A Multi-Scale Modeling Concept and Computational Tools for the Integrative Analysis of Stationary Metabolic Data

Eric von Lieres[1], Sören Petersen[2], Wolfgang Wiechert[3]

[1]Institute of Biotechnology, Research Centre Jülich, 52425 Jülich, Germany, e.von.lieres@fz-juelich.de (corresponding author).

[2]Degussa AG, 33790 Halle/Westfalen, Germany, soeren.petersen@degussa.com.

[3]Department of Simulation, University of Siegen, 57068 Siegen, Germany, wiechert@simtec.mb.uni-siegen.de.

Summary

Metabolic Engineering aims at the systematic analysis and targeted manipulation of the metabolism in biotechnologically utilized microorganisms [1,2]. Recently, consistent stationary *in vivo* data sets of intracellular metabolite concentrations, fluxes and specific enzyme activities have become available for this purpose [3,4]. For the integrative analysis of the metabolic data at hand, a novel multi-scale modeling concept, computer-algebraic methods and efficient numerical algorithms are proposed. The available metabolic data are typically afflicted with comparatively large measurement errors. Therefore, reliable comprehensive error estimations are essential for the reasonable interpretation of consecutive outcomes, such as simulation results.

The concepts, methods and algorithms are first presented as universal methods and subsequently applied to the anaplerotic regulation in lysine-producing *Corynebacterium glutamicum*. A multi-scale model is set up, fitted to the available experimental data and validated by the prediction of further experiments. This model is capable of forecasting the quantitative effect of changes in the specific activity of anaplerotic enzymes, namely phosphoenolpyruvate carboxylase, pyruvate carboxylase and phosphoenolpyruvate carboxykinase, on lysine productivity and yield.

1 Introduction

The metabolism of organisms in wildlife is regulated so as to synthesize intracellular metabolites mostly in the amounts needed for maintenance and growth, whereas metabolite overproduction would clearly be an evolutionary disadvantage. Systems Biology copes with the systemic comprehension of the cellular components and their interplay [5]. Here, as well as in the context of Metabolic Engineering, modeling and simulation play a major role [6]. In basic research, qualitative and quantitative knowledge of the cellular metabolism is gained. Industrial research emphasizes the optimization of cellular metabolism in order to improve the yield of biotechnological production processes.

Bacteria are widely used for the production of fine chemicals, such as the essential amino acid lysine. Essential amino acids are widely used as food and feed additives, since they cannot be produced by mammalian metabolism and therefore need to be supplied with nutrition [7]. Today, approximately 500.000 tons of lysine – worth more than 500 million Euros – are turned over yearly on the world marked [8]. Nevertheless, the margins are low and even small changes in the efficiency of production processes generally have a significant impact on profit [9].

1.1 Data Basis

The bacterium *Corynebacterium glutamicum* is employed for large scale lysine production [10]. The mutant MH20-22B excretes significant amounts of this amino acid into the extracellular environment [11]. Several metabolic pathways involved with lysine production in this organism, particularly the anaplerotic pathways, were examined experimentally in detail (figure 1). For this purpose, the cellular metabolism was adjusted to four different stationary states by imposition of definite genetic manipulations and environmental conditions in continuous culture. The phosphoenolpyruvate carboxykinase gene was deleted as well as overexpressed; and the growth rate was varied for the unmodified strain.

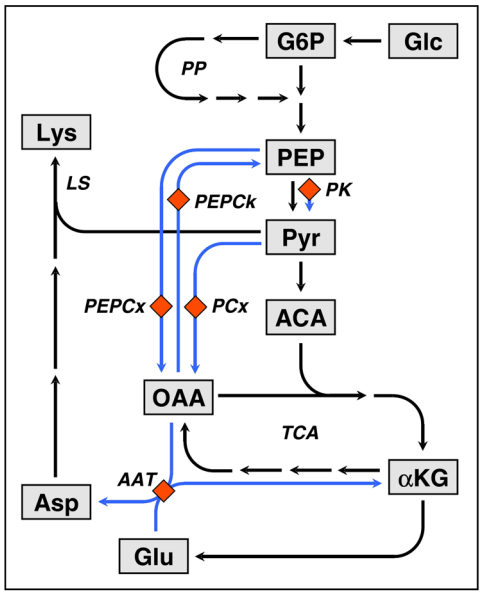

Figure 1: Anaplerosis and lysine synthesis in *Corynebacterium glutamicum*. Extracellular Metabolites: glucose (Glc) and lysine (Lys). Intracellular metabolites: glucose-6-phosphate (G6P), phosphoenolpyruvate (PEP), pyruvate (Pyr), oxaloacetate (OAA), aspartate (Asp), acetyl coenzyme A (ACA), α-ketoglutarate (aKG) and glutamate (Glu). Enzymes: pyruvate kinase (PK), phosphoenolpyruvate carboxylase (PEPCx), phosphoenolpyruvate carboxykinase (PEPCk), pyruvate carboxylase (PCx) and aspartate aminotransferase (AAT). Pathways: pentose phosphate pathway (PP), tricarboxylic acid cycle (TCA) and lysine synthesis (LS).

