Parameter Balancing in Kinetic Models of Cell Metabolism[†]

Timo Lubitz, Marvin Schulz, Edda Klipp,* and Wolfram Liebermeister*

Humboldt-Universität zu Berlin, Institut für Biologie, Theoretische Biophysik,
Invalidenstrasse 42, D-10115 Berlin

Received: September 14, 2010; Revised Manuscript Received: October 20, 2010

Kinetic modeling of metabolic pathways has become a major field of systems biology. It combines structural information about metabolic pathways with quantitative enzymatic rate laws. Some of the kinetic constants needed for a model could be collected from ever-growing literature and public web resources, but they are often incomplete, incompatible, or simply not available. We address this lack of information by parameter balancing, a method to complete given sets of kinetic constants. Based on Bayesian parameter estimation, it exploits the thermodynamic dependencies among different biochemical quantities to guess realistic model parameters from available kinetic data. Our algorithm accounts for varying measurement conditions in the input data (pH value and temperature). It can process kinetic constants and state-dependent quantities such as metabolite concentrations or chemical potentials, and uses prior distributions and data augmentation to keep the estimated quantities within plausible ranges. An online service and free software for parameter balancing with models provided in SBML format (Systems Biology Markup Language) is accessible at www.semanticsbml.org. We demonstrate its practical use with a small model of the phosphofructokinase reaction and discuss its possible applications and limitations. In the future, parameter balancing could become an important routine step in the kinetic modeling of large metabolic networks.

Introduction

The complex, dynamic behavior of cell metabolism can be simulated by mathematical models. Metabolic pathway models consist of enzymatic reactions described by their stoichiometry, the enzymatic rate laws, and their kinetic constants (such as, for instance, equilibrium constants or catalytic constants). The more we know about these quantities, the more reliably we can simulate the metabolic dynamics. Kinetic laws of individual enzymes have been studied experimentally for about 100 years,[1] and metabolic control theory,[2] a theoretical apparatus for the analysis of metabolic systems, has been developed since the 1970s. Recently, comprehensive web databases, advances in high-throughput experiments, and inexpensive computing power have led to a new interest in metabolic modeling. In particular, the numerous large-scale metabolic networks reconstructed from sequenced genomes[3−5] call for automatic routines that can fill these networks with enzymatic rate laws and turn them into dynamic models.

Unfortunately, the enzymatic mechanisms and the rate laws of most enzymes are unknown, and it is laborious to determine them exclusively by enzyme assays. A pragmatic solution is to substitute missing kinetic laws by standard rate laws, such as mass-action kinetics, generalized mass-action kinetics,[6] or linlog kinetics.[7,8] Here, we will use the common modular rate law,[9] a generalized version of the reversible Michaelis−Menten rate law, suitable for any reaction stoichiometry and accounting for various types of allosteric regulation. Once a metabolic network and enzymatic rate laws have been chosen, we need numerical values for the kinetic constants. This can be a challenge, especially for large networks. Modelers can find known kinetic constants in published models, in the literature, or in public web resources such as Sabio-RK,[10] Brenda,[11] and NIST.[12,13] As pointed out by Alberty,[14] varying conditions such as pH or salt concentrations can be taken into account by describing biochemical reactants and reactions in terms of transformed thermodynamic quantities. In the future, automated enzyme assays might provide more kinetic data, but they will still not reach the speed at which metabolic networks are reconstructed from newly sequenced genomes. Available kinetic data may not be suited for a model if they are contradictory or measured under inappropriate conditions (e.g., pH values and temperatures). Furthermore, data collected from various sources are very unlikely to represent a thermodynamically consistent set. Since incompleteness of the kinetic constants remains a major obstacle, methods for guessing unknown kinetic constants or adjusting the known values will become increasingly important.

