Benchmarking Evolutionary Algorithms on Convenience Kinetics Models of the Valine and Leucine Biosynthesis in *C. glutamicum*

Andreas Dräger*‡ Marcel Kronfeld* Jochen Supper* Hannes Planatscher* Jørgen B. Magnus† Marco Oldiges† Andreas Zell*

***Abstract*— An important problem in systems biology is parameter estimation for biochemical system models. Our work concentrates on the metabolic subnetwork of the valine and leucine biosynthesis in *Corynebacterium glutamicum*, an anaerobic actinobacterium of high biotechnological importance. Using data of an *in vivo* experiment measuring 13 metabolites during a glucose stimulus-response experiment we investigate the performance of various Evolutionary Algorithms on the parameter inference problem in biochemical modeling. Due to the inconclusive information on the reversibility of the reactions in the pathway, we develop both a reversible and an irreversible differential equation model based on the recent convenience kinetics approach. As the reversible model allows better approximation on the whole, we use it to analyze the impact of different settings on four especially promising EAs. We show that Particle Swarm Optimization as well as Differential Evolution are useful methods for parameter estimation on convenience kinetics models outperforming Genetic Algorithm and Evolution Strategy approaches and nearly reaching the quality of independent spline approximations on the raw data.**

I. Introduction

A rife task in systems biology is the identification and inference of metabolic network models. In many studies cellular reaction systems were translated into sets of differential equations to derive network properties, predict long term behaviour or perform steady-state analysis [1]–[5].

To construct a mathematical model of a biochemical reaction network, many formalisms have been released based on the generalized mass-action rate law [6]. Any such model contains parameters to be fitted to the available data, which can be generated experimentally or *in silico*. Despite phenomenological models, the parameter values can be measured *in vitro*. However, this procedure is time consuming, expensive and often impractical. The application of computational methods to optimize the model parameters regarding the fit error has therefore become a challenging task in the model identification process. The approach is based on the assumption that, during evolution, enzymes were optimized for specific environmental demands. Thus, the model parameters represent the optimal thermodynamical enzyme properties.

A very recent approach to model biochemical reaction networks is the so called *convenience kinetics*, which describes a default rate equation for cases in which the exact reaction mechanism remains unknown [7]. This is the case for many larger networks available in databases like KEGG or MetaCyc [8], [9]. Based on the trimolecular random order Michaelis-Menten formalism the convenience kinetics gives an approximation of the exact process. The linlog kinetics [10], another approximative formalism, was applied to model the L-valine (Val) and L-leucine (Leu) metabolism in the industrial producer species *C. glutamicum* by Magnus *et al.* [4]. Due to the economic importance of this pathway and the high quality of the measured data, this model system provides a valuable reference for further investigations on the applicability of different modeling approaches and optimization procedures.

Evolutionary Algorithms are known to handle highly nonlinear optimization problems and have been applied sucessfully to biochemical systems in several studies [11], [12]. In this work we constructed two alternative seven dimensional systems of differential equations based on the convenience kinetics formalism and applied the following EAs to optimize the model parameters: multi-start Hill Climbers, Simulated Annealing, binary and real valued Genetic Algorithm, standard and covariance matrix adaption Evolution Strategy, Particle Swarm Optimization and Differential Evolution. Standard settings of these optimization methods were explored and varied systematically on the model systems aiming for determining the most appropriate approach with its specific settings to obtain suitable parameters for metabolic networks based on convenience kinetics and *in vivo* data.

II. Methods

A. The System under Study

1) The Biochemical Model: Fig. 1 shows the biochemical reactions of the Val/Leu biosynthesis in *C. glutamicum* according to the MetaCyc database.

Our consideration of the pathway starts with pyruvate (Pyr) which is consumed to form 2-ketoisovalerate (KIV) in two reaction steps. There are two different ways to form Val and one to convert KIV to 2-isopropylmalate (2IPM). The latter is the starting substance for the Leu production in four following reaction steps. Both Val and Leu can be used for biomass production or can be pumped out of the cell if not needed. Here we only consider the industrially interesting transport. In four feedback loops Val and Leu downregulate their own production rate. The transport of Leu and Val across the cell wall is actually performed by the same enzyme, so that both substrates compete with each other. However, for modeling purposes two distinct reactions are necessary in which the competition is included as inhibition.

*Center for Bioinformatics Tübingen (ZBIT), 72076 Tübingen, Germany
†Forschungszentrum Jülich, Institute of Biotechnology, Germany
‡Corresponding author, andreas.draeger@uni-tuebingen.de

1-4244-1340-0/07/$25.00

Since the reaction $2\,\mathrm{IPM} \rightleftharpoons 3\,\mathrm{IPM}$ is fast, it is assumed to be in equilibrium. This and the two following reactions $3\,\mathrm{IPM} + \mathrm{NAD}^+ \longrightarrow 2\,\mathrm{I_3OS} + \mathrm{NADH_2}$ as well as (2S)-2-isopropyl-3-oxosuccinate ($2\,\mathrm{I_3OS}$) $\longrightarrow$ 2-ketoisocaproate (KIC) $+\mathrm{CO_2}$ that only depend on the concentration of 2 IPM were lumped together introducing the symbol IPM for both derivates. The KEGG database mentions two additional reaction steps not included in METACYC: Pyr reacts to 2-hydroxyethyl-thio-dipyrophosphate first before forming (S)-2-acetolactate (AcLac) which then turns over in 3-hydroxy-3-methyl-2-oxobutanoate before it further reacts to (R)-2,3-dihydroxy-3-metylbutanoate (DHIV).

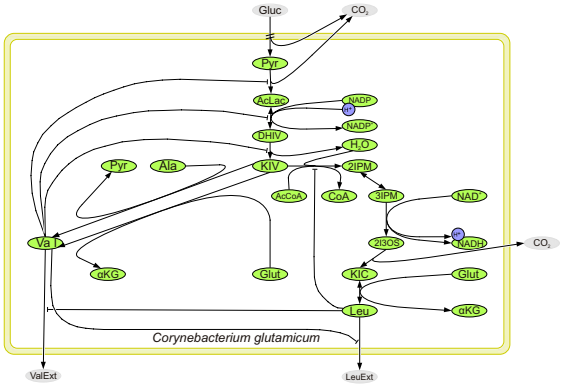


Fig. 1. Process diagram of the Val and Leu synthesis in *C. glutamicum* Metabolites outside the cell are not directly included in the model system. Both amino acids can be formed from Pyr, the end product of the glycolysis.