Each of the subsequent measurements was carried out with a sample from one of these cell cultures. This way, a data set of unique consistency and extent was created, covering intracellular metabolite concentrations, fluxes and specific enzyme activities [3,4]. Namely, the intracellular concentrations of α-ketoglutarate, aspartate, glutamate, pyruvate and phosphoenolpyruvate, the specific activities of aspartate aminotransferase, pyruvate kinase, phosphoenolpyruvate carboxykinase and carboxylase, the corresponding fluxes *in vivo* as well as the pyruvate carboxylase flux were determined (figure 1). The extracellular fluxes were derived from the cultivation data.

The experimental results are supplemented by data taken from literature. Today, the intracellular stoichiometry is well known [12]. However, the structure of the regulatory network is far from being entirely identified. On the other hand, the precursor metabolite demand for biomass synthesis is known with sufficient accuracy, whereas most kinetic parameters are only vaguely known. Despite the rapid progress of experimental methods, the data basis available today is still relatively small and inaccurate. Therefore, the proper integration and analysis of these data is a challenge for modeling and simulation.

1.2 Simulation Task

Generally, modeling and simulation are powerful tools, but they are pointless unless applied to specific questions [13]. Consequently, a typical case, the optimization of lysine production with *Corynebacterium glutamicum* MH20-22B, is worked out exemplarily. Lysine production extracts oxaloacetate from the citrate cycle (figure 1). The anaplerotic reactions compensate this drain by the conversion of pyruvate and phosphoenolpyruvate. The enzymes pyruvate and phosphoenolpyruvate carboxylase operate virtually irreversible *in vivo*. On the other hand, a significant reverse flux is catalyzed by phosphoenolpyruvate carboxykinase [3]. The latter lowers the concentration of the precursor oxaloacetate and evokes a so-called energy-consuming futile cycle.

The experiments show an increase or decrease in lysine productivity when the phosphoenol-pyruvate carboxykinase gene is deleted or overexpressed, respectively [3]. This observation gives rise to the hypothesis that the positive effect could be intensified by an amplification of the pyruvate or phosphoenolpyruvate carboxylase activities (figure 1). In the following, a modeling strategy and mathematical methods are presented, which are capable of testing this hypothesis quantitatively, based on the available data. The results are compared with the outcome of actual experiments.

2 Methods

Different concepts have been developed for modeling cellular metabolism, ranging from mechanistic (white-box) to phenomenologic (black-box) approaches [14-18]. In principle, mechanistic models can explain and predict the function of whole cells. However, one obtains complex systems with up to 1.000 rate and balance equations. Furthermore, precise mathe-matical formulae, including regulatory structures, are known for rather few enzymes [6]. Mechanistic models tend to include far too many parameters to be estimable from the available data. On the other hand, phenomenological models solely interpolate between observations and do not take underlying mechanisms into account. Hence, they are not able to explain or predict cellular function at all.

2.1 Multi-Scale Modeling Concept

For the integrative analysis of the available data for *Corynebacterium glutamicum*, we combine different modeling approaches [19,20]. The advantages of both mechanistic and phenomenologic approaches are merged into a multi-scale modeling concept (gray-box) while their disadvantages are avoided. This concept is not restricted to *Corynebacterium glutamicum* or lysine production. Indeed, the proposed approach is universally suitable for modeling cellular metabolism in stationary state, including the regulatory system, based on a given simulation task.

The multi-scale approach distinguishes from classical modeling concepts by the choice of a focus. Selected enzymes in this center of interest are mechanistically modeled in detail, while metabolic pathways outside the focus are simplified. Additionally, the kinetic rate expressions are related to each other by stoichiometric balance equations. Furthermore, the model is supplemented by phenomenological relations in case the rate and balance equations do not suffice to calculate all unknowns. Kinetic and stoichiometric parameters are taken from literature, as far as reliable *in vitro* data is obtainable. Finally, the remaining parameters are estimated as the model is fitted to the *in vivo* measurement data.

When choosing the model focus, it is crucial to describe the mechanisms responsible for effects under investigation as complete and detailed as possible. Moreover, the choice of

focus is influenced by data availability, since the model parameters must either be taken from literature or identified based on the measurement data at hand. The operation of focused enzymes is modeled rather accurately by kinetic expressions, whereas outside the focus, the function of metabolic pathways as a whole is described by much simpler formal kinetic expressions. Alternatively, terms of Biochemical Systems Theory or Cybernetic Modeling may be applied off focus [21,22].

Database mining serves to reveal phenomenological relations from experimental data, regardless of biological meaning. The relations need not be linear, but should be considered meaningful by an experienced biologist, to prevent correlations by chance. Phenomenological relations are primarily suited for the interpolation of measurement data. Extrapolation is rarely sensible and only if the existence of an appropriate mechanism is at least supposed. On the other hand, the multi-scale approach can be applied to study unknown mechanisms. For example, the enzyme kinetics may be extended by generic terms to investigate the regulatory influence of potential effectors.

