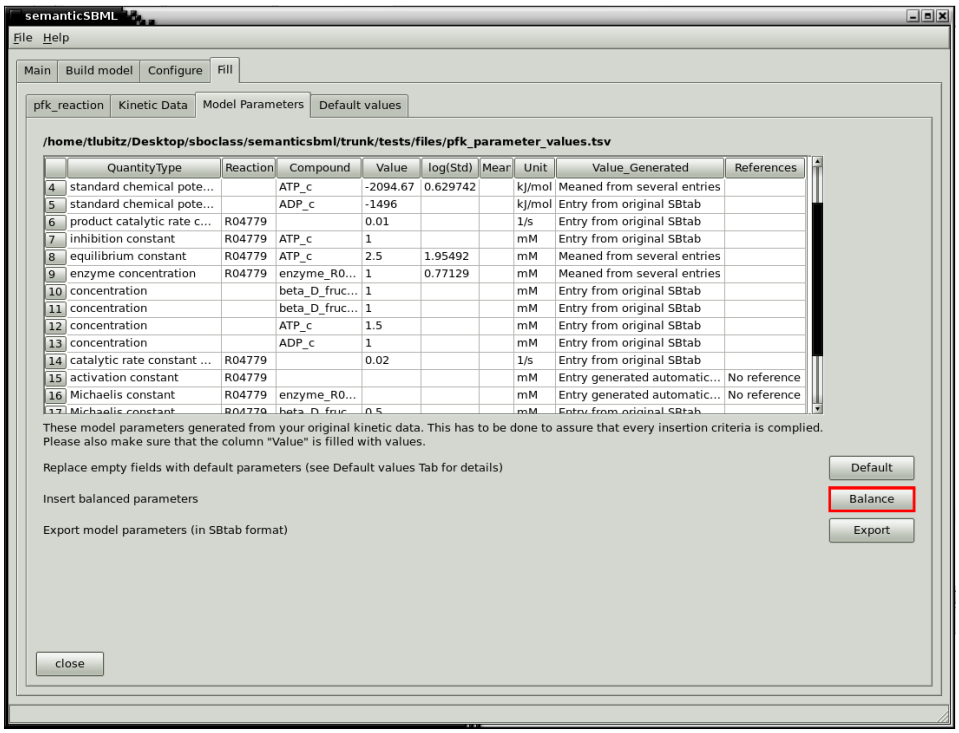

Figure A.7.: The Balance button starts the parameter balancing.

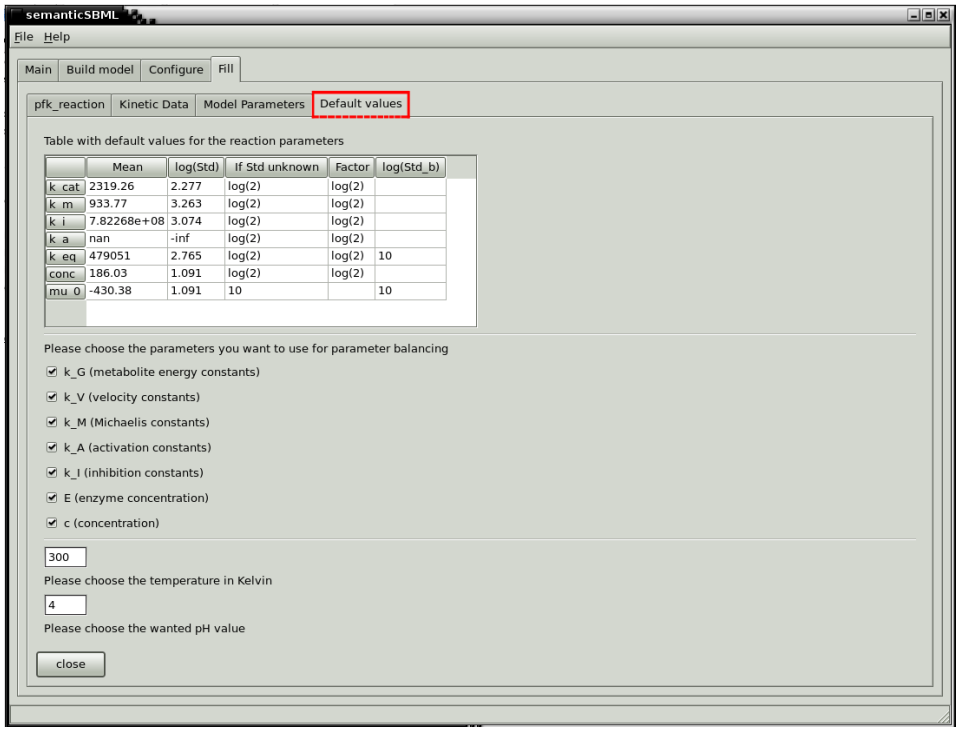

Figure A.8.: The subtab “Default values” also holds the options for parameter balancing.

regression into the parameter balancing: the given values from the SBtab are compared to the favoured values, resulting in Δ values that influence the balancing process (see Chapter 2.6.4). After the balancing of the parameters or setting of default values, the SBtab file is ready to be inserted into the SBML model.

A.2.6. Integration of Model Parameters Into an SBML Model

After having generated an insertable SBtab parameter file the user is able to integrate the content of the file into the SBML model. The user has to change back to the overview subtab of Fill (see Chapter A.2.1 and Figure A.9), where a new button for insertion is added to this tab as soon as an insertable SBtab parameter file has been generated.