TABLE I

THE REACTION SYSTEM IN MORE DETAIL

The reactions in KEGG were lumped together to result in this reaction scheme which is in accordance with METACYC besides the question of irreversibility, apart from R_2 and R_9, which are reversible.

No.	Reaction	Enzyme	Inhib.
R_1	$2\,\mathrm{Pyr} \longrightarrow \mathrm{AcLac} + \mathrm{CO_2}$	AHAS	Val
R_2	$\mathrm{AcLac} + \mathrm{NADPH_2} \rightleftharpoons \mathrm{DHIV} + \mathrm{NADP^+}$	AHAIR	Val
R_3	$\mathrm{DHIV} \longrightarrow \mathrm{KIV} + \mathrm{H_2O}$	DHAD	Val
R_4	$\mathrm{KIV} + \mathrm{Glut} \longrightarrow \mathrm{Val} + \alpha\mathrm{KG}$	$\mathrm{BCAAT_{ValB}}$	
R_5	$\mathrm{KIV} + \mathrm{Ala} \longrightarrow \mathrm{Val} + \mathrm{Pyr}$	$\mathrm{BCAAT_{ValC}}$	
R_6	$\mathrm{Val} \longrightarrow \mathrm{Val_{ext}}$	$\mathrm{Trans_{Val}}$	Leu
R_7	$\mathrm{KIV} + \mathrm{AcCoA} \longrightarrow \mathrm{IPM} + \mathrm{CoA}$	IPMS	Leu
R_8	$\mathrm{IPM} + \mathrm{NAD^+} \longrightarrow \mathrm{KIC} + \mathrm{NADH_2} + \mathrm{CO_2}$	IPMDH	
R_9	$\mathrm{KIC} + \mathrm{Glut} \rightleftharpoons \mathrm{Leu} + \alpha\mathrm{KG}$	$\mathrm{BCAAT_{LeuB}}$	
R_{10}	$\mathrm{Leu} \longrightarrow \mathrm{Leu_{ext}}$	$\mathrm{Trans_{Leu}}$	Val

2) Glucose Stimulus-Response Experiment: A glucose shock was caused after a 10 min starvation period by adding this central nutrient to the culture medium as described in detail by Magnus *et al.* [4]. Over a time span of 25 s, beginning 4 s before the glucose pulse, 47 samples were taken for 13 metabolites on the pathway starting at the state of Pyr. For technical reasons $\mathrm{NADH_2}$ and $\mathrm{NADPH_2}$ as well as AcetylCoA and CoA could not be measured with a high degree of exactness. Thus, Magnus *et al.* [4] suggested to take into account that both couples $\mathrm{NAD^+}$ and $\mathrm{NADH_2}$ as well as $\mathrm{NADP^+}$ and $\mathrm{NADPH_2}$ follow a conservation relation so that the total amount of both coupled metabolites remains constant during the 25 s of interest. Thus, $\mathrm{NADH_2} = 0.8\,\mathrm{mM} - [\mathrm{NAD^+}]$ and $[\mathrm{NADPH_2}] = 0.04\,\mathrm{mM} - [\mathrm{NADP^+}]$. We assume a constant pool of the other two central metabolites, which does not vary over the considered time span. The steady-state concentrations of the seven metabolites to be simulated [4] serve as initial values for the models.

B. Mathematical Models

We benchmarked the following two alternative modeling approaches on this pathway, which will be explained in more detail in the following sections:

1) Convenience kinetics with three Michaelis-Menten equations (CK MM), reversible
2) CK MM, irreversible

The change of the metabolite concentrations over time can be calculated by linear combination of the stochiometric matrix $\mathbf{N}$ describing the topology of the reaction system (Tab. I) with the vector of reaction velocities $\mathbf{v}$ that depends on the vector of reacting species $\mathbf{S}$ and the parameter vector $\mathbf{p}$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{S} = \mathbf{N}\mathbf{v}(\mathbf{S}(t), t, \mathbf{p}).$$

For the Val/Leu biosynthesis this yields the equations listed in Tab. II.

TABLE II

LINEAR COMBINATION OF THE REACTION VELOCITIES

$$\begin{array}{lcllcl} \frac{\mathrm{d}}{\mathrm{d}t}[\mathrm{AcLac}] & = & v_1 - v_2 & \frac{\mathrm{d}}{\mathrm{d}t}[\mathrm{DHIV}] & = & v_2 - v_3 \\ \frac{\mathrm{d}}{\mathrm{d}t}[\mathrm{KIV}] & = & v_3 - v_4 - v_5 - v_7 & \frac{\mathrm{d}}{\mathrm{d}t}[\mathrm{IPM}] & = & v_7 - v_8 \\ \frac{\mathrm{d}}{\mathrm{d}t}[\mathrm{Val}] & = & v_4 + v_5 - v_6 & \frac{\mathrm{d}}{\mathrm{d}t}[\mathrm{KIC}] & = & v_8 - v_9 \\ \frac{\mathrm{d}}{\mathrm{d}t}[\mathrm{Leu}] & = & v_9 - v_{10} & & & \end{array}$$

1) Reversible Convenience Kinetics (CK MM, rev.): The reactions R_3, R_6 and R_{10} follow a two molecular Michaelis-Menten reaction mechanism. Eq. (1) gives the general equation for bimolecular enzyme reactions of S and E forming product P and the catalyzer E inhibited by I as a special case of the generalized mass-action kinetics.

$$v_j = \frac{\frac{v_+^{\max}}{K_{\mathrm{S}}^{\mathrm{M}}}[\mathrm{S}] - \frac{v_-^{\max}}{K_{\mathrm{P}}^{\mathrm{M}}}[\mathrm{P}]}{1 + \frac{[\mathrm{I}]}{K^{\mathrm{I}a}} + \left(\frac{[\mathrm{S}]}{K_{\mathrm{S}}^{\mathrm{M}}} + \frac{[\mathrm{P}]}{K_{\mathrm{P}}^{\mathrm{M}}}\right)\left(1 + \frac{[\mathrm{I}]}{K^{\mathrm{I}b}}\right)} \qquad (1)$$

In case of R_3 there might be a reverse reaction. Both of the remaining reactions are assumed to be irreversible as they describe the transport of Val and Leu out of the cell. To avoid numerical problems, the inhibition constants in Michaelis-Menten kinetics were transformed into their reciprocals $K^{\mathrm{I}a|b\prime} = \frac{1}{K^{\mathrm{I}a|b}}$. This modification allows us to model any kind of inhibition [6] as limits of Eq. (1):

- competetive (for $0 < K^{\mathrm{I}a} < \infty$, $K^{\mathrm{I}b} \to \infty$)
- noncompetetive (for $0 < K^{\mathrm{I}a} = K^{\mathrm{I}b} < \infty$) and
- uncompetetive (for $K^{\mathrm{I}a} \to \infty$, $0 < K^{\mathrm{I}b} < \infty$).