Here, we present parameter balancing, an approach to infer complete and consistent sets of model parameters from incomplete, inconsistent kinetic data. This is only possible due to mutual dependencies between the kinetic constants and other model parameters, arising from their definitions or from thermodynamic laws (Wegscheider conditions[15] and Haldane relationships). In a simple approach, incomplete kinetic data sets could be complemented by inserting all available values into the model and adding other quantities that can be directly computed from them. However, this might leave parameters undetermined and would not eliminate inconsistencies between the original data values. As a better strategy, we determine parameter sets that are consistent *by construction* and resemble the original data as closely as possible. Since these values may not be uniquely determined, we have to restrict them to plausible ranges and quantify the remaining uncertainties by error bars. We have previously suggested[16] implementing this parameter balancing approach in a Bayesian statistical framework[17] and based on standard rate laws.[9] The critical step is to identify a

---

Kinetic Models of Cell Metabolism

J. Phys. Chem. B, Vol. 114, No. 49, 2010 **16299**

set of independent model quantities that can be freely selected during parameter fitting, sampling, or optimization and from which all remaining quantities can be easily computed. After establishing this dependence scheme for a wide range of kinetic constants and metabolic quantities, we obtain a relatively general and simple data integration method that can guarantee consistent parameter sets. Here, we use more general formulas and present an interactive online workflow for models in SBML format (Systems Biology Markup Language,[18] www.sbml.org), then illustrate its use with an example case.

Methods

Parameter balancing exploits the fact that kinetic model parameters often share mutual dependencies, either by their definition or because of thermodynamic constraints.[16] For kinetic models with modular rate laws[9] and many other rate laws, the model parameters can be split into two subsets: a set of independent *basis quantities* that can be chosen arbitrarily and a set of *derived quantities* that can be computed from linear combinations of the basis quantities. As a simple example, let us consider the concentration and the amount of a substance within a cellular compartment. The amount a can be computed from the concentration c and the compartment volume V by the formula $a = cV$. If we choose (e.g., guess, sample, fit, or optimize) all three model parameters independently, this dependence will be violated. Of course, this problem is easy to avoid: we just need to treat a as a *derived quantity* to be computed from the *basis quantities* c and V. On a logarithmic scale, we obtain the simple linear *dependence scheme* $\ln a = \ln c + \ln V$. In parameter balancing, we do exactly the same thing, but treat all quantities of a (possibly large) kinetic model at the same time. In addition, we consider not only dependencies arising from quantity definitions, as in this simple example, but also more involved dependencies arising from the laws of thermodynamics. The resulting dependence scheme consists of many linear equations, emerging from a thermodynamic and kinetic analysis of the rate laws. For practical calculations, these equations are represented by a large, sparse matrix to be constructed from the metabolic network.

Parameter balancing exploits this partition of the parameter set into basis and derived quantities. Initially, the basis quantities are estimated from data by a linear regression model that follows directly from the dependence scheme. Following this, the dependence scheme is used again to determine the dependent model quantities. The estimation is based on Bayesian statistics, combining the data with prior distributions for all model parameters, which can be selected to incorporate general knowledge about plausible values. Accordingly, parameter balancing yields not only point estimates but also posterior distributions, from which mean values, variances, correlations, and even random samples of all relevant quantities can be obtained. A detailed description is given in the Supporting Information. On the basis of our practical experience, we have extended the original parameter balancing approach in three major ways:

(1) Our model quantities comprise not only kinetic constants, but possibly also metabolite and enzyme concentrations for one or more metabolic states. From their values, we can derive other state-dependent quantities; in particular, the chemical potentials and reaction affinities. The reaction affinities describe how strongly chemical reactions deviate from their equilibrium and predetermine the reaction directions.

(2) In biochemical reactions, individual protonation states of a molecule ("chemical species") are usually subsumed in a single
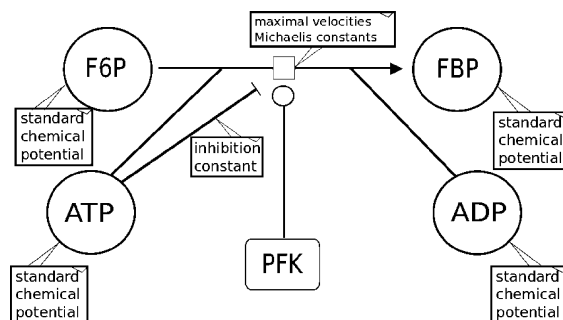

Figure 1. The PFK reaction as a network model. The model structure is defined by the sum formula F6P + ATP ↔ FBP + ADP with the molecular species F6P (fructose 6-phosphate), FBP (fructose 1,6-bisphosphate), ATP (adenosine triphosphate), and ADP (adenosine diphosphate). In our parameter balancing workflow, the stoichiometry is provided as an SBML file, and kinetic constants (shown in flags) are read from a separate data file (see Table 1). The constants shown suffice to define a common modular rate law,[9] which we use as a standard rate law.