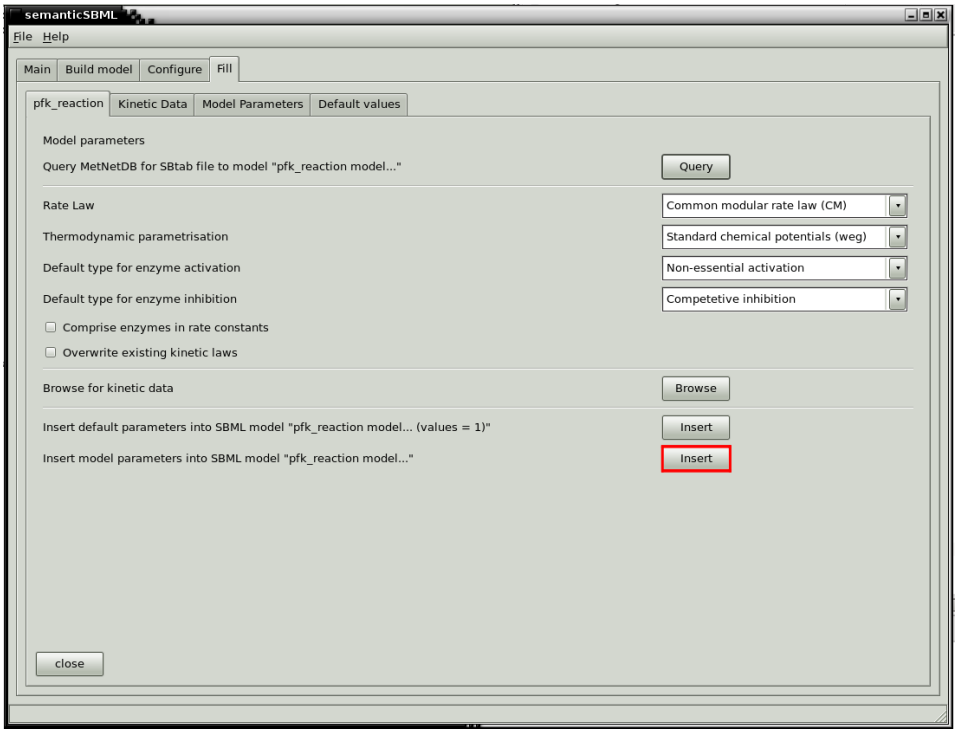

Figure A.9.: A new button has appeared in the Fill overview: insertion of generated model parameters.

A.3. Tables and Visualizations of the Parameter Balancings

A.3.1. Model Data Information

	PFK reaction	Citric acid cycle	Glycolysis model	Threonine model
Number of Reactions	1	12	17	7
Number of Species	4	14	26	11
MetNetDB: Gibbs free energies of formation	20	60	92	39
MetNetDB: Michaelis constants	14	1497	305	230
MetNetDB: Inhibitory constants	8	394	23	90
MetNetDB: Species concentrations	19	68	89	40
MetNetDB: Equilibrium constants	-	32	193	5
MetNetDB: Turnover rates	-	201	23	85
MetNetDB data: Number of Organisms	24	44	51	19
Sabio-RK: Michaelis constants	1			
Sabio-RK: Species concentration	1			
Teusink data: Equilibrium constants			10	
Teusink data: Maximal velocities			16	
Teusink data: Michaelis constants			37	
Teusink data: Inhibitory constants			4	
Chassagnole data: Equilibrium constants				7
Chassagnole data: Maximal velocities				5
Chassagnole data: Michaelis constants				23
Chassagnole data: Inhibitory constants				5
Chassagnole data: Species concentrations				9

Table A.2.: The model data information for the four used models. The table comprises information on the network (reactions and species), the quantitative information of values derived from the prior distribution of MetNetDB, and the availability of original values for the models, taken from the underlying publications, and the Sabio-RK.

A.4. Input Model Data for PFK Reaction

QuantityType	Reaction	Compound	Value	Std	Unit
Gibbs free energy of formation	-	F6P	-908.76	-	kJ/mol
...					
Gibbs free energy of formation	-	F16P	-909.60	-	kJ/mol
...					
Gibbs free energy of formation	-	ATP	-253.55	-	kJ/mol
...					
Gibbs free energy of formation	-	ADP	-1356.14	-	kJ/mol
Michaelis constant	pfk-reaction	ATP	0.009	0.001	mM
...					
Michaelis constant	pfk-reaction	ADP	0.34	-	mM
inhibitory constant	pfk-reaction	ATP	2.5	-	mM
...					
concentration	-	ATP	2.641	-	mM
...					
concentration	pfk-reaction	ADP	0.823	-	mM
...					
concentration of enzyme	pfk-reaction	-	33477	-	m/cell

Table A.3.: The initial kinetic parameter set in form of SBtab KineticData. There are several circumstances to focus on: not every single parameter is shown due to lack of space (I am provided with 5 standard chemical potential values for each species, 6 concentrations of enzyme for ATP, 10 concentration values for ATP and 9 for ADP, 43 inhibitory constants for ATP, 12 Michaelis constants for ATP and 2 for ADP). Since I have more than one measured value for many kinetic parameters, these values are averaged by the method introduced in 2.6.3. It is shown that some parameters are providing a standard deviation and some do not. While standard chemical potentials do not need to hold a reaction, Michaelis constants need a reference for species **and** for the reaction. Due to lack of space the fructose-6-phosphate is abbreviated by F6P, the fructose-1,6-bisphosphate is abbreviated F16P. The concentrations of enzyme for ATP are holding values of the unit "molecules per cell", which cannot be converted to mM by the program so far. These values have to be ignored.

A.4.1. List of Required Kinetic Parameters for the PFK Reaction

QuantityType	Reaction	Compound
Gibbs free energy of formation	-	fructose-6-phosphate
Gibbs free energy of formation	-	fructose-1,6-bisphosphate
Gibbs free energy of formation	-	ATP
Gibbs free energy of formation	-	ADP
catalytic rate constant geometric mean	PFK-reaction -	
Michaelis constant	PFK-reaction	fructose-6-phosphate
Michaelis constant	PFK-reaction	fructose-1,6-bisphosphate
Michaelis constant	PFK-reaction	ATP
Michaelis constant	PFK-reaction	ADP
inhibitory constant	PFK-reaction	ATP
concentration	-	fructose-6-phosphate
concentration	-	fructose-1,6-bisphosphate
concentration	-	ATP
concentration	-	ADP
concentration of enzyme	PFK-reaction phosphofructokinase	
equilibrium constant	PFK-reaction	-
substrate catalytic rate constant	PFK-reaction	-
product catalytic rate constant	PFK-reaction	-
maximal velocity forward	PFK-reaction	-
maximal velocity backward	PFK-reaction	-
reaction affinity	PFK-reaction	-

Table A.4.: The complete kinetic parameter set of the PFK-reaction in SBtab format "KineticData": A '-' indicates rowfields, where no information is required. Retrieving numeric values for all of these parameters is the aim of the parameter balancing.