The presented approach includes ranges of scales in many respects:

1) Rate equations are set up on different scales. Local metabolic functions in the focus are described in detail by mechanistic enzyme kinetics, whereas the global metabolic context outside the focus is represented by formal kinetic expressions and other approximations.

2) Balance equations are set up on different scales around metabolites, metabolic branches and the whole system. They serve to embed the rate equations in the global metabolic context and to connect them to the extracellular fluxes.

3) Generally, the focus of interest is modeled reasonably detailed, while the metabolic context is merely approximated or described phenomenologically. This way, local regulatory structures are related to global system properties, such as lysine productivity.

In comparison with mechanistic models for whole cells, multi-scale models are less complex and include fewer parameters. Nevertheless, for a specific simulation task, all relevant metabolic functions are taken into account and, according to significance and data availability, modeled in detail, approximated or described phenomenologically. Multi-scale models are capable of predicting global effects of local changes in the model focus.

2.2 Equation Preprocessing

Multi-scale models consist of linear stoichiometric equations, nonlinear kinetic equations and unstructured phenomenological relations. A complete set of model equations defines the variables, representing fluxes and metabolite concentrations, uniquely as a typically implicit function of the parameters, which are known a priori or to be estimated as described below. The specific enzyme activities are handled as known parameters. Throughout parameter estimation, the model equations are solved repeatedly, hence, their necessarily numeric solution ought to be fast.

The numerical performance is improved by computer-algebraic processing of the implicit nonlinear equation system. Advantage is taken of the equations' known structure. First, we eliminate as many explicit variables as possible, such as fluxes defined by kinetic expressions from the equation system. Second, the linear stoichiometry is resolved for further fluxes, which are then explicit and are as well eliminated from the system. Third, the resulting equations are automatically simplified. The remaining variables are implicit and to be determined numerically.

For given parameters, the implicit equations serve to calculate closing conditions, which are zero for the correct implicit variables. The iterative solution of these equations may be carried out by optimization, minimizing the absolute closing conditions. For given parameters and implicit variables, the explicit equations, which were eliminated on the first hand, serve to calculate the remaining variables. Finally, simulated and real measurements are compared in order to identify the unknown parameters (figure 2).

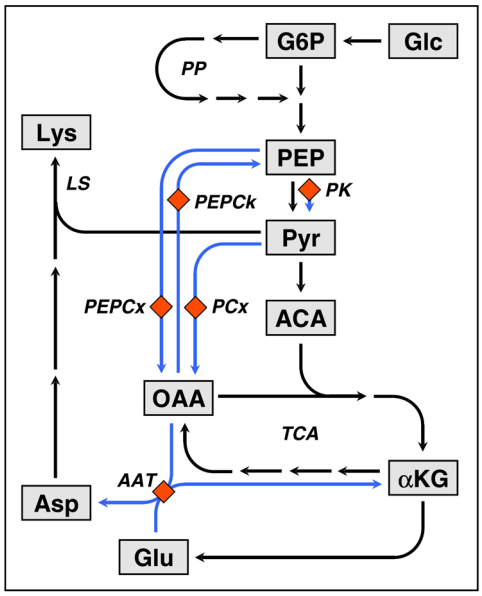

Figure 2: Nested iterations are caused by the iterative solution of implicit model equations and the estimation of unknown parameters from measurement data. Besides the measurement data, known parameters and initial values must be provided.

2.3 Parameter Identification

The model is fitted simultaneously to several data sets stemming from different metabolic states (figure 3). This way, it is possible to estimate kinetic information from stationary data. It is advisable to weight the individual residuals with estimations of the corresponding measurement errors, since the experimental data are typically of heterogeneous quality. In most cases, least squares estimation is adequate for parameter identification, while sophisticated estimators are occasionally helpful to reduce the influence of outliers [23]. Alternatively, the data may be tested for outliers by cross validation.

Nested iterations are caused by straightforward parameter estimation with a subordinated numerical model solution. For each step of the optimization process, several instances of the model are entirely solved, which is computationally inefficient. The overall performance is significantly improved by a fusion of these iterative processes, which is achieved by penalized optimization. The absolute closing conditions of all model instances are added to the residual for parameter optimization. To speed up convergence, the closing conditions are emphasized by a technical factor. The residual's optimal value is essentially not influenced by this method.

Initial values must be at hand for the implicit variables and unknown parameters. At the beginning of the denested procedure, the simulated measurements and consequently, the calculated residuals are no more than rough approximations, since the initial values do not solve the model equations. Nevertheless, these intermediate results are sufficient for the first optimization steps. As the optimization proceeds, all model solutions are approximated with substantial accuracy. Thus, the parameters are identified for the numerically solved model.