"reactant" concentration. As pointed out by Alberty,[14,19,20] these reactants should be described by *transformed* thermodynamic quantities, which effectively depend on the pH value and on salt concentrations. In our approach, quantities such as equilibrium constants, Gibbs free energies, and chemical potentials are given as transformed values, depending both on temperature and on pH. If experimental values stem from measurements under incompatible conditions, the discrepancies can be corrected during parameter balancing.

(3) Although prior distributions can help to define plausible ranges for the basis quantities, they are not applicable to the derived quantities. As a consequence, whenever few data on these quantities are available, their large uncertainties lead to unreasonable balancing results. We address this problem by augmenting the experimental data set with fictitious pseudo values, playing a role similar to the prior distributions. They allow the modeler to control the variance of the independent values and, thus, the reliability of the whole estimate.

A detailed description of the original method and all new features is given in the Supporting Information.

For practical use, we have implemented an interactive workflow for SBML models, allowing the user to balance the parameters and to replace or complete kinetic rate laws in the model. Three kinds of information are needed as an input: the network structure, which is obtained from the SBML file, the mathematical rate laws, which are chosen from the list of modular rate laws,[9] and a table with collected kinetic constants and other relevant data (for an example, see Figure 1 and Table 1). In the model, enzymes and allosteric activators and inhibitors need to be listed as reaction modifiers and specified by SBO terms. The quantity types are defined as in the Systems Biology Ontology,[21] and the names of the reactions and species refer to IDs in the SBML model. All this helps the process to run in an almost fully automated manner. For further details, see the Supporting Information and our online documentation at www.semanticsbml.org.

At the end of the workflow, kinetic rate laws with a set of balanced parameters are inserted into the SBML model. As a result of balancing, most parameters are represented by normal distributions of their logarithmic values. When converting these values back to nonlogarithmic scale, we obtain log-normal distributions and need to carefully distinguish between their arithmetic and geometric mean values. To insert the most probable parameter set into a model, we choose the median

TABLE 1: Extract of the Input Data Table for the Phosphofructokinase Reaction[a]

quantity type	SBML reaction	SBML species	mean	std	unit	ref
standard chemical potential		F6P	−1316.55		kJ/mol	Alberty
standard chemical potential		FBP	−2206.14		kJ/mol	Alberty
standard chemical potential		ATP	−2292.28		kJ/mol	Alberty
standard chemical potential		ADP	−1425.17		kJ/mol	Alberty
inhibitory constant	PFK	ATP	0.396		mM	Brenda
concentration		FBP	1.94		mM	pseudo value
concentration		F6P	0.97		mM	pseudo value
concentration		ATP	1.5		mM	Brenda
concentration		ADP	0.81		mM	Brenda
concentration of enzyme		ATP	0.003608		mM	yeastGFP
equilibrium constant	PFK		0.08			Nissler et al.
Michaelis constant	PFK	F6P	0.66		mM	Brenda
Michaelis constant	PFK	FBP	12.5		mM	Brenda
Michaelis constant	PFK	ATP	0.1		mM	Brenda
Michaelis constant	PFK	F6P	0.945		mM	Brenda

[a] For abbreviations, see Figure 1. Data taken from Alberty,[20] yeastGFP,[29] Nissler et al.,[30] NIST,[13] Brenda.[11]

TABLE 2: Balancing Result for the Phosphofructokinase Reaction

quantity type	SBML reaction	SBML species	mean	std	median	unit
standard chemical potential		FBP	−2191.919	30.232	−2191.919	kJ/mol
standard chemical potential		ATP	−2292.069	30.232	−2292.069	kJ/mol
standard chemical potential		F6P	−1321.474	30.232	−1321.474	kJ/mol
standard chemical potential		ADP	−1415.064	30.232	−1415.064	kJ/mol
catalytic rate constant geometric mean	PFK		13.148	25.602	6.006	1/s
Michaelis constant	PFK	F6P	0.630	0.306	0.567	mM
Michaelis constant	PFK	FBP	9.956	4.829	8.957	mM
Michaelis constant	PFK	ATP	0.103	0.050	0.093	mM
Michaelis constant	PFK	ADP	0.835	0.405	0.752	mM
inhibitory constant	PFK	ATP	0.389	0.103	0.376	mM
concentration		F6P	0.775	1.175	0.427	mM
concentration		FBP	1.454	2.204	0.801	mM
concentration		ATP	1.478	0.393	1.428	mM
concentration		ADP	0.769	0.458	0.660	mM
concentration of enzyme	PFK		0.003	0.001	0.003	mM
equilibrium constant	PFK		0.080	0.040	0.072	
substrate catalytic rate constant	PFK		45.309	102.940	18.253	1/s
product catalytic rate constant	PFK		4.906	11.146	1.976	1/s
forward maximal velocity	PFK		0.016	0.039	0.006	mM/s
reverse maximal velocity	PFK		0.150	0.357	0.058	mM/s
chemical potential		ATP	−2291.18	30.239	−2291.18	kJ/mol
chemical potential		F6P	−1323.597	30.354	−1323.597	kJ/mol
chemical potential		ADP	−1416.099	30.263	−1416.099	kJ/mol
chemical potential		FBP	−2192.472	30.354	−2192.472	kJ/mol
reaction affinity	PFK		0.003	0.005	0.003	kJ/mol