For instance, by setting $K^{\mathrm{I}a|b\prime} = 0$ we obtain the same effect as if $K^{\mathrm{I}a|b} \to \infty$. For sake of simplicity we omit the prime symbol in the following equations.

$$v_3 = \frac{\frac{v^{\max}_{+1}}{K^{\mathrm{M}}_{[\mathrm{DHIV}]}}[\mathrm{DHIV}] - \frac{v^{\max}_{-1}}{K^{\mathrm{M}}_{[\mathrm{KIV}]}}[\mathrm{KIV}]}{1+K^{\mathrm{I}a}_1[\mathrm{Val}]+\left(\frac{[\mathrm{DHIV}]}{K^{\mathrm{M}}_{[\mathrm{DHIV}]}}+\frac{[\mathrm{KIV}]}{K^{\mathrm{M}}_{[\mathrm{KIV}]}}\right)\left(1+K^{\mathrm{I}b}_1[\mathrm{Val}]\right)} \quad (2)$$

$$v_6 = \frac{v^{\max}_{+2}[\mathrm{Val}]}{K^{\mathrm{M}}_{[\mathrm{Val}]}+[\mathrm{Val}]+\left(K^{\mathrm{M}}_{[\mathrm{Val}]}K^{\mathrm{I}a}_2+K^{\mathrm{I}b}_2[\mathrm{Val}]\right)[\mathrm{Leu}]} \quad (3)$$

$$v_{10} = \frac{v^{\max}_{+3}[\mathrm{Leu}]}{K^{\mathrm{M}}_{[\mathrm{Leu}]}+[\mathrm{Leu}]+\left(K^{\mathrm{M}}_{[\mathrm{Leu}]}K^{\mathrm{I}a}_3+K^{\mathrm{I}b}_3[\mathrm{Leu}]\right)[\mathrm{Val}]} \quad (4)$$

Recently, the convenience kinetics was suggested by Liebermeister *et al.* [7], designed to be a standard formalism for any enzyme reaction where the exact mechanism is unknown or as an approximation of the real kinetics. The formalism was derived from the trimolecular Michaelis-Menten reaction with a random order mechanism. It was shown that this formalism is able to describe any reaction mechanism in a reasonable way [7]. The general equation of the convenience kinetics for reaction j reads

$$v_j = \frac{k^{\mathrm{cat}}_{+j}\prod_i\left(\frac{S_i}{K^{\mathrm{M}}_{ji}}\right)^{n^-_{ij}} - k^{\mathrm{cat}}_{-l}\prod_i\left(\frac{S_i}{K^{\mathrm{M}}_{ji}}\right)^{n^+_{ij}}}{\prod_i\sum_{m=0}^{n^-_{ij}}\left(\frac{S_i}{K^{\mathrm{M}}_{ji}}\right)^m + \prod_i\sum_{m=0}^{n^+_{ij}}\left(\frac{S_i}{K^{\mathrm{M}}_{ji}}\right)^m - 1} \cdot[\mathrm{E}_j]\prod_m h_{\mathrm{A}}(S_m, K^{\mathrm{A}}_{jm})^{w^+_{jm}} h_{\mathrm{I}}(S_m, K^{\mathrm{I}}_{jm})^{w^-_{jm}} \quad (5)$$

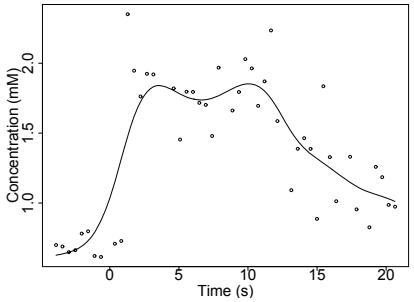
with h_{A} and h_{I} being functions for activation or inhibition, respectively, the turnover rates $k^{\mathrm{cat}}_{\pm j}$ and the matrices $\mathbf{W}^{\pm}$ containing positive entries for the connectivity of the metabolites as well as K^{M}_{ji} being a constant analogous to the Michaelis-Menten constant K^{M} [7]. Function h_{A} can be modeled in two alternative ways [7], for inhibition, which plays an important role in the system under study,

$$h_{\mathrm{I}}(S_i, K^{\mathrm{I}}) = \frac{K^{\mathrm{I}}}{K^{\mathrm{I}}+S_i} = \frac{1}{1+\frac{S_i}{K^{\mathrm{I}}}} = \frac{1}{1+K^{\mathrm{I}\prime}S_i} \quad (6)$$

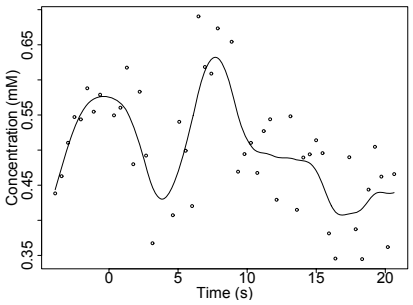
was suggested and herein applied. Eq. (5) is also a special case of the generalized mass-action kinetics. The product $[\mathrm{E}_j]k^{\mathrm{cat}}_{\pm j}$ was lumped into one parameter $k^{\mathrm{cat}\prime}_{\pm j}$ for all j assuming all enzyme concentrations to remain constant.

Applying Eq. (5) to reaction system R_1 through R_{10} yields the Eq. system (7) – (13). The three reactions that follow the traditional monomolecular Michaelis-Menten mechanism are modeled using Eq. (1). The reactions R_6 and R_{10} are considered irreversible as described before. The whole system contains 59 parameters. The stochiometric matrix has full column rank. Hence, the parameters $k^{\mathrm{cat}}_{\pm j}$ can be estimated directly without violating thermodynamic constraints.

2) Irreversible Convenience Kinetics: By setting all product concentrations apart from R_2 and R_9 to zero, we obtained an irreversible version of this model containing 41 parameters.



(a) Pyruvate (Pyr)



(b) NAD^+

Fig. 2. Representing external metabolites using approximation splines
Six of the 13 measured metabolites were considered external and approximated with splines which are not shown for Alanine (Ala), α-ketoglutarate (αKG), glutamate (Glut) and $NADP^+$.