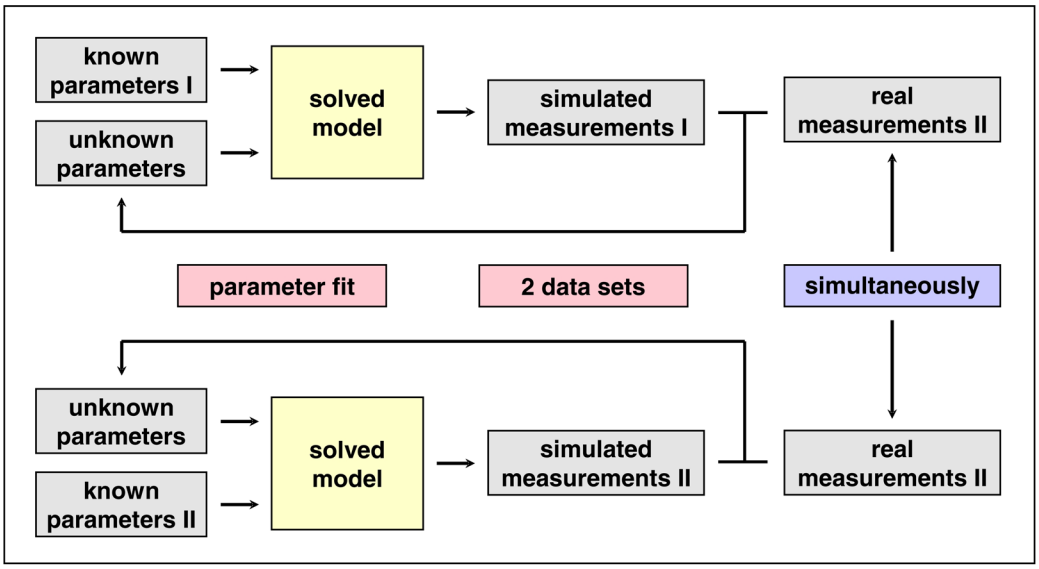

Figure 3: The parameters are identified by simultaneous fit of several model instances to different sets of stationary metabolic data. The model instances are adjusted to the metabolic states by the known parameters, such as growth rate and specific enzyme activities. For *Corynebacterium glutamicum* MH20-22B, four data sets are at hand.

Initial values are usually estimated from experimental data or adopted from literature. They typically range one order of magnitude around the sought parameters. Accordingly, the optimum is first approximated using the gradient method. This approximation is then refined with Newton's method [24]. The convergence is considerably enhanced by a transformation of the unknowns on a logarithmic scale. This way, the optimizers sensitivity is automatically adjusted to the parameters' orders of magnitude. Furthermore, the search space is restricted to positive, i.e., biologically meaningful, values.

2.4 Sensitivity Analysis

The calculation of sensitivities is a principal tool for system analysis [25]. First, the model equations are derived implicitly with respect to the parameters. The resulting sensitivities describe the variables response to small parameter changes. They are employed for parameter identification, control analysis and the prediction of further experiments. Second, the control coefficients are once more derived, with respect to the data and parameters, for the purpose of error estimation. Third, again for error estimation, the derivatives of the estimated parameters, with respect to the experimental data and known parameters, are calculated by subsequent explicit and implicit derivation of the residual for parameter estimation.

By the aid of computer-algebra, all occurring derivatives are calculated symbolically. This approach gets along without discretization and demands less computational power than numerical methods, which are equally accurate for suitable step widths. However, the determination of a suitable step width is computationally expensive and numerical derivation is ill posed. Hence, numerical noise is amplified disproportionately for small step widths [26].

2.5 Error Estimation

Metabolic data are typically heterogeneous and commonly afflicted with large errors. In other words, data quality is not at its best. The model propagates errors from the experimental data and known parameters to the identified parameters and further to the variables and control coefficients. Thus, comprehensive error estimation is indispensable for the serious inter-

pretation of simulation results. Estimated errors are reasonably specified as standard deviations. Although the accuracy of measurements is usually known, literature data are frequently taken for precise knowledge. However, the data basis should be supplemented by realistic error estimations, because neglect of errors leads to unreliable results.

Errors estimations are computed not only for the identified parameters and matching variables, but also for the control coefficients and for the simulations of further experiments. The consideration of imprecisely "known" parameters increases computational complexity, but helps to avoid overoptimistic results. The errors are estimated by linearization of the model equations using symbolic derivatives and by Monte Carlo simulation. Unlike the Gaussian law, Monte Carlo methods take account of the nonlinear model structure. Both results are compared.

Monte Carlo simulation is a statistical and computationally expensive method. Both experimental data and known parameters are interpreted as samples of normally distributed quantities. The mean values are estimated by the data itself, while the standard deviations are known from error estimation. According to these distributions, different sets of random data are generated. Next, the parameters are identified for each data set and from the distributed results, and mean values and standard deviations are computed.

2.6 Implementation

The methods and algorithms are implemented prototypically in the computer-algebra system Mathematica [27]. Separate modules are created for model solution, sensitivity analysis, parameter identification, error estimation and statistical simulation. The experimental and literature data, as well as definitions of the variables, parameters and model equations, are filed centrally. Thus, changes of the data and model structure are easily realized.