values of the nonlogarithmic parameters, which are identical to the geometric mean. In addition, we can sample additional logarithmic parameter sets from the normal distribution, convert them back by applying the exponential function, and insert them into the model.

In the Supporting Information, we discuss a number of possible extensions, such as handling of identical species in different compartments, electrochemical potentials, cell compartments with different pH values, forward and reverse reaction rates, use of correlated priors, use of equilibrium constants as basis quantities, and prescribed reaction directions. We plan to include these features in a future version of the workflow.

Results

Implementation As an Interactive Workflow. We have implemented parameter balancing as an open source software written in Python. An interactive online version is accessible at www.semanticsbml.org. The Web site also contains the documentation and a number of example models, including the phosphofructokinase reaction discussed below. At the beginning

of the workflow, the user uploads an SBML model,[18] defining the network structure (see Figure 1) as well as a formatted data table (see Table 1) listing the known kinetic constants and other quantities relevant for the model. After uploading both files, the data can be filtered for a specific source organism, and missing standard errors are completed by default values. Then, a table of relevant model quantities is produced, with blank rows where data are missing and averaged values where more than one data point was available. The standard chemical potentials or equilibrium constants measured under different temperature or pH are not averaged, but kept as separate values. As a second source of information, we use prior distributions, describing our general expectation about the quantity types. Such priors can be derived from the typical ranges of kinetic constants found in the literature.

On the basis of the prior distributions and pseudo values chosen by the user, a set of model parameters is determined by parameter balancing (see Table 2). For comparison, the user can also choose between several simpler methods for data completion: (i) completing all missing quantities by the prior

Kinetic Models of Cell Metabolism

J. Phys. Chem. B, Vol. 114, No. 49, 2010 **16301**

TABLE 3: Comparison of Balancing Results to Literature and Web Resource Results[a]

	V_{max}	c_{F6P}	c_{FBP}	c_{ATP}	c_{ADP}	k^{eq}
Rizzi et al.[36]	2.33 mM/s					
Hynne et al.[26]	0.76 mM/s	0.49 mM	4.64 mM	2.1 mM	1.5 mM	∞
Teusink et al. (measured)[22]	0.68 U/mg Protein^{-1}	0.62 mM	5.51 mM	2.52 mM	1.32 mM	∞
Teusink et al. (predicted)[22]	3.05 mM/s	0.16 mM	0.98 mM	2.52 mM	1.29 mM	80
Sabio-RK[10]			0−5 mM			
Brenda[11]						
before balancing		0.97 mM	1.94 mM	1.5 mM	0.81 mM	0.08
parameter balancing	0.006 mM/s	0.427 mM	0.801 mM	1.428 mM	0.660 mM	0.072

[a] Maximal forward velocity, species concentrations, and the equilibrium constant are shown.

TABLE 4: Comparison of Balancing Results to Literature and Web Resource Results: Michaelis Constants and Inhibitory Constant

	k^M_{F6P}	k^M_{FBP}	k^M_{ATP}	k^M_{ADP}	k^I_{ATP}
Rizzi et al.[36]	0.008 mM		0.25 mM	0.36 mM	
Hynne et al.[26]					
Teusink et al. (measured)[22]	0.1 mM		0.71 mM		
Teusink et al. (predicted)[22]					
Sabio-RK[10]		19.2 mM			1.09 mM
Brenda[11]	0.66 mM	12.5 mM	0.1 mM	0.945 mM	
before balancing	0.66 mM	12.5 mM	0.1 mM	0.945 mM	0.396 mM
parameter balancing	0.567 mM	8.957 mM	0.093 mM	0.752 mM	0.376 mM