3) Representing external Metabolites with Splines: As suggested by Magnus *et al.* [4], metabolites whose concentrations cannot be explained in terms of the model are considered external, i. e., they are an input to the model but involved in several other reactions outside this system (Fig. 1) and approximated using cubic splines smoothing the measurments. To weight all measurements equally, all ω_i were set to 1. Due to the different ranges of the concentrations of the six metabolites it is not possible to find one appropriate degree of smoothness λ that leads to equally smooth curves. Hence, we transformed all concentrations into the range $[0, 1]$, set $\lambda = 1$, computed the spline coefficients and retransformed them back into the original range (Fig. 2).

C. Fitness Function and Search Space Restrictions

Due to the differences in the concentrations of certain metabolites the Euclidian distance between the model values and the measurements is not applicable: Metabolites in higher concentration would dominate the fitness over lower concentrated ones. Minimizing the relative squared error (RSE, cf. Eq. 14) overcomes these limitations. The first sum

$$v_1 = \frac{\frac{k^{\text{cat}}_{+1}\cdot[\text{AHAS}]\cdot K^{\text{I}}_1}{\left(K^{\text{M}}_{[\text{Pyr}]1}\right)^2}[\text{Pyr}]^2 - \frac{k^{\text{cat}}_{-1}\cdot[\text{AHAS}]\cdot K^{\text{I}}_1}{K^{\text{M}}_{[\text{AcLac}]1}}[\text{AcLac}]}{\left(1+\frac{[\text{Pyr}]}{K^{\text{M}}_{[\text{Pyr}]1}}+\left(\frac{[\text{Pyr}]}{K^{\text{M}}_{[\text{Pyr}]1}}\right)^2+\frac{[\text{AcLac}]}{K^{\text{M}}_{[\text{AcLac}]1}}\right)\left(K^{\text{I}}_1+[\text{Val}]\right)} \tag{7}$$

$$v_2 = \frac{\frac{k^{\text{cat}}_{+2}\cdot[\text{AHAIR}]\cdot K^{\text{I}}_2}{K^{\text{M}}_{[\text{AcLac}]2}\cdot K^{\text{M}}_{[\text{NADPH}_2]1}}[\text{AcLac}][\text{NADPH}_2] - \frac{k^{\text{cat}}_{-2}\cdot[\text{AHAIR}]\cdot K^{\text{I}}_2}{K^{\text{M}}_{[\text{DHIV}]1}\cdot K^{\text{M}}_{[\text{NADP}^+]1}}[\text{DHIV}][\text{NADP}^+]}{\left(\left(1+\frac{[\text{AcLac}]}{K^{\text{M}}_{[\text{AcLac}]2}}\right)\left(1+\frac{[\text{NADPH}_2]}{K^{\text{M}}_{[\text{NADPH}_2]1}}\right)+\left(1+\frac{[\text{DHIV}]}{K^{\text{M}}_{[\text{DHIV}]1}}\right)\left(1+\frac{[\text{NADP}^+]}{K^{\text{M}}_{[\text{NADP}^+]1}}\right)-1\right)\left(K^{\text{I}}_2+[\text{Val}]\right)} \tag{8}$$

$$v_4 = \frac{\frac{k^{\text{cat}}_{+3}\cdot[\text{BCAAT}_{\text{ValB}}]}{K^{\text{M}}_{[\text{KIV}]1}\cdot K^{\text{M}}_{[\text{Glut}]1}}[\text{KIV}][\text{Glut}] - \frac{k^{\text{cat}}_{-3}\cdot[\text{BCAAT}_{\text{ValB}}]}{K^{\text{M}}_{[\text{Val}]1}\cdot K^{\text{M}}_{[\alpha\text{KG}]1}}[\text{Val}][\alpha\text{KG}]}{1+\frac{[\text{KIV}]}{K^{\text{M}}_{[\text{KIV}]1}}+\frac{[\text{Glut}]}{K^{\text{M}}_{[\text{Glut}]1}}+\frac{[\text{KIV}][\text{Glut}]}{K^{\text{M}}_{[\text{KIV}]1}\cdot K^{\text{M}}_{[\text{Glut}]1}}+\frac{[\text{Val}]}{K^{\text{M}}_{[\text{Val}]1}}+\frac{[\alpha\text{KG}]}{K^{\text{M}}_{[\alpha\text{KG}]1}}+\frac{[\text{Val}][\alpha\text{KG}]}{K^{\text{M}}_{[\text{Val}]1}\cdot K^{\text{M}}_{[\alpha\text{KG}]1}}} \tag{9}$$

$$v_5 = \frac{\frac{k^{\text{cat}}_{+4}\cdot[\text{BCAAT}_{\text{ValC}}]}{K^{\text{M}}_{[\text{KIV}]2}\cdot K^{\text{M}}_{[\text{Ala}]1}}[\text{KIV}][\text{Ala}] - \frac{k^{\text{cat}}_{-4}\cdot[\text{BCAAT}_{\text{ValC}}]}{K^{\text{M}}_{[\text{Val}]2}\cdot K^{\text{M}}_{[\text{Pyr}]2}}[\text{Val}][\text{Pyr}]}{1+\frac{[\text{KIV}]}{K^{\text{M}}_{[\text{KIV}]2}}+\frac{[\text{Ala}]}{K^{\text{M}}_{[\text{Ala}]1}}+\frac{[\text{KIV}][\text{Ala}]}{K^{\text{M}}_{[\text{KIV}]2}\cdot K^{\text{M}}_{[\text{Ala}]1}}+\frac{[\text{Val}]}{K^{\text{M}}_{[\text{Val}]2}}+\frac{[\text{Pyr}]}{K^{\text{M}}_{[\text{Pyr}]2}}+\frac{[\text{Val}][\text{Pyr}]}{K^{\text{M}}_{[\text{Val}]2}\cdot K^{\text{M}}_{[\text{Pyr}]2}}} \tag{10}$$