Computer-algebra systems feature sophisticated algorithms for equation manipulation, including symbolic derivation and equation simplification, while common numerical routines are provided as well. As interpreted languages, they are particularly well suited for prototype design, where the speed of numerical computations is of subordinated priority. Furthermore, a wide range of routines is provided for the graphical representation of simulation results.

3 Results

A multi-scale model is set up and applied to integrate the data at hand for lysine producing *Corynebacterium glutamicum*. The individual equations, variables, parameters, experimental and literature data are exhaustively described in a previous publication [20]. A rephrasal of the entire modeling process is not in the scope of this contribution. The differences in identified parameters, estimated errors and subsequent results are due to an enlargement of the utilized data basis and the supposition of realistic errors for all data from literature.

3.1 Multi-Scale Model

The model focus is chosen to embrace pyruvate kinase, phosphoenolpyruvate carboxylase, pyruvate carboxylase, phosphoenolpyruvate carboxykinase and aspartate aminotransferase (figure 1), which are modeled mechanistically by enzyme kinetics. In *Corynebacterium glutamicum* MH20-22B, lysine synthesis is highly deregulated and hence described en bloc by a formal kinetic expression. Balance equations are set up for: phosphoenolpyruvate, pyruvate, aspartate, lysine, acetyl coenzyme A and NADPH pools, for the tricarboxylic acid cycle with oxaloacetate, the pentose phosphate pathway with upper glycolysis and for the entire cell.

Phenomenological relations are established for pyruvate and oxaloacetate, α-ketoglutarate and oxaloacetate, as well as glutamate.

To ensure parameter identifiability, the resulting model is slightly modified. The final model interconnects 20 variables by 20 equations, 17 of which are explicitly solved and algebraically eliminated prior to numerical model solution. 12 variables are related to intracellular metabolite concentrations and fluxes, for which experimental data is available for 4 metabolic states. From this data, 15 unknown parameters are estimated (figure 4), while 21 parameters are taken from literature.

Further experimental data are available for the growth rate and 4 specific enzyme activities. They are used as known parameters for each of the 4 model instances. Each metabolic state is characterized by only four known parameters, namely the growth rate and specific activities of phosphoenolpyruvate carboxykinase, pyruvate kinase and aspartate aminotransferase, while the specific activity of phosphoenolpyruvate carboxylase does not vary significantly.

3.2 Parameter Identification

For the estimation of initial values, isolated model equations are first individually fitted to the experimental data. The model parameters are then calculated by optimization and by using the Monte Carlo method (figure 4). In the first case, propagated errors are estimated according to the Gaussian law. The $\chi 2$-test is passed with 84 percent confidence. In the second case, mean values and standard deviations are determined statistically. Samples outside a physiologically reasonable region in parameter space are rejected.

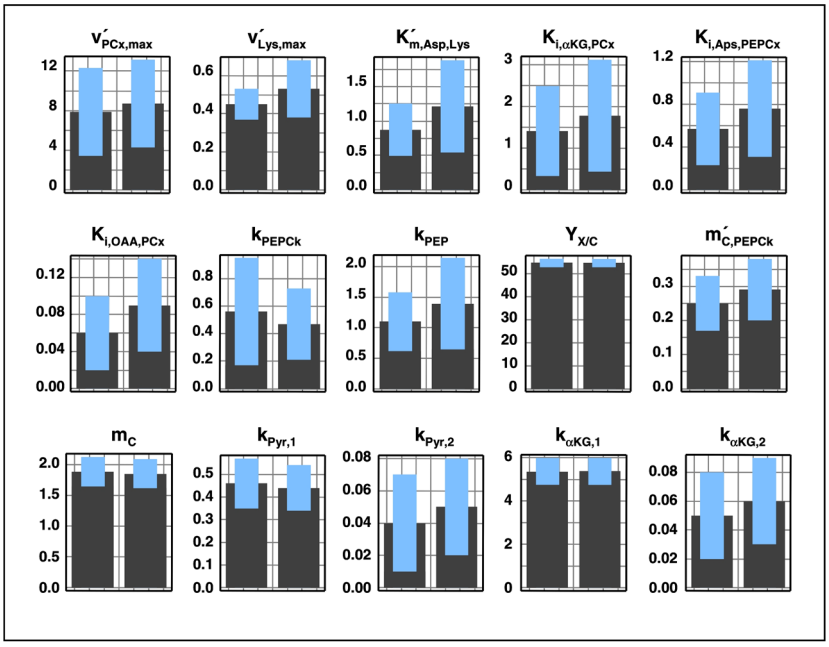

Figure 4: The model parameters and error estimations as determined by a single parameter identification (left bars) and Monte Carlo simulation (right bars). Detailed descriptions of the individual parameters are given in a previous publication [20].

Figure 4 depicts mean values and standard deviations as estimated by different mathematical methods. The variation of most mean values is insignificant against the background of the corresponding error estimations. Nevertheless, errors estimated by Monte Carlo simulation, with one plain exception, tend to be significantly larger than estimated by the Gaussian law, which is in all likelihood due to the nonlinear model structure.