means and widths (for missing basis quantities) or by pseudo values and their standard errors (for missing derived quantities), leading to complete, but inconsistent parameter sets; (ii) completing all missing basis quantities by prior values and computing all derived quantities by the dependence scheme; (iii) completing all missing derived quantities by pseudo values, followed by balancing without pseudo values; or (iv) balancing without pseudo values. Afterward, the user can choose a rate law from the list of modular rate laws[9] that will be inserted into the SBML file. The parameter values represent either median values or random samples from the posterior distribution. Finally, the balanced quantities and the completed model are exported, respectively, as a data table and as a fully parametrized SBML file.

Parameter Balancing for the Phosphofructokinase Reaction. We have tested parameter balancing with medium-scale models of central metabolism: yeast glycolysis models by Teusink et al.[22] and Hynne et al.[23] and a model of metabolism in pancreatic beta cells.[24] The original models, the collected data, and the balancing results for all models can be found at www.semanticsbml.org in the online documentation. For simplicity, we consider here a small model of the phosphofructokinase reaction, a key step in upper glycolysis:

$$\beta\text{-D-fructose 6-phosphate} + \text{ATP} \leftrightarrow$$
$$\beta\text{-D-fructose 1,6-bisphosphate} + \text{ADP}$$

The enzyme phosphofructokinase (PFK), which transfers a phosphate group from ATP (adenosine triphosphate) to fructose 6-phosphate (see Figure 1), has been studied extensively, but the databases Brenda,[11] NIST,[13] and Sabio-RK[10] do not contain a complete kinetic data set for a reversible rate law. Therefore, we pretend that the kinetic law of PFK is unknown, apply parameter balancing to the pooled data from several organisms, and compare the results to parameters from published kinetic models for the Baker's yeast *Saccharomyces cerevisiae*.[22,25,26]

Our SBML model for the PFK reaction, including MIRIAM-compliant annotations, was automatically constructed from the KEGG[27] reaction identifier R04779 with the tool seman-ticSBML[28] (accessible at www.semanticsbml.org). Then, we

collected data from the databases Sabio-RK,[10] Brenda,[11] NIST,[13] and yeastGFP[29] and from publications by Nissler et al.,[30] Albe et al.,[31] and Alberty.[20] The preprocessed data are shown in Table 1. For the concentrations of fructose 6-phosphate and fructose 1,6-bisphosphate, we inserted pseudo values (arithmetic mean values arising from geometric mean values 0.5 and 1 mM, respectively, with a broad standard deviation of 0.5 for the decadic logarithms). Some of the values were obtained by averaging over several data values measured in different organisms. The set of balanced parameters, obtained with the default workflow settings, is shown in Table 2.

A comparison with values from the literature and existing models is shown in Tables 3 and 4. The maximal velocity of the phosphofructokinase, mainly determined by a broad prior on catalytic constants and by an experimentally derived count number of the enzyme molecules, remains within a sensible range of the values from the literature (0.006 mM/s). We find the equilibrium constant to be much lower (0.072) than the one estimated by Teusink et al.[22](80), which is clearly due to the small input value (0.08) from Nissler et al.[30] Finally, the balanced inhibition constant for ATP (0.376 mM) lies within a sensible range of the one found in Sabio-RK[10] (1.09 mM), but is nonetheless a little lower, since the input data value (0.396 mM) is lower, as well.

The equilibrium constants play a central role in parameter balancing and their numerical values primarily arise from measured values and from known standard chemical potentials. To incorporate their dependence on pH values, we reran the balancing with input data containing equilibrium constants measured at different temperatures and pH values (see Table 5). Instead of just averaging over them (as for other duplicate values), parameter balancing can adjust the values to a certain target temperature and pH value chosen by the user. The aim is, of course, to describe in vivo conditions considered in the model. When choosing a target temperature of 300 K and a target pH value of 7, we obtained a balanced equilibrium constant $k^{eq} = 0.19397$. Since the input measurement conditions (average pH, 7.7; average temperature, 303.82 K) differ from the desired conditions (pH, 7; temperature, 300 K), the input value for the equilibrium constant ($k^{eq} = 0.02923$) changes