$$v_7 = \frac{\frac{k^{\text{cat}}_{+5}\cdot[\text{IPMS}]\cdot K^{\text{I}}_3}{K^{\text{M}}_{[\text{KIV}]3}\cdot K^{\text{M}}_{[\text{AcCoA}]1}}[\text{KIV}][\text{AcCoA}] - \frac{k^{\text{cat}}_{-5}\cdot[\text{IPMS}]\cdot K^{\text{I}}_3}{K^{\text{M}}_{[\text{IPM}]1}\cdot K^{\text{M}}_{[\text{CoA}]1}}[\text{IPM}][\text{CoA}]}{\left(1+\frac{[\text{KIV}]}{K^{\text{M}}_{[\text{KIV}]3}}+\frac{[\text{AcCoA}]}{K^{\text{M}}_{[\text{AcCoA}]1}}+\frac{[\text{KIV}][\text{AcCoA}]}{K^{\text{M}}_{[\text{KIV}]3}\cdot K^{\text{M}}_{[\text{AcCoA}]1}}+\frac{[\text{IPM}]}{K^{\text{M}}_{[\text{IPM}]1}}+\frac{[\text{CoA}]}{K^{\text{M}}_{[\text{CoA}]1}}+\frac{[\text{IPM}][\text{CoA}]}{K^{\text{M}}_{[\text{IPM}]1}K^{\text{M}}_{[\text{CoA}]1}}\right)\left(K^{\text{I}}_3+[\text{Leu}]\right)} \tag{11}$$

$$v_8 = \frac{\frac{k^{\text{cat}}_{+6}\cdot[\text{IPMDH}]}{K^{\text{M}}_{[\text{IPM}]2}\cdot K^{\text{M}}_{[\text{NAD}^+]1}}[\text{IPM}][\text{NAD}^+] - \frac{k^{\text{cat}}_{-6}\cdot[\text{IPMDH}]}{K^{\text{M}}_{[\text{KIC}]1}\cdot K^{\text{M}}_{[\text{NADH}_2]1}}[\text{KIC}][\text{NADH}_2]}{1+\frac{[\text{IPM}]}{K^{\text{M}}_{[\text{IPM}]2}}+\frac{[\text{NAD}^+]}{K^{\text{M}}_{[\text{NAD}^+]1}}+\frac{[\text{IPM}][\text{NAD}^+]}{K^{\text{M}}_{[\text{IPM}]2}\cdot K^{\text{M}}_{[\text{NAD}^+]1}}+\frac{[\text{KIC}]}{K^{\text{M}}_{[\text{KIC}]1}}+\frac{[\text{NADH}_2]}{K^{\text{M}}_{[\text{NADH}_2]1}}+\frac{[\text{KIC}][\text{NADH}_2]}{K^{\text{M}}_{[\text{NADH}_2]1}K^{\text{M}}_{[\text{KIC}]1}}} \tag{12}$$

$$v_9 = \frac{\frac{k^{\text{cat}}_{+7}\cdot[\text{BCAAT}_{\text{LeuB}}]}{K^{\text{M}}_{[\text{KIC}]2}\cdot K^{\text{M}}_{[\text{Glut}]2}}[\text{KIC}][\text{Glut}] - \frac{k^{\text{cat}}_{-7}\cdot[\text{BCAAT}_{\text{LeuB}}]}{K^{\text{M}}_{[\text{Leu}]1}\cdot K^{\text{M}}_{[\alpha\text{KG}]2}}[\text{Leu}][\alpha\text{KG}]}{1+\frac{[\text{KIC}]}{K^{\text{M}}_{[\text{KIC}]2}}+\frac{[\text{Glut}]}{K^{\text{M}}_{[\text{Glut}]2}}+\frac{[\text{KIC}][\text{Glut}]}{K^{\text{M}}_{[\text{KIC}]2}\cdot K^{\text{M}}_{[\text{Glut}]2}}+\frac{[\text{Leu}]}{K^{\text{M}}_{[\text{Leu}]1}}+\frac{[\alpha\text{KG}]}{K^{\text{M}}_{[\alpha\text{KG}]2}}+\frac{[\text{Leu}][\alpha\text{KG}]}{K^{\text{M}}_{[\text{Leu}]1}\cdot K^{\text{M}}_{[\alpha\text{KG}]2}}} \tag{13}$$

runs over all dimensions of $\hat{\mathbf{x}}$ describing the model output at each sample time τ_t. T is the number of measurements taken and $\mathbf{X} = (x_{ti})$ the given data matrix.

$$f_{\text{RSE}}(\hat{\mathbf{x}}, \mathbf{X}) = \sum_{i=1}^{\dim(\hat{\mathbf{x}})}\sum_{t=1}^{T}\left(\frac{\hat{x}_i(\tau_t) - x_{ti}}{x_{ti}}\right)^2 \tag{14}$$

The fitness (Eq. 14) was used in several publications for similar problems [12]. An implementation of the fourth order Runge-Kutta method solved the ordinary differential equation systems to obtain $\hat{\mathbf{x}}(\tau_t)$ for every t.

In biology the parameter space is limited to values greater than or equal to zero and cannot exceed the diffusion rate. So to restrict the search space for the optimizers, we limited parameter values to the range $[0, 2000]$, still covering 98.748 % of all known kinetic parameters in the BRENDA database [13]. In more detail, 99.958 % of all K^{I} values, 99.957 % of all K^{M} values and 96.328 % of all k^{cat} values are lower than 2000. All known parameters in BRENDA are greater than or equal to zero. To avoid division by zero in some parameters, the range was set to $[\varepsilon, 2000]$ with $\varepsilon = 10^{-8}$ there. For transformed parameters in Michaelis-Menten equations (Sec. II-B.1) the range was limited to $[0, 10^8]$. Only 0.962 % of all K^{I} and 0.004 % of all K^{M} values in BRENDA are reported to be lower than 10^{-8}.

Previous Monte Carlo searches showed that initialization plays an important role due to the high nonlinearity of both considered models. Parameter values chosen completely by chance often lead to instable systems. Hence, all parameters were initialized with low values, assuming that large parameter values are rather infrequent in nature. This assumtion is also supported by the entries of the BRENDA database [14], showing that 64.807 % of the known parameters are $\leqslant 2$. A Gaussian distribution with $\mu = \sigma = 1$ guarantees low initial values and ensures stable initial populations. Each parameter was set to the boundary values if it broke any of the search space restrictions.

D. Standard Settings for the Optimization Algorithms

Using the JAVAEVA framework[1] [15] we tested the following Evolutionary Algorithms on the inference problem:

- (Multi-start) Hill Climber (HC), the number of starts varying from 1, 10, 25, 50, 100 to 250. All used Gaussian mutation with a fixed standard deviation of $\sigma = 0.2$ and a mutation probability $p_m = 1.0$.
- Simulated Annealing (SA) with $\alpha = 0.1$ and an initial temperature of $T = 5$ using a linear annealing schedule and a population size of 250.
- Binary Genetic Algorithm (binGA) with one-point mutation, $p_m = 0.1$, and one-point crossover, $p_c = 0.7$.