A.4.2. Check for the Validity of the Thermodynamic Dependencies

The thermodynamic dependencies explained in Chapter 2.6.2 have to be valid by construction of the dependence matrix R_θ^x. Referring to the parameter balancing results of Table 4.2, the dependencies are tested. The dependence between the equilibrium constants, the turnover rates, and the Michaelis constants are described in the logarithmic form of the Haldane relationship (for the sakes of clarity, the units of the Gibbs free energies are omitted in the following equations).

$$\begin{aligned} \ln k_l^{\mathrm{eq}} &= \ln k_{+l}^{\mathrm{cat}} - \ln k_{-l}^{\mathrm{cat}} + \sum_i n_{il} \ln k_{il}^{\mathrm{M}} \\ 4.59 &= 4.28 - 5.53 + 5.84\checkmark \end{aligned} \tag{A.1}$$

The Haldane relationship is valid. Another parameter type dependence concerns the forward and backward turnover rates. It denotes

$$\begin{aligned}
\ln k_{+l}^{\text{cat}} &= \ln k_l^{\text{V}} - \frac{1}{2}\sum_i n_{il}((G_i^{(0)})/RT + \ln k_{li}^{\text{M}}) \\
\ln(2.8e+06) &= \ln(426.76) - \frac{1}{2}(-1061.63/RT - 0.625 \\
&\quad -1695.03/RT - 0.491 + 1101.83/RT + 0.491 + 1643.38/RT + 3.968) \\
4.28 &= 4.907 + (-0.627) \\
&= 4.28, \checkmark
\end{aligned} \tag{A.2}$$

and

$$\begin{aligned}
\ln k_{-l}^{\text{cat}} &= \ln k_l^{\text{V}} + \frac{1}{2}\sum_i n_{il}((G_i^{(0)})/RT + \ln k_{li}^{\text{M}}) \\
\ln(9.8e+06) &= \ln(426.76) + \frac{1}{2}(-1061.63/RT - 0.625 \\
&\quad -1695.03/RT - 0.491 + 1101.83/RT + 0.491 + 1643.38/RT + 3.968) \\
5.53 &= 4.907 - (-0.623). \checkmark
\end{aligned} \tag{A.3}$$

The turnover rate dependence is valid as well, at last the maximal velocity dependence needs to be confirmed.

$$\begin{aligned}
\ln v_{+l}^{\max} &= \ln E_l + \ln k_l^{\text{V}} - \frac{1}{2}\sum_i n_{il}((G_i^{(0)})/RT + \ln k_{li}^{\text{M}}) \\
&= \ln E_l + \ln k_{+l}^{\text{cat}} \\
3.43 &= -0.851 + 4.28, \checkmark
\end{aligned} \tag{A.4}$$

and

$$\begin{aligned}
\ln v_{-l}^{\max} &= \ln E_l + \ln k_l^{\text{V}} + \frac{1}{2}\sum_i n_{il}((G_i^{(0)})/RT + \ln k_{li}^{\text{M}}) \\
4.68 &= -0.851 + 5.53. \checkmark
\end{aligned} \tag{A.5}$$

Both of the maximal velocities have been confirmed. Last but not least, the newly introduced concentration dependent Gibbs energies G and reaction affinities A are tested for their validity concerning the thermodynamic dependencies.

$$
\begin{array}{rcl}
G_i & = & G_i^{(0)} + \ln c_i \cdot RT \\
\hline
G_1 & = & G_1^{(0)} + \ln c_1 \cdot RT \\
-1641.68 & = & -1643.38 + 0.68 \cdot RT \checkmark \\
\hline
G_2 & = & G_2^{(0)} + \ln c_2 \cdot RT \\
-1103.96 & = & -1101.83 - 0.85 \cdot RT \checkmark \\
\hline
G_3 & = & G_3^{(0)} + \ln c_3 \cdot RT \\
-1061.17 & = & -1061.63 + 0.19 \cdot RT \checkmark \\
\hline
G_4 & = & G_4^{(0)} + \ln c_4 \cdot RT \\
-1697.15 & = & -1695.03 - 0.85 \cdot RT, \checkmark
\end{array}
\tag{A.6}
$$

and

$$
\begin{array}{rcl}
\ln A_l & = & -\sum_i n_{il} G_i^{(0)} - \sum_i n_{il} \ln c_i \cdot RT \\
5.09 & = & -(-1061.63 - 1695.03 + 1101.83 + 1643.38) \\
 & & -(0.19 - 0.85 - 0.85 + 0.68) \cdot RT \\
 & = & 4.59 + 0.498 \cdot RT. \checkmark
\end{array}
\tag{A.7}
$$

All the thermodynamic dependencies are proved to be valid for the parameter balancing results of the PFK reaction.

A.4.3. Balancing Results for the PFK Reaction Using a Pseudo Equilibrium Constant

QuantityType	Reaction	Compound	Value	Std	Unit
Gibbs free energy of formation	-	F6P	-1103.06	9.17	kJ/mol
Gibbs free energy of formation	-	ATP	-1644.6	9.17	kJ/mol
Gibbs free energy of formation	-	F16P	-1693.81	9.17	kJ/mol
Gibbs free energy of formation	-	ADP	-1060.41	9.17	kJ/mol
equilibrium constant	**PFK-reaction**	**-**	**1837.77**	**24357.97**	
turnover rate forward	PFK-reaction	-	4671.12	112811.74	1/s
turnover rate backward	PFK-reaction	-	2287.27	55239.56	1/s
maximal velocity forward	PFK-reaction	-	4329.55	227076.64	1/s
maximal velocity backward	PFK-reaction	-	2120.01	111190.68	1/s
reaction affinity	PFK-reaction	-	14769.59	9561418.03	

Table A.5.: Parameter balancing output for the case of a provided pseudo equilibrium constant (mean = 2980.96, std = 888610.0). The output values have significantly improved, the formerly huge values are disposed.

A.4.4. Parameter Balancing With a Provided Equilibrium Constant

If the user is providing an equilibrium constant in the input data, the whole estimation results become more appropriate than elsewise. This can be seen in Table A.6: the estimated values in this table show that the results become much better if the user provides an equilibrium constant.