TABLE 5: Input Data for Different Measurement Conditions (pH and temperature)

quantity type	SBML reaction	SBML species	mean	std	unit	pH	temperature	ref
standard chemical potential		FBP	−2206.14		kJ/mol			Alberty
standard chemical potential		ATP	−2292.28		kJ/mol			Alberty
standard chemical potential		F6P	−1316.55		kJ/mol			Alberty
standard chemical potential		ADP	−1425.17		kJ/mol			Alberty
equilibrium constant	PFK		0.0029			8	303.15	Nissler et al.
equilibrium constant	PFK		0.08			7	298.15	Nissler et al.
equilibrium constant	PFK		0.0048			8	310.15	Nissler et al.

significantly after parameter balancing. When we choose target conditions closer to the data measurement conditions (pH, 7.7; temperature, 304 K), the balanced value $k^{eq} = 0.19393$ moves slightly closer to the data value.

Finally, we tested what would happen without any direct information on equilibrium constants. When we remove all data and pseudo values for the equilibrium constant, we obtain abnormally high values for the majority of the derived parameters. Even though equilibrium constants can in principle be obtained from standard chemical potentials, this result indicates that the remaining uncertainty may be extremely large, and pseudo values are an efficient way to limit the ranges of balanced values.

Parameter balancing can also be applied to models of bigger size. Due to the large amounts of data that are produced, the result tables are not included in this publication. Instead, the detailed kinetic data for the models of Teusink et al.[22] (17 reactions, computing time 0.29 s), Hynne et al.[23] (24 reactions, computing time 0.50 s), and Jiang et al.[24] (45 reactions, computing time 2.32 s) can be found, downloaded, and used on our Web site www.semanticsbml.org. Nevertheless, the rising resource demands of bigger models can be a limiting factor. In the Supporting Information, we discuss several possibilities to break down the calculations into manageable pieces.

Discussion

The example of the phosphofructokinase reaction shows that parameter balancing can produce consistent estimates on kinetic constants in plausible ranges and close to available input data. In the spirit of Bayesian statistics, we do not estimate the unknowns by averaging over the data, but by fitting the data with a statistical model; in our case, produced from the dependence scheme. One advantage is that such a model can handle not only uncertainties of individual quantities, but also correlated uncertainties arising from parameter dependencies. Of course, this approach strongly depends on the collected input data and on the chosen prior distributions. In situations when input data are missing, we found that pseudo constants can significantly improve the results. As a side effect, balancing will shift all the values—even if data are available for a certain quantity—toward the center of the prior distribution and toward the pseudo values, as can be seen in the example shown in Tables 3 and 4. If this effect is unwanted, there are two possibilities to avoid it: on one hand, experimental values can be fixed by assigning small standard errors to them; on the other hand, our workflow also provides the possibility to replace unknown values by prior or pseudo values without further balancing. This, however, will lead to inconsistent parameter sets.

The median posterior values obtained from balancing can be used as point estimates, but we can also sample parameter sets from the posterior distribution. Studying models with sampled parameters can be an emergency solution if few data are available, but it is also a convenient way to explore the dynamics

in metabolic networks for a wide range of kinetic parameters and to discern the influences of network structure and kinetics on the dynamic behavior.

Parameter balancing is a general approach that could be applied to any physical model as long as all parameters fit into a linear dependence scheme. For kinetic models, we could establish linear dependencies for most relevant quantities, considering some of them on logarithmic scale. The choice of model parameters to be balanced is not fixed, but can depend on the specific situation (for a variety of possible variants, see the Supporting Information). In the scheme presented here, the equilibrium constants are derived from standard chemical potentials, which allows us to incorporate data or predictions of Gibbs free energies of formation; a general alternative, with even fewer parameters, would be to express all equilibrium constants by a set of independent equilibrium constants (see the Supporting Information). However, since model identifiability is guaranteed by the usage of prior distributions, keeping the number of parameters small is usually not crucial.

Some of the constants (for instance, the inhibition constants) are independent of all other parameters and could therefore be balanced separately, but as long as the parameter set is not too large, it turned out to be practical to account for all kinetic constants and metabolic data in one large dependence matrix. The reaction rates would be the most important target but do not fit into the scheme. In the future, their signs and the individual forward and backward rates could be used as input data. Once a subset of the flux directions have been predefined, available data on metabolite levels can directly contribute to the balancing of kinetic constants and thereby improve the estimation results.