[1] http://www-ra.informatik.uni-tuebingen.de/software/JavaEvA

- Real valued Genetic Algorithm (realGA) with global mutation, $p_m = 0.1$ and UNDX crossover, $p_c = 0.8$. Both GAs used tournament selection with a group size of 8 in a population of 250 individuals.
- Standard Evolution Strategy (stdES) as (5,25)-ES with global mutation, $p_m = 0.8$ and discrete one-point crossover, $p_c = 0.2$.
- Evolution Strategy with covariance matrix adaption (cmaES) as (5,25)-ES, $p_m = 1.0$, no crossover. Both ESs used deterministic best-first selection to choose the next generation.
- Constricted Particle Swarm Optimization (PSO) setting $\phi_1 = 2.05, \phi_2 = 2.05, \chi = 0.73$ and using star topology and a population size of 100.
- Differential Evolution (DE) with the scheme DE/current-to-best/1 [16] setting $\lambda = F = 0.8$, $CR = 0.5$ and a population size of 100.

For all algorithms with population sizes lower than 250 individuals, a pre-population with 250 parameter vectors was generated and the best were selected to generate the initial population. This step is crucial to obtain comparable results for algorithms with different population sizes [12]. Every setting was repeated 20 times with 100,000 fitness evaluations per run.

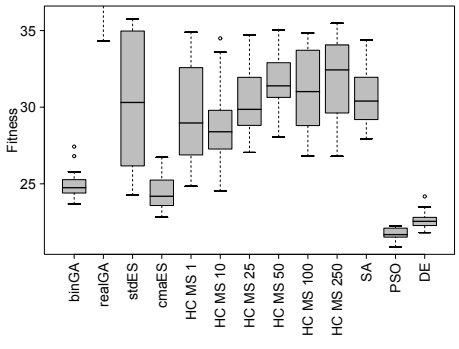
E. Preliminary Comparison

Figures 3(a) and 3(b) show the results of the preliminary comparison for the rev. and irrev. CK MM, respectively. Even though the different Hill Climbers produce lower fitness values in the irreversible case, the global evolutionary optimizers reach notably better fits using the reversible model. The realGA fails badly on both models, whereas four algorithms show to be interesting for further analysis: binGA, PSO and DE produce similarly good results on both models. The cmaES is not competitive in the irreversible case, however it is among the most successful in the reversible case (Tab. III). We therefore concentrate on binGA, cmaES, PSO as well as DE and take a closer look on parameter settings for the reversible model.

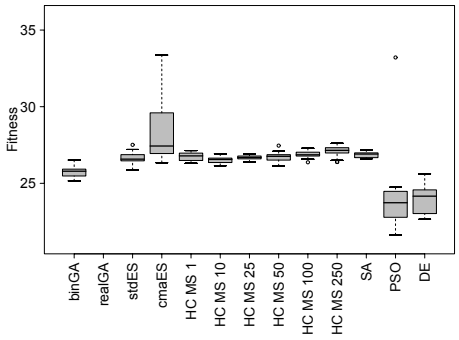
F. Parameter Analysis for the Evolutionary Algorithms

We first studied the influence of different mutation and crossover operators on the binGA. A grid search over the operators provided by the EA framework was performed with 20 runs per setting at 100,000 fitness evaluations each and population size 100. Next to standard one-point, n-point (setting $n = 3$) and uniform crossover the bit-simulated crossover scheme was tested as well. We tried no mutation, one-point and adaptive mutation, an operator which modifies individual mutation probabilities similar to ES step-size adaptation. For the most successful operator combination, different population sizes have then been tested.

For the cmaES we employed a grid search over the population parameters $\mu \in \{1, 5, 10, 25, 50, 125\}$ and $\lambda \in \{25, 50, 125, 250, 500\}$ rejecting combinations where $\mu \geqslant \lambda$ and keeping p_m at 1 and p_c at 0.



(a) reversible



(b) irreversible

Fig. 3. Comparison of the different Evolutionary Algorithms applied to both models

In the preliminary test, the PSO with standard parameters performed best on average, still we tested the alternative settings $\phi_1 = 2.8$ and $\phi_2 = 1.3$ as suggested in [17] as well as additional population sizes, namely $\{25, 50, 250, 500\}$.

For the DE approach another grid search was performed, alltogether testing values for $f, \lambda \in \{0.5, 0.8\}$ and $CR \in \{0.3, 0.5, 0.9\}$. For the most promising parameter set, the population size was varied additionally in $\{50, 250, 500, 1000\}$.

G. Hardware Configuration

All experiments were run on a cluster with 16 AMD dual Opteron CPUs with 2.4 GHz, 1 MB level 2 cache and 2 GB RAM per node under the Sun Grid Engine and JVM 1.5.0 with Scientific Linux 4 as operating system. An experiment with 20 runs took a computation time of approximately 1.5 h.

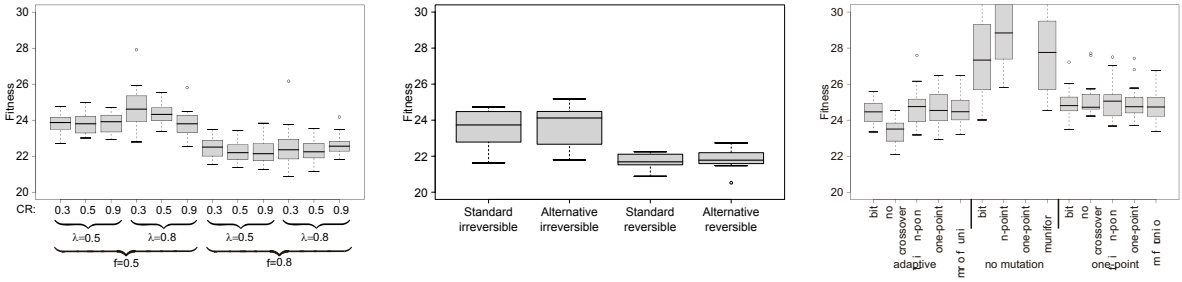
III. Results

The grid search over the GA operators revealed that adaptive mutation without crossover performs slightly better than the other combinations (Fig. 4). When testing this

TABLE III
PRELIMINARY TEST RESULTS

Listed are four single best and average best algorithms, respectively.

CK MM, reversible					CK MM, irrreversible				
Min.	Algorithm	Average	Std. Dev.	Algorithm	Min.	Algorithm	Average	Std. Dev.	Algorithm
20.882	PSO	21.773	0.352	PSO	21.632	PSO	23.968	0.931	DE
21.821	DE	22.633	0.562	DE	22.651	DE	23.991	2.415	PSO
22.829	cmaES	24.341	1.026	cmaES	25.152	binGA	25.761	0.331	binGA
23.687	binGA	24.960	0.910	binGA	25.888	stdES	26.539	0.210	HC MS 10



(a) DE (b) PSO (c) binGA

Fig. 4. Benchmark of different parameters for DE and PSO, and operators of binGA

configuration for different population sizes, a value of 500 suggests a minor advantage, cf. Fig. 6(c).