QuantityType	Reaction	Compound	Value	Std	Unit
Gibbs free energy of formation	-	F6P	-1105.04	8.97	kJ/mol
Gibbs free energy of formation	-	ATP	-1646.59	8.97	kJ/mol
Gibbs free energy of formation	-	F16P	-1691.82	8.97	kJ/mol
Gibbs free energy of formation	-	ADP	-1058.42	8.97	kJ/mol
equilibrium constant	**PFK-reaction**	**-**	**0.93**	**0.16**	
turnover rate forward	PFK-reaction	-	2029.42	22949.98	1/s
turnover rate backward	PFK-reaction	-	1161.34	13133.14	1/s
maximal velocity forward	PFK-reaction	-	1881.02	46306.38	1/s
maximal velocity backward	PFK-reaction	-	1076.42	26498.86	1/s
reaction affinity	PFK-reaction	-	13.16	302.53	

Table A.6.: Parameter balancing output for the case of a provided equilibrium constant by the user (mean = 0.53, std = 0.17). The output values have significantly improved in their sensibility, huge values are disposed.

A.4.5. Input Data for the Balancing With Temperature and pH Value Regression

QuantityType	Reaction	Compound	Value	Std	Unit	pH	temperature
Gibbs free energy of formation	-	F16P	-1947.01	287.56	kJ/mol	-	-
Gibbs free energy of formation	-	ATP	-1885.03	287.56	kJ/mol	-	-
Gibbs free energy of formation	-	F6P	-1235.17	287.56	kJ/mol	-	-
Gibbs free energy of formation	-	ADP	-1286.93	287.56	kJ/mol	-	-
...							
equilibrium constant	PFK-reaction	-	0.0029	10	-	8	303.15
equilibrium constant	PFK-reaction	-	0.08	10	-	7	298.15
equilibrium constant	PFK-reaction	-	0.0048	10	-	8	310.15
...							

Table A.7.: Input parameter set taken from Brenda and Nissler et al. The focus is on the equilibrium constant of the PFK reaction that is provided with different temperatures and pH values on multiple entries.

A.4.6. Balancing Results for the PFK Reaction With Temperature and pH Value Regression

QuantityType	Reaction	Compound	Value	Std	Unit	μ(ln)	σ(ln)
Gibbs free energy of formation	-	F16P	-1685.9	9.04	kJ/mol	-	-
Gibbs free energy of formation	-	ATP	-1652.51	9.04	kJ/mol	-	-
Gibbs free energy of formation	-	F6P	-1110.96	9.04	kJ/mol	-	-
Gibbs free energy of formation	-	ADP	-1052.5	9.04	kJ/mol	-	-
...							
equilibrium constant	PFK-reaction	-	**0.0002**	**0.001**	-	2.54	3.33
...							

Table A.8.: Parameter estimation including pH value and temperature regression. The desired pH value is 8, the desired temperature is 310 K. The equilibrium constant is provided in the input data with different circumstances (see table A.7).

A.4.7. Balancing Results for the PFK Reaction With Only Gibbs Energies and Equilibrium Constants as Input Data

QuantityType	Reaction	Compound	Value	Std	Unit
Gibbs free energy of formation	-	F16P	-1691.82	8.97	kJ/mol
Gibbs free energy of formation	-	ATP	-1646.59	8.97	kJ/mol
Gibbs free energy of formation	-	F6P	-1105.04	8.97	kJ/mol
Gibbs free energy of formation	-	ADP	-1058.42	8.97	kJ/mol
geometric mean rate constant	PFK-reaction	-	-426.76	1277.38	1/s
Michaelis constant	PFK-reaction	F6P	2.18	7.47	mM
Michaelis constant	PFK-reaction	F16P	2.18	7.47	mM
Michaelis constant	PFK-reaction	ATP	2.18	7.47	mM
Michaelis constant	PFK-reaction	ADP	2.18	7.47	mM
inhibitory constant	PFK-reaction	ATP	2.6	7.85	mM
concentration of enzyme	PFK-reaction	PFK	0.93	1.78	mM
concentration	PFK-reaction	F6P	0.93	1.78	mM
concentration	PFK-reaction	F16P	0.93	1.78	mM
concentration	PFK-reaction	ATP	0.93	1.78	mM
concentration	PFK-reaction	ADP	0.93	1.78	mM
equilibrium constant	PFK-reaction	-	0.59	0.16	
turnover rate forward	PFK-reaction	-	2029.42	22949.98	1/s
turnover rate backward	PFK-reaction	-	1161.34	13133.14	1/s
maximal velocity forward	PFK-reaction	-	1881.02	46306.38	1/s
maximal velocity backward	PFK-reaction	-	1076.42	26498.96	1/s
Gibbs free energy	-	F16P	-1693.94	9.49	kJ/mol
Gibbs free energy	-	ATP	-1648.71	9.49	kJ/mol
Gibbs free energy	-	F6P	-1107.17	9.49	kJ/mol
Gibbs free energy	-	ADP	-1060.54	9.49	kJ/mol
reaction affinity	PFK-reaction	-	13.16	302.53	1/s

Table A.9.: Balancing result that is produced from only Gibbs free energies of formation and equilibrium constants as input.