3.3 Control Analysis

In this section, control coefficients denote derivatives of simulation outcomes with respect to model parameters. They are normalized to quote causes and effects as percentages of the corresponding absolute values. Productivity refers to the flux of product excretion, whereas yield denotes the quotient of product excretion and substrate uptake fluxes. The optimization of product yield is a common objective of metabolic engineering, since many manipulations afflict the affected organism's ability to grow. Control coefficients provide valuable information about regulation *in vivo* and indicate targets for strain improvement, albeit that metabolic regulation is typically distributed over several enzymes [28,29].

Control analysis gives first answers to the question of how lysine synthesis is influenced by the anaplerotic reactions (table 1). As supposed, lysine productivity and yield can be enhanced by increasing the specific activity of phosphoenolpyruvate carboxylase or pyruvate carboxylase. Conversely, high specific activities of phosphoenolpyruvate carboxykinase diminish lysine production, as already known and reproduced by the corresponding control coefficients. The estimated errors are notably small, even though rather large errors are assumed for the underlying literature data.

Enzyme	Productivity	Yield
PEPCk	- 0.17 ± 0.02	- 0.16 ± 0.02
PEPCx	0.10 ± 0.02	0.07 ± 0.01
PCx	0.17 ± 0.04	0.12 ± 0.03

Table 1: Normalized control coefficients for lysine productivity and yield with respect to the specific enzyme activities of phosphoenolpyruvate carboxykinase (PEPCk), phosphoenol-pyruvate carboxylase (PEPCx) and pyruvate carboxylase (PCx). The errors are estimated according to the Gaussian law.

By control analysis, the model is validated qualitatively with regard to its predefined objective and focus. However, control coefficients describe local system behavior and infinitesimal alterations of cellular metabolism are experimentally not feasible. Hence, the model is validated quantitatively by comparison of virtual and real experiments.

3.4 Virtual Experiments

Once the known and identified parameters are determined, the model is ready for virtual experiments. Real experiments may influence the specific enzyme activities indirectly, e.g., by changing gene expression levels, whereas virtual experiments are realized by direct alterations of the corresponding activities. Virtual experiments are used to predict global effects of finite alterations within the model focus. The comparability of virtual and real experiments is guaranteed by experimental determination of the operative factors for specific enzyme activity alterations. In literature, data are available for lysine producing *Corynebacterium glutamicum*, featuring overexpressed and deleted phosphoenolpyruvate carboxylase and pyruvate carboxylase genes (table 2).

Enzyme	Factor	Productivity	Yield	Reference
PEPCx	7.5	+ 32 ± 7	+ 19 ± 4	+ 23 ± 8
PEPCx	0	- 14 ± 3	- 10 ± 2	± 0
PCx	11	+ 62 ± 23	+ 31 ± 10	+ 47 ± 10
PCx	0	- 42 ± 30	- 35 ± 30	- 59 ± 4

Table 2: Virtual experiments (productivity and yield) in comparison with real experiments (references [30-32]) in which phosphoenolpyruvate carboxylase (PEPCx) and pyruvate carboxylase (PCx) encoding genes were deleted and overexpressed. Factor indicates the operative alteration of the specific activity for the corresponding enzyme. Simulated and measured effects are stated in percentages. The errors are estimated according to the Gaussian law.

The four real experiments described in literature were carried out as batch cultures, which were stopped after a predefined time to determine the extracellular lysine concentration. Although neither yield nor productivity values are explicitly provided, comparison with model predictions is possible. If the carbon source was utterly metabolized, the change in extracellular lysine concentration would be equivalent to the change in yield. If the carbon source remained partly unmetabolized, the measured lysine concentration would rather be correlated with productivity. Consequently, the experimental data are expected to rest between the predictions, which is the case for both overexpression experiments. The deletion experiment for phosphoenolpyruvate carboxylase showed no determinable effect, but with unpublished accuracy. The pyruvate carboxylase deletion experiment is predicted within the range of the comparably large error estimations.

4 Discussion

We present a novel modeling concept and associated computational methods for the integrative analysis of stationary metabolic data. Based on experimental *in vivo* data for lysine producing *Corynebacterium glutamicum*, the influence of anaplerotic regulation on product formation is predicted quantitatively.

4.1 Modeling Concept

The multi-scale modeling approach is designed for the integration of metabolic data stemming from stationary *in vivo* experiments and literature. The stipulation of a focus is normally consistent with the objectives of modeling and simulation. The focused enzymes are modeled mechanistically and embedded in the global metabolic context via the stoichiometry. This approach permits parameter identification based on relatively few, inaccurate and heterogeneous data sets. Furthermore, it is possible to predict global effects of changes in the model focus. The predictive power of multi-scale models is hardly restricted by the limited use of phenomenological relations.