The cmaES performs best in the (25,50)-configuration (Fig. 5) but does not achieve the excellence of the other optimizers. Without further investigation we attribute this to the highly deceptive target function with the typically high ES selection pressure making it hard to escape local minima.

The alternative parameter setting for PSO did not show advantageous (Fig. 4(b)). The standard population size of 25 for PSO was best on average (Fig. 6(b)), confirming the ability of PSO to work fine with small swarms, compared to other EAs, also due to the fact that PSO does not apply classical selection.

DE shows best average results for $f = 0.8$, $\lambda = 0.5$ and $CR = 0.9$, squeezing the average performance close to a fitness of 22 (Fig. 4(a)). The population size of 100 turns out to be most effective for this setting (Fig. 6(a)).

The accordingly tuned settings were also applied to the irreversible model (Tab. IV) and, besides PSO, showed improvements. Still the overall best fitness (19.974) was found for the reversible model using PSO with $\phi_1 = \phi_2 = 2.05$ and a population size of 50 (Tab. IV).

The resulting model systems for the parameter values yielding the lowest fitness values are plotted in Fig. 7. Splines were added for a better visualization to indicate a plausible

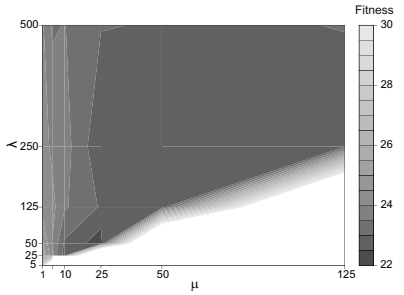


Fig. 5. Impact of μ and λ on the performance of the cmaES

fit (RSE of the splines: 19.670). These were constructed using the settings described in Sec. II-B.3. The irreversible system is often unable to follow the dynamic behavior of the measurements and results in straight lines, which are locally optimal but have low biological significance.

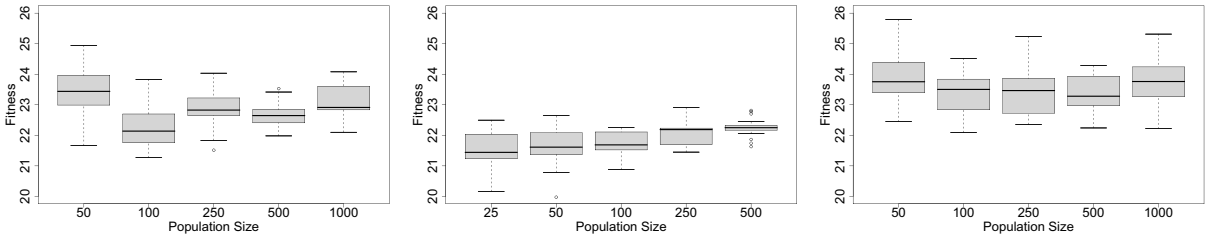
IV. CONCLUSIONS

In this study we conducted a systematic benchmark on the problem of parameter estimation for two convenience kinetics models on the Val/Leu biosynthesis in *C. glutamicum*. Out of the eight different optimization procedures tested, four showed outstanding performance approximating the *in vivo*

TABLE IV

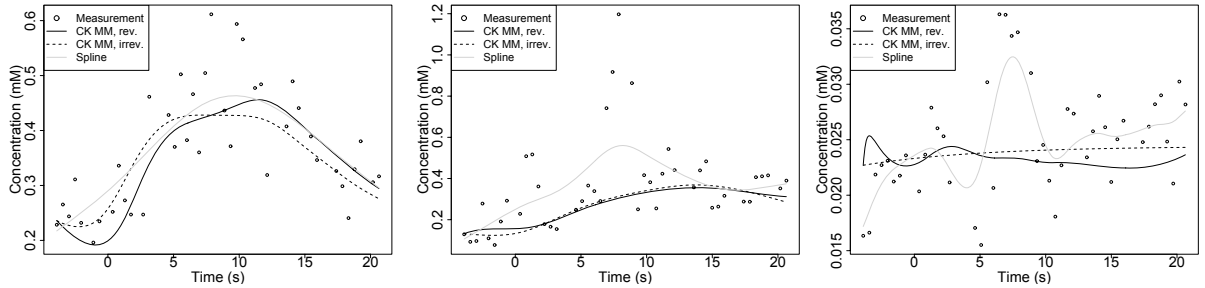
STATISTICS ON THE MOST SUCCESSFUL RUNS OF EACH MAIN OPTIMIZER

CK MM, rev.			CK MM, irrev.			Algorithm	Population Size
Minimum	Average	Std. dev.	Minimum	Average	Std. dev.		
22.092	23.353	0.666	25.040	25.346	0.176	binGA, adaptive MUT, no CO	100
21.431	23.222	1.066	25.632	28.055	2.697	cmaES, $p_c = 0, p_m = 1$	$(10, 50)$
20.862	22.499	1.119	21.763	23.603	1.268	DE, $f = \lambda = 0.8$, $CR = 0.3$	100
19.974	21.620	0.638	21.747	24.468	3.082	PSO, $\phi_1 = \phi_2 = 2.05$	50

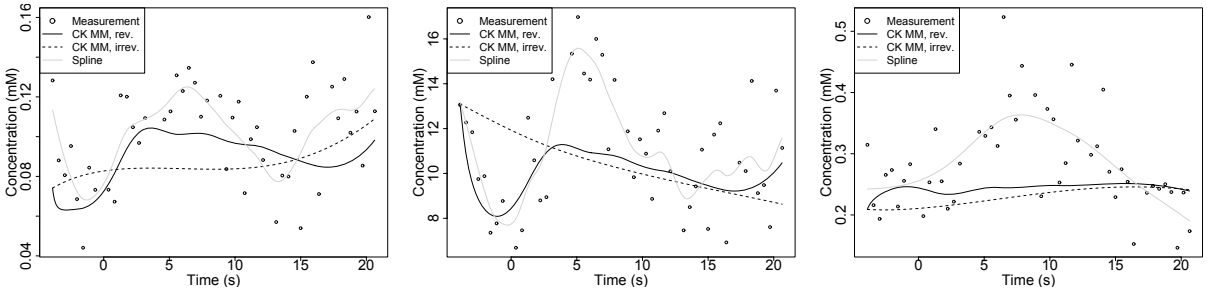


(a) DE (b) PSO (c) binary GA

Fig. 6. Results of different population sizes for binGA, PSO and DE, 20 runs per experiment.