A.5. Balancing Results for the Glycolysis Model

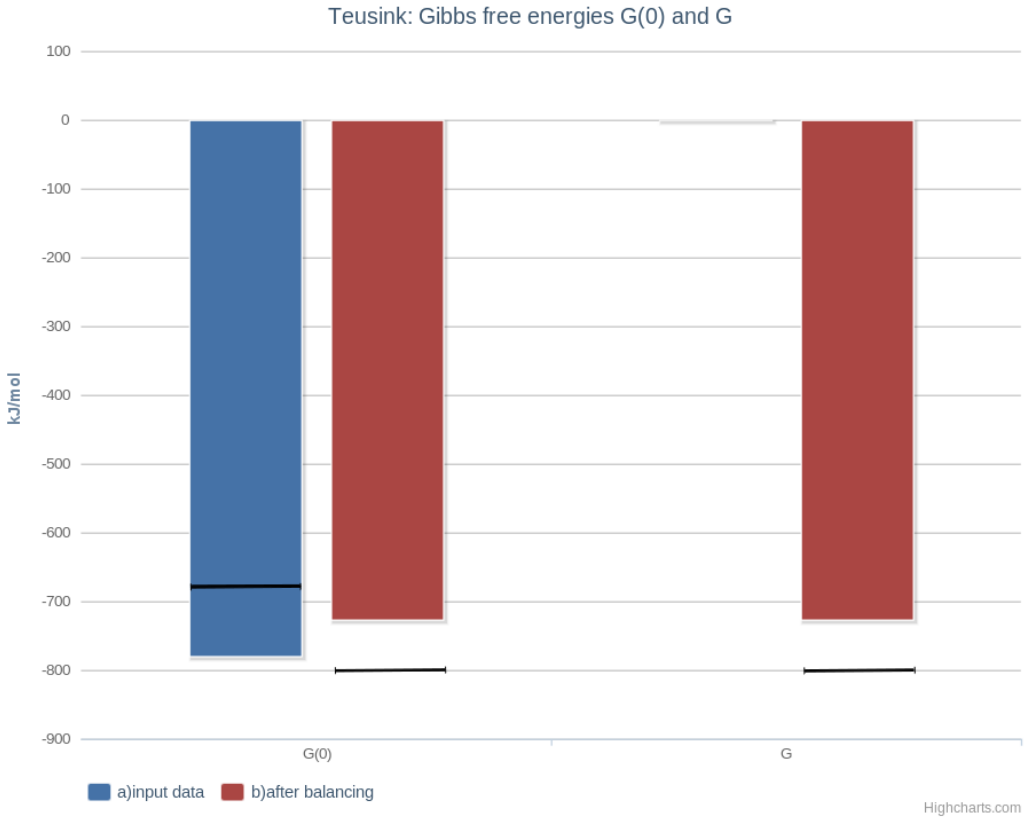

Figure A.10.: Teusink: Average Gibbs free energies. Gibbs free energies of formation before and after the balancing process, Gibbs free energies, before and after. The black bar marks the median.

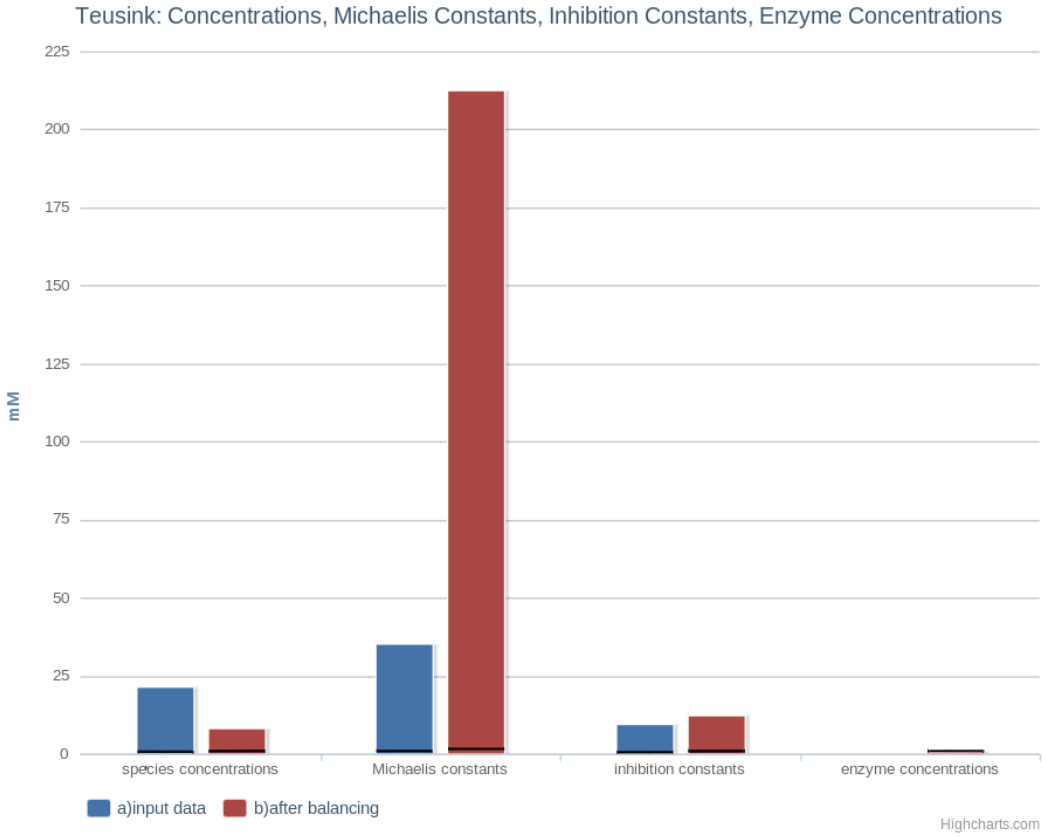

Figure A.11.: Teusink: Average concentrations of species, average Michaelis constants, average inhibition constants, average enzyme concentrations. The black bar marks the median.