The modeling process is naturally subdivided as follows: Initially, the biological objective and a matching focus need to be defined. Next, the mechanistic rate and balance equations are set up. Then, the metabolic context is described much less detailed, using formal kinetic and other approximate expressions. Finally, the remaining degrees of freedom are eliminated by phenomenological relations.

4.2 Computational Tools

The computational effort for a numerical model solution is considerably reduced by the aid of computer algebra, condensing the model from 20 to 3 implicit equations for each instance. The efficiency of parameter identification is significantly improved by fusion of all iterative processes and transformation of all unknowns on logarithmic scale. Optimization methods and Monte Carlo simulation are then used for parameter identification and error estimation. The Monte Carlo method is computationally expensive, but has the benefit of exploring non-linear system behavior.

Many of the presented methods, particularly the symbolic determination of various sensitivities, would be unfeasible without the aid of modern computer-algebra. Today, the feasibility of computer-algebraic methods, especially implicit derivation, for complex non-linear equation systems is limited by the available memory rather than by processor performance.

4.3 Predictive Simulation

A multi-scale model is created specifically to explore the impact of anaplerotic regulation in *Corynebacterium glutamicum* on lysine production. This model is validated qualitatively by control analysis and quantitatively by prediction of effects caused by finite changes in the specific activities of anaplerotic enzymes, namely phosphoenolpyruvate carboxylase and pyruvate carboxylase, on lysine productivity and yield.

The supplementary data employed for model validation are naturally not considered for parameter identification. Besides model validation, virtual experiments are functional for the exploration of system behavior and for optimal design of further experiments [6]. By the aid of appropriate models, experiments are carried out much faster and are less expensive *in silico* than *in vivo*.

4.4 Outlook

A promising alternative approach for the integration of metabolic data is provided by Bayes statistics. In this framework, it is possible to exploit vague a priori knowledge of the "unknown" parameters, such as their sign and order of magnitude, in addition to the customary data. A posteriori distributions for the parameters are determined by the Markov Chain Monte Carlo method, substituting parameter optimization. Furthermore, the nonlinearly propagated errors are estimated statistically.

Another important issue is identifiability: For mathematical reasons, individual parameters could possibly be unidentifiable. Hence, it is advisable to analyze parameter identifiability prior to and during optimization. Both topics will be addressed in forthcoming publications.

4.5 Conclusions

In this contribution, the entire course from the development of a specific biological objective over modeling and method design to simulation and model validation is systematically worked out. A multi-scale approach is proposed for the integrative analysis of stationary metabolic data. Methods and algorithms for the systematic solution of implicit nonlinear model equations, symbolic determination of sensitivities, efficient parameter identification and error estimation are presented.

The framework is applied to the anaplerotic regulation in lysine producing *Corynebacterium glutamicum*. The available experimental and literature data are integrated with a multi-scale

model, which is validated by comparison of virtual and real experiments. The specific activities of focused anaplerotic enzymes are correlated with lysine productivity and yield. It is shown that quantitative simulations are possible even against the background of rather defensive error estimations.

5 References

[1] J. E. Bailey. Towards a Science of Metabolic Engineering. Science, 252:668-1674, 1991.

[2] G. N. Stephanopoulos, A. A. Aristidou and J. Nielsen. Metabolic Engineering: Principles and Methodologies. New York: Academic Press, 1998.

[3] S. Petersen, A. A. de Graaf, L. Eggeling, M. Möllney, W. Wiechert and H. Sahm. In vivo Quantification of Parallel and Bidirectional Fluxes in the Anaplerosis of Corynebacterium glutamicum. Journal of Biological Chemistry, 275(46):35932-35941, 2000.

[4] S. Petersen, C. Mack, A. A. de Graaf, C. Riedel, J. Eikmanns and H. Sahm. Metabolic Consequences of Altered Phosphoenolpyruvate Carboxylase Activity in Corynebacterium glutamicum Reveal Anaplerotic Regulation Mechanism In Vivo. Metabolic Engineering, 3:344-361, 2001.

[5] H. Kitano. Perspective on Systems Biology. New Generation Computing, 18:199-216, 2000.

[6] W. Wiechert. Modeling and Simulation: Tools for Metabolic Engineering. Journal of Biotechnology, 94:37-63, 2002.

[7] D. L. Nelson and M. M. Cox. Lehninger Principles of Biochemistry, New York: Worth Publishers, 2000.

[8] K. H. Drauz, B. Hoppe, A. Kleemann, H.-P. Krimmer, W. Leuchtenberger and C. Weckbecker. Amino Acids. in Ullmann's Encyclopedia of Industrial Chemistry (6th Edition, Electronic Release), Weinheim: Wiley - VCH, 2002.

[9] L. Eggeling and H. Sahm. L-Glutamate and L-Lysine: Traditional Products with Impetuous Developments. Applied Microbiology and Biotechnology, 52:146-153, 1999.

[10] W. Liebl. The genus Corynebacterium - nonmedical. 1157-1171 in A. Balows, H. G. Trüper, M. Dworkin, W. Harder and K. H. Schleifer (Editors). The Prokaryotes. Berlin: Springer Verlag, 1992.