Arguably, the most critical point in our approach is the use of heterogeneous data from different sources. Ideally, all kinetic constants used in a model should stem from measurements under the same, standardized, nearly in vivo conditions. Since such data are rarely available, modelers often need to utilize heterogeneous data collected from literature and databases, acknowledging that this may cause various problems. First, kinetic constants measured in vitro and in vivo differ due to different pH values, temperatures, salt concentrations, or other factors. Although we attempt to take these conditions into account, it may be difficult to apply corrections, since the exact conditions in living cells are not known. Second, most kinetic models neglect the complexity of the cell (e.g., molecular crowding, channeling, inhomogeneous localization of enzymes). Since the kinetic parameters in such models describe an effective behavior (e.g., averaged over different cell compartments), they will differ from the values measured in vitro. Third, there are different conventions for reaction formulas (e.g., H^+ ions and H_2O molecules may or may not be listed) and about the standard concentration c^0 (in our case, 1 mM). Since the definition of transformed equilibrium constants crucially depends on the reaction formulas, it is important that model and data sources use the same, appropriate conventions. Finally, if kinetic

constants are taken from existing models, their values may become invalid out of this specific context. Thus, especially for poorly identifiable parameters, it is important to consider the uncertainties in the original estimations.

Facing these difficulties, parameter balancing follows a pragmatic approach: even if data have a low quality, we may still use them as clues about unknown parameters. This is why uncertainty ranges play a central role here. If a quantity has been measured under the wrong conditions, we may account for this by increasing its standard error. On one hand, this will decrease the weight of this data point in the balancing process; on the other hand, the uncertainty of all related balanced parameters will increase, reflecting our precautionary approach.

In some cases, we may need to assign very small uncertainties to some of the input data. For instance, if an SBML model contains kinetic laws and we intend to replace only some of them, we have to make sure that the new rate laws are compatible with those pre-existing rate laws that will remain in the model. As a key precondition, all equilibrium constants in the resulting model need to satisfy the Wegscheider conditions.[15] To ensure this, one can determine the equilibrium constants of the existing rate laws and use them as input data (with zero standard error) in parameter balancing.

Due to the large uncertainties in standard chemical potentials and metabolite concentrations, it is unlikely that models obtained from parameter balancing will directly show realistic stationary flux distributions. In the future, this could be enforced by a subsequent fit to "omics" data[16] or by predefining the signs of reaction affinities. These signs will induce dependencies between kinetic and metabolic quantities; for instance, between the standard chemical potentials and the metabolite concentrations. If several metabolic states are treated within the same dependence scheme, the resulting method would resemble network-embedded thermodynamic analysis[32] and allow to use the results of previous fluxome and metabolome analysis (by flux-balance methods including thermodynamic constraints[32−35]) as input data for parameter balancing.

Conclusions

Parameter balancing yields consistent and complete parameter sets for kinetic models, as a potential starting point for further modeling. It can integrate incomplete and contradictory input data and respects constraints implied by common modeling assumptions (standard rate laws; reactants in ideal, dilute solution). If few input data are available, parameter balancing can help to quantify the remaining uncertainties, whereas with more and higher-quality data, the predictions become more trustworthy and precise. The resulting posterior distribution can be used to define parameter ranges, to sample possible parameter sets, or to be reused as a prior for following rounds of parameter balancing.

References and Notes

(1) Michaelis, L.; Menten, M. L. *Biochem. Z.* **1913**, *49*, 334–336.

(2) Heinrich, R.; Schuster, S. *The Regulation of Cellular Systems*; Chapman & Hall: New York, 1996.

(3) Feist, A.; Henry, C.; Reed, J.; Krummenacker, M.; Joyce, A.; Karp, P.; Broadbelt, L.; Hatzimanikatis, V.; Palsson, B. *Mol. Syst. Biol.* **2007**, *3*, 121.

(4) Herrgård, M.; et al. *Nat. Biotechnol.* **2008**, *26*, 1155–1160.

(5) Feist, A.; Palsson, B. *Nat. Biotechnol.* **2008**, *26*, 659–667.

(6) Savageau, M. *J. Theor. Biol.* **1970**, *26*, 215–226.

(7) Visser, D.; Heijnen, J. *Metab. Eng.* **2003**, *5*, 164–176.

(8) Visser, D.; Schmid, J.; Mauch, K.; Reuss, M.; Heijnen, J. *Metab. Eng.* **2004**, *6*, 378–390.