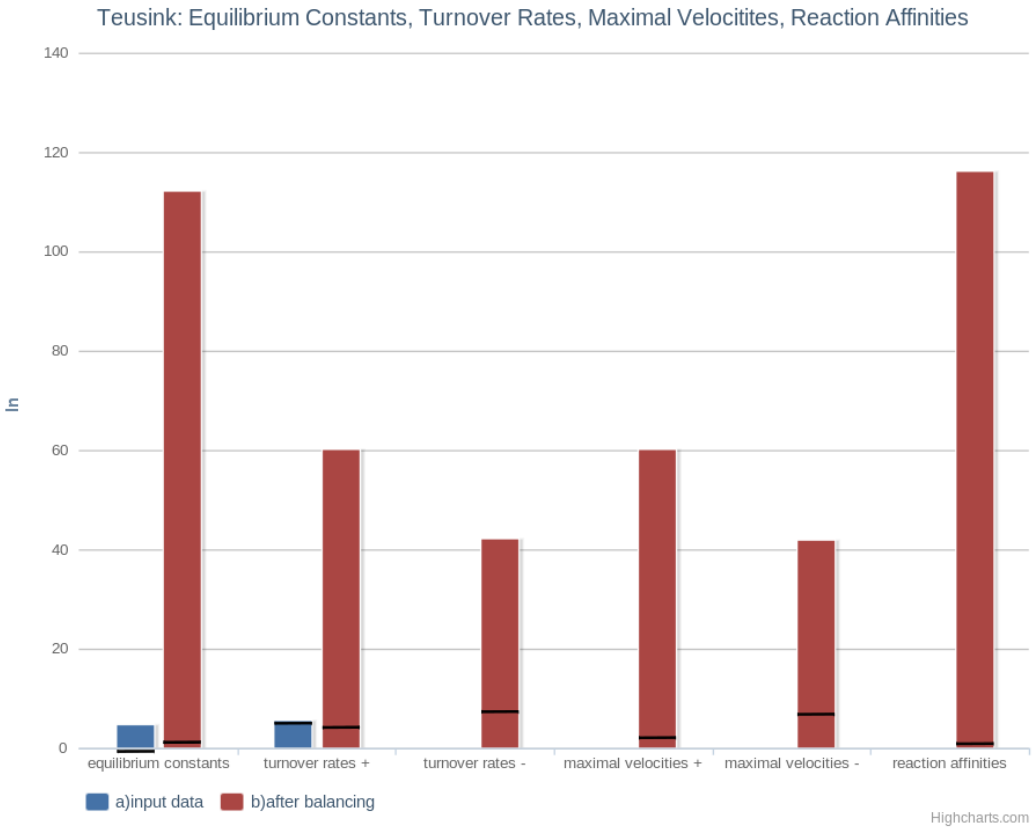

Figure A.12.: Teusink: Shown are the means of the logarithmic values of the average equilibrium constants, average turnover rate forward, average turnover rate backward, average maximal velocity forward, average maximal velocity backward, average reaction affinity. The black bar marks the median.

A.6. Balancing Results for the Threonine Model

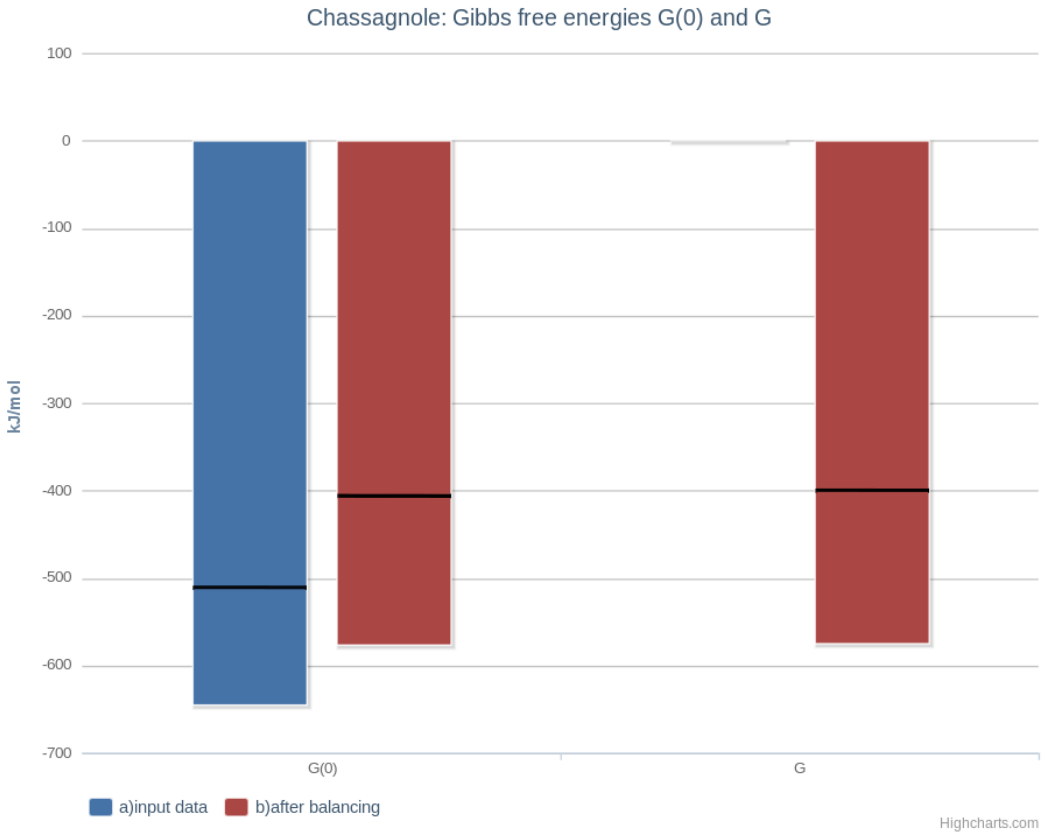

Figure A.13.: Chassagnole: Average Gibbs free energies; Gibbs free energies of formation before and after the balancing process. Concentration dependent Gibbs free energies, before and after. The black bar marks the median.