[11] L. Eggeling. Biology of L-Lysine overproduction by Corynebacterium glutamicum. Aminoacids, 6:261-272, 1994.

[12] G. Michal (Editor). Biochemical Pathways. Heidelberg: Spektrum Akademischer Verlag, 1999.

[13] J. E. Bailey. Life is complicated. 41-47 in A. J. Cornish-Bowden and M L. Cárdenas (Editors). Technological and Medical Implications of Metabolic Control Analysis, Dordrecht: Kluwer Academic Publishers, 2000.

[14] M. M. Domach, S. K. Leung, R. E. Cahn, G. G. Cocks and M. L. Schuler. Computer Model for Glucose-Limited Growth of a Single Cell of Escherichia coli B/r-A. Biotechnology and Bioengineering, 26:203-216, 1984.

[15] J. Schaff, C. Fink, B. Slepchenko, J. Carson and L. Loew. A General Computational Framework for Modeling Cellular Structure and Function. Biophysical Journal, 73:1135-1146, 1997.

[16] H. J. L. van Can, H. A. B. te Braake, C. Hellinga, K. C. A. M. Luyben and J. J. Heijnen. An Efficient Model Development Strategy for Bioprocesses Based on Neural Networks in Macroscopic Balances. Biotechnology and Bioengineering, 54(6):549-566, 1997.

[17] M. Tomita, K. Hashimoto, K. Takahashi, T. S. Shimizu, Y. Matsuzaki, F. Miyoshi, K. Saito, S. Tanida, K. Yugi, J. C. Venter and C. A. Hutchison. E-CELL: Software Environment for Whole-Cell Simulation. Bioinformatics, 15:72-84, 1999.

[18] B. Teusink, J. Passarge, C. A. Reijenga, E. Esgalhado, C. C. van der Weijden, M. Schepper, M. C. Walsh, B. M. Bakker, K. van Dam, H. V. Westerhoff and J. L. Snoep. Can Yeast Glycolysis be Understood in Terms of In Vitro Kinetics of the Constituent Enzymes? Testing Biochemistry. European Journal of Biochemistry, 267:5313-5329, 2000.

[19] E. von Lieres, S. Petersen and W. Wiechert. Integrating Fluxes, Metabolite Concentrations and Enzyme Activities for the Predictive Modeling of Metabolic Regulation. Castelvecchio Pascoli: Metabolic Engineering IV (Workshop on Computational Methods), 2002.

[20] S. Petersen, E. von Lieres, A. A. de Graaf, H. Sahm and W. Wiechert. A Multi-Scale Approach for the Predictive Modeling of Metabolic Regulation. to appear in B. N. Kholodenko, H. V. Westerhoff (Editors). Metabolic Engineering in the Post Genomic Era. Wymondham: Horizon Scientific Press, 2004.

[21] E. O. Voit (Editor). Canonical Nonlinear Modeling: S-System Approach to Understanding Complexity. New York: Reinhold van Nostrand, 1991.

[22] J. Varner and D. Ramkrishna. Metabolic Engineering from a Cybernetic Perspective. 1. Theoretical Preliminaries. Biotechnology Progress, 15:407-425, 1999.

[23] D. Birkes and Y. Dodge. Alternative Methods of Regression. New York: John Wiley and Sons, 1993.

[24] H. M. Antia. Numerical Methods for Scientists and Engineers. Basel: Birkhäuser Verlag, 2002.

[25] A. Satelli, K. Chan and E. M. Scott (Editors). Sensitivity Analysis. New York: John Wiley and Sons, 2000.

[26] A. K. Louis. Inverse und Schlecht Gestellte Probleme. Stuttgart: Teubner, 1989.

[27] S. Wolfram. The Mathematica Book. Cambridge: Cambridge University Press, 1999.

[28] R. Heinrich and S. Schuster. The Regulation of Cellular Systems. New York: Chapman and Hall, 1996.

[29] D. A. Fell. Understanding the Control of Metabolism. London: Portland Press, 1997.

[30] J. Cremer, L. Eggeling and H. Sahm. Control of the Lysine Biosynthesis Sequence in Corynebacterium glutamicum as Analyzed by Overexpression of the Individual Corresponding Genes. Applied and Environmental Microbiology, 57:1746-1752, 1991.

[31] P. G. Peters-Wendisch, B. J. Eikmanns, G. Thierbach, B. Bachmann and H. Sahm. Phosphoenolpyruvate Carboxylase in Corynebacterium glutamicum is Dispensable for Growth and Lysine Production. FEMS Microbiology Letters, 112:269-274, 1993.

[32] P. G. Peters-Wendisch, B. Schiel, V. F. Wendisch, E. Katsoulidis, B. Möckel, H. Sahm and B. J. Eikmanns. Pyruvate carboxylase is a Major Bottleneck for Glutamate and Lysine Production by Corynebacterium glutamicum. Journal of Molecular Microbiology and Biotechnology, 3(2):295-300, 2001.