(9) Liebermeister, W.; Uhlendorf, J.; Klipp, E. *Bioinformatics* **2010**, *26*, 1528.

(10) Wittig, U.; Golebiewski, M.; Kania, R.; Krebs, O.; Mir, S.; Weidemann, A.; Anstein, S.; Saric, J.; Rojas, I. In *Data Integration in the Life Sciences*.; Springer: New York, 2006; Chapter SABIO-RK: integration and curation of reaction kinetics data.

(11) Schomburg, I.; Chang, A.; Ebeling, C.; Gremse, M.; Heldt, C.; Huhn, G.; Schomburg, D. *Nucleic Acids Res.* **2004**, *32*, D431 Database Issue.

(12) Goldberg, R. N. *J. Phys. Chem. Ref. Data* **1999**, *28*, 931.

(13) Goldberg, R. N.; Tewari, Y.; Bhat, T. *Bioinformatics* **2004**, *20*, 2874–2877.

(14) Alberty, R. *Biochem. Biophys. Acta* **1994**, *1207*, 1–11.

(15) Wegscheider, R. *Z. Phys. Chemie.* **1902**, *39*, 257–303.

(16) Liebermeister, W.; Klipp, E. *Theor. Biol. Med. Modell.* **2006**, *3*, 42.

(17) Gelman, A.; Carlin, J. B.; Stern, H. S.; Rubin, D. *Bayesian Data Analysis*; Chapman & Hall: New York, 1997.

(18) Hucka, M.; et al. *Bioinformatics* **2003**, *19*, 524.

(19) Alberty, R.; Cornish-Bowden, A. *Trends Biochem. Sci.* **1993**, *18*, 288–291.

(20) Alberty, R. *Arch. Biochem. Biophys.* **1998**, *353*, 116–130.

(21) Le Novere, N. *BMC Neuroscience* **2006**, *7*, S11.

(22) Teusink, B.; Passarge, J.; Reijenga, C.; Esgalhado, E.; van der Weijden, C.; Schepper, M.; Walsh, M.; Bakker, B.; van Dam, K.; Westerhoff, H.; Snoep, J. *Eur. J. Biochem.* **2000**, *267*, 5313–5329.

(23) Hynne, F.; Danø, S.; Sørensen, P. *Biophys. Chem.* **2001**, *94*, 121–163.

(24) Jiang, F.; Cram, D.; DeAizpurua, H.; Harrison, L. *Diabetes* **1999**, *48*, 722.

(25) Rizzi, M.; Baltes, M.; Theobald, U.; Reuss, M. *Biotechnol. Bioeng.* **1997**, *55*, 592–608.

(26) Hynne, F.; Danø, S.; Sørensen, P. *Biophys. Chem.* **2001**, *94*, 121–163.

(27) Kanehisa, M.; Goto, S. *Nucleic Acids Res.* **2000**, *28*, 27.

(28) Krause, F.; Uhlendorf, J.; Lubitz, T.; Klipp, E.; Liebermeister, W. *Bioinformatics* **2010**, *26*, 421–422.

(29) Huh, W. K.; Falvo, J. V.; Gerke, L. C.; Carroll, A. S.; Howson, R. W.; Weissman, J. S.; O'Shea, E. K. *Nature* **2003**, *425*, 686–691.

(30) Nissler, K.; Otto, A.; Schellenberger, W.; Hofmann, E. *Biochem. Biophys. Res. Commun.* **1983**, *111*, 294–300.

(31) Albe, K. R.; Butler, M. H.; Wright, B. E. *J. Theor. Biol.* **1990**, *143*, 163–195.

(32) Kümmel, A.; Panke, S.; Heinemann, M. *Mol. Syst. Biol.* **2006**, *2006*, 0034.

(33) Zamboni, N.; Kümmel, A.; Heinemann, M. *BMC Bioinf.* **2008**, *9*, 199.

(34) Beard, D.; Liang, S.; Qian, H. *Biophys. J.* **2002**, *83*, 79–86.

(35) Hoppe, A.; Hoffmann, S.; Holzhütter, H. *BMC Syst. Biol.* **2007**, *1*, 23.

(36) Rizzi, M.; Theobald, U.; Querfurth, E.; Rohrhirsch, T.; Baltes, M.; Reuss, M. *Biotechnol. Bioeng.* **1996**, *49*, 316–327.