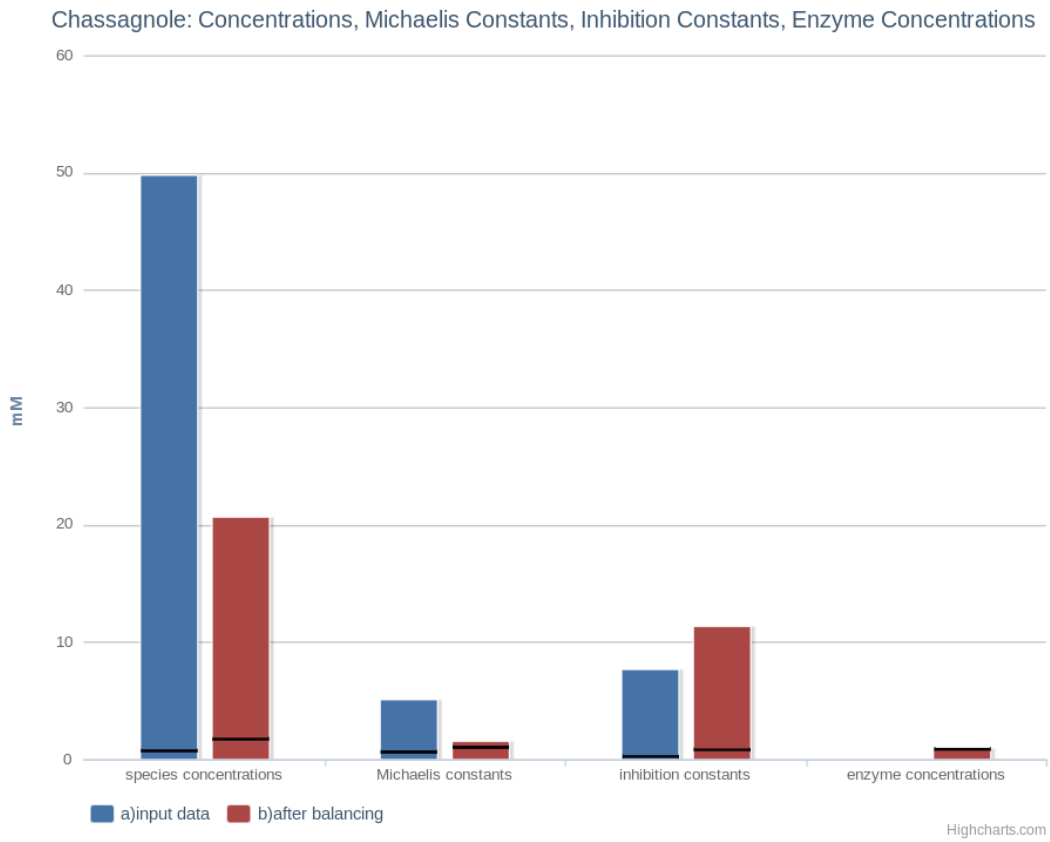

Figure A.14.: Chassagnole: Average concentrations of species, average Michaelis constants, average inhibition constants, average enzyme concentrations. The black bar marks the median.

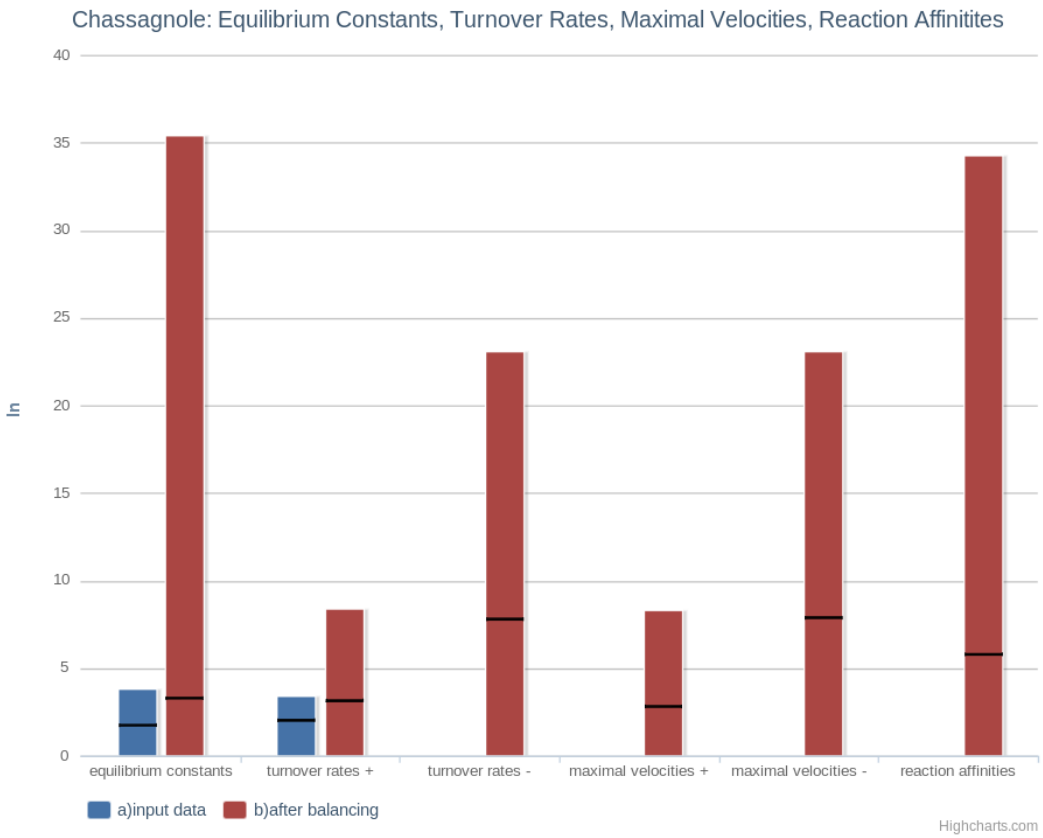

Figure A.15.: Chassagnole: Shown are the means of the logarithmic values of equilibrium constants, average turnover rate forward, average turnover rate backward, average maximal velocity forward, average maximal velocity backward, average reaction affinity. The black bar marks the median.

A.6.1. Comparing Model Data Information

	PFK reaction	Glycolysis model	Threonine model
Input data: Gibbs free energies average	-	-780.2	-643.88
Input data: Gibbs free energies median	-	-679.01	-509.1
Output data: Gibbs free energies average	-	-723.7	-578.11
Output data: Gibbs free energies median	-	-803.22	-407.6
Original data: Gibbs free energies average	-	-	-
Original data: Gibbs free energies median	-	-	-
Input data: Species concentrations average	-	21.1	49
Input data: Species concentrations median	-	3.2	2.1
Output data: Species concentrations average	0.93	8.45	21
Output data: Species concentrations median	0.93	3.9	1.9
Original data: Species concentrations average	0-5	-	2.49
Original data: Species concentrations median	0-5	-	2.22
Input data: Michaelis constants average	-	34.11	5
Input data: Michaelis constants median	-	2.2	1.5
Output data: Michaelis constants average	2.18	210.98	1
Output data: Michaelis constants median	2.18	0.3	0.9
Original data: Michaelis constants average	19.2	-	0.87
Original data: Michaelis constants median	19.2	-	0.22
Input data: Inhibitory constants average	-	8.6	8.2
Input data: Inhibitory constants median	-	1.6	0.5
Output data: Inhibitory constants average	-	10.1	15.2
Output data: Inhibitory constants median	-	1.8	0.3
Original data: Inhibitory constants average	-	1	1.92
Original data: Inhibitory constants median	-	1	0.39

	PFK reaction	Glycolysis model	Threonine model
Input data: Equilibrium constants average	-	6	7.1
Input data: Equilibrium constants median	-	0.2	-0.6
Output data: Equilibrium constants average	-	92	15.2
Output data: Equilibrium constants median	-	-0.5	0.9
Original data: Equilibrium constants average	-	7.48	10.62
Original data: Equilibrium constants median	-	1.1	0.09
Input data: Turnover rates average	-	6.45	3.1
Input data: Turnover rates median	-	6.11	2.2
Output data: Turnover rates average	-	59.1	8.3
Output data: Turnover rates median	-	5.5	3.01
Original data: Turnover rates average	-	-	-
Original data: Turnover rates median	-	-	-
Input data: Maximal velocities average	-	-	-
Input data: Maximal velocities median	-	-	-
Output data: Maximal velocities average	-	59.1	11.6
Output data: Maximal velocities median	-	5	4.2
Original data: Maximal velocities average	-	5.48	-2.21
Original data: Maximal velocities median	-	5.48	-2.3

Table A.10.: The comparison of the averages and means of the input data, the balanced output data and the data from the original publications and Sabio-RK.

References

Aitchison, J. & Brown, J. (1969). *The lognormal distribution*. Cambridge Univ. Pr.

Barthelmes, J., Ebeling, C., Chang, A., Schomburg, I., & Schomburg, D. (2007). BRENDA, AMENDA and FRENDA: the enzyme information system in 2007. *Nucleic Acids Research*, *35*(Database issue), D511.

Borger, S. (2008). *Parametrising kinetic models of biolog0ical networks*. PhD thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät I.

Bornstein, B., Keating, S., Jouraku, A., & Hucka, M. (2008). LibSBML: an API Library for SBML. *Bioinformatics*, *24*(6), 880.

Chassagnole, C., Rais, B., Quentin, E., Fell, D., & Mazat, J. (2001). An integrated study of threonine-pathway enzyme kinetics in Escherichia coli. *Biochemical Journal*, *356*(Pt 2), 415.

Chen, S., Zhao, T., Gordon, G., & Murphy, R. (2007). Automated image analysis of protein localization in budding yeast. *Bioinformatics*, *23*(13), i66.

Degtyarenko, K., de Matos, P., Ennis, M., Hastings, J., Zbinden, M., McNaught, A., Alcantara, R., Darsow, M., Guedj, M., & Ashburner, M. (2008). ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic acids research*, *36*(Database issue), D344.

Ellson, J., Gansner, E., Koutsofios, L., North, S., & Woodhull, G. (2001). Graphviz—open source graph drawing tools. In *Graph Drawing*, (pp. 594–597). Springer.

Gelman, A. (2004). *Bayesian data analysis*. CRC press.

Hoppe, A., Hoffmann, S., & Holzhuetter, H. (2007). Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks. *BMC Systems Biology*, *1*(1), 23.

Hucka, M., Finney, A., Sauro, H., Bolouri, H., Doyle, J., Kitano, H., et al. (2003). The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, *19*(4), 524.

Kanehisa, M. & Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic acids research*, *28*(1), 27.

Kitano, H. (2002). Computational systems biology. *Nature*, *420*(6912), 206–210.

Krause, F., Uhlendorf, J., Lubitz, T., Schulz, M., Klipp, E., & Liebermeister, W. (2010). Annotation and merging of SBML models with semanticSBML. *Bioinformatics*, *26*(3), 421.

Kriete, A. & Eils, R. (2006). *Computational systems biology*. Academic Press.

Laibe, C. & Le Novère, N. (2007). MIRIAM Resources: tools to generate and resolve robust cross-references in Systems Biology. *BMC Systems Biology*, *1*(1), 58.

Le Novère, N., Finney, A., Hucka, M., Bhalla, U., Campagne, F., Collado-Vides, J., Crampin, E., Halstead, M., Klipp, E., Mendes, P., et al. (2005). Minimum information requested in the annotation of biochemical models (MIRIAM). *Nature biotechnology*, *23*(12), 1509–1515.

Liebermeister, W. & Klipp, E. Sbtab: a standardised table format for systems biology. unpublished.

Liebermeister, W. & Klipp, E. (2006). Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. *Theoretical Biology and Medical Modelling*, *3*(1), 41.

Liebermeister, W., Uhlendorf, J., & Klipp, E. (2010). Modular rate laws for enzymatic reactions: thermodynamics, elasticities, and implementation. *Bioinformatics*, *27*.

Nazaret, C., Heiske, M., Thurley, K., & Mazat, J. (2009). Mitochondrial energetic metabolism: a simplified model of TCA cycle with ATP production. *Journal of theoretical biology*, *258*(3), 455–464.

Nissler, K., Otto, A., Schellenberger, W., & Hofmann, E. (1983). Similarity of activation of yeast phosphofructokinase by AMP and fructose-2, 6-bisphosphate. *Biochemical and Biophysical Research Communications*, *111*(1), 294–300.

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmuller, U., & Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, *25*(15), 1923.

Schomburg, I., Chang, A., Hofmann, O., Ebeling, C., Ehrentreich, F., & Schomburg, D. (2002). BRENDA: a resource for enzyme data and metabolic information. *Trends in biochemical sciences*, *27*(1), 54–56.

Snoep, J. & Olivier, B. (2002). Java Web Simulation (JWS); a web based database of kinetic models. *Molecular biology reports*, *29*(1), 259–263.

Teusink, B., Passarge, J., Reijenga, C., Esgalhado, E., van der Weijden, C., Schepper, M., Walsh, M., Bakker, B., van Dam, K., Westerhoff, H., et al. (2000). Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. *European Journal of Biochemistry*, *267*(17), 5313–5329.

Wittig, U., Golebiewski, M., Kania, R., Krebs, O., Mir, S., Weidemann, A., Anstein, S., Saric, J., & Rojas, I. (2006). SABIO-RK: integration and curation of reaction kinetics data. In *Data integration in the life sciences*, (pp. 94–103). Springer.

Zamboni, N., Kuemmel, A., & Heinemann, M. (2008). anNET: a tool for network-embedded thermodynamic analysis of quantitative metabolome data. *BMC bioinformatics*, *9*(1), 199.

Zi, Z. & Klipp, E. (2006). SBML-PET: a Systems Biology Markup Language-based parameter estimation tool. *Bioinformatics*, *22*(21), 2704.

B. Eidesstattliche Erklärung

Hiermit erkläre ich, die vorliegende Arbeit selbstständig ohne fremde Hilfe verfasst und nur die angegebene Literatur und Hilfsmittel verwendet zu haben.

Timo Lubitz (Verfasser)
Berlin, den 01. Mai 2010