(a) Acetolactate (AcLac) (b) Dihydroxyisovalerate (DHIV) (c) Isopropylmalate (IPM)



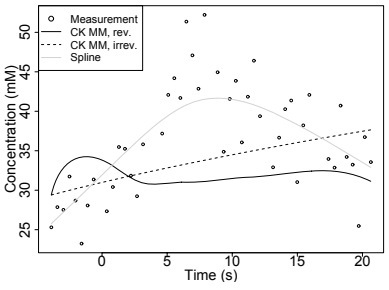
(d) 2-Ketoisocaproate (KIC) (e) 2-Ketoisovalerate (KIV) (f) L-Leucine (Leu)

Fig. 7. The best fit of the different models

data. Although the concentration trace of some metabolites could not be simulated with a high degree of exactness, other traces approached the quality of independent spline approximations.

The splines, however, are uncoupled and do not underly any biological restraints, thus having no explanatory power. With regard to more complex metabolic systems, we showed that the convenience kinetics is an appropriate standard formalism when exact knowledge of the underlying mechanism is not available. This is often the case in large reaction databases like KEGG or MetaCyc. Especially the reversible variant allowed good reconstruction of the experimental data while the irreversible alternative often produced implausible straight lines. We conclude that secondary reactions not incorporated in our model system interfere with this pathway. In our case, this can only be represented by the assumption of reversibility. Thereby the model is granted the necessary degree of freedom to come close to the observed data.

Among the Evolutionary Algorithms tested, four reached notable results on the parameter inference problem. The binGA and cmaES can however not fully compete with tuned PSO and DE approaches. The best parameter set found by PSO on the reversible model had an RSE of 19.974, close to the RSE of 19.670 of the spline approximation. In the related study of Spieth *et al.* [11] on *in silico* data mainly from gene regulatory network models, ES were found to be most successful for parameter estimation followed by DE, while binGA and PSO performed poorly. However, for parameter estimation in metabolic models based on *in vivo* data, the ES was not adequate but outperformed by DE and PSO. DE seems to be a reasonable choice for both kinds of biological models. For future extensions of this model system towards more complex metabolic networks and their parameter estimation, the results of this study will be a valuable basis.



(g) L-Valine (Val)

Fig. 7. The best fit of the different models

ACKNOWLEGEMENTS

We are grateful to Ralf Takors, Klaus Beyreuther and Oliver Kohlbacher. A. D. and J. S. are funded by the NGFN-II EP project no. 0313323.

REFERENCES

[1] R. Guthke, W. Schmidt-Heck, G. Pless, R. Gebhardt, M. Pfaff, J. C. Gerlach, and K. Zeilinger, "Dynamic Model of Amino Acid and Carbohydrate Metabolism in Primary Human Liver Cells," in *VII International Symposium on Biological and Medical Data Analysis*, 2006.

[2] E. Klipp, B. Nordlander, R. Kruger, P. Gennemark, and S. Hohmann, "Integrative model of the response of yeast to osmotic shock," *Nature Biotechnology*, vol. 23, no. 8, pp. 975–982, August 2005. [Online]. Available: http://dx.doi.org/10.1038/nbt1114

[3] C. Chassagnole, N. Noisommit-Rizzi, J. W. Schmid, K. Mauch, and M. Reuss, "Dynamic modeling of the central carbon metabolism of *Escherichia coli*," *Wiley Periodicals, Inc.*, pp. 54–73, January 2002 2002.

[4] J. B. Magnus, D. Hollwedel, M. Oldiges, and R. Takors, "Monitoring and Modeling of the Reaction Dynamics in the Valine/Leucine Synthesis Pathway in *Corynebacterium glutamicum*," *Biotechnology Progress*, vol. 22, no. 4, pp. 1071–1083, 2006. [Online]. Available: http://dx.doi.org/10.1021/bp060072f

[5] R. Guthke, K. Zeilinger, S. Sickinger, W. Schmidt-Heck, H. Buentemeyer, K. Iding, J. Lehmann, M. Pfaff, G. Pless, and J. C. Gerlach, "Dynamics of amino acid metabolism of primary human liver cells in 3D bioreactors," *Bioprocess Biosystem Engineering*, vol. 28, no. 5, pp. 331–340, April 2006.

[6] R. Heinrich and S. Schuster, *The Regulation of Cellular Systems*. 115 Fifth Avenue New York, NY 10003: Chapman and Hall, 1996.

[7] W. Liebermeister and E. Klipp, "Bringing metabolic networks to life: convenience rate law and thermodynamic constraints." *Theor. Biol. Med. Model.*, vol. 3, p. 41, December 2006. [Online]. Available: http://dx.doi.org/10.1186/1742-4682-3-41

[8] M. Kanehisa, S. Goto, M. Hattori, K. F. Aoki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, and M. Hirakawa, "From genomics to chemical genomics: new developments in KEGG," *Nucl. Acids Res.*, vol. 34, no. suppl_1, pp. D354–357, 2006. [Online]. Available: http://nar.oxfordjournals.org/cgi/content/abstract/34/suppl_1/D354

[9] R. Caspi, H. Foerster, C. A. Fulcher, R. Hopkinson, J. Ingraham, P. Kaipa, M. Krummenacker, S. Paley, J. Pick, S. Y. Rhee, C. Tissier, P. Zhang, and P. D. Karp, "MetaCyc: a multiorganism database of metabolic pathways and enzymes." *Nucl. Acids Res.*, vol. 34, no. Database issue, pp. D511–D516, January 2006. [Online]. Available: http://dx.doi.org/10.1093/nar/gkj128

[10] D. Visser and J. Heijnen, "The mathematics of metabolic control analysis revisited," *Metabolic Engineering*, vol. 4, pp. 114–123, 05 2002. [Online]. Available: http://www.idealibrary.com

[11] C. Spieth, N. Hassis, F. Streichert, J. Supper, K. Beyreuther, and A. Zell, "Comparing Mathematical Models on the Problem of Network Inference," in *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2006)*, 2006.

[12] C. Spieth, R. Worzischek, F. Streichert, J. Supper, N. Speer, and A. Zell, "Comparing Evolutionary Algorithms on the Problem of Network Inference," in *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2006)*, 2006.

[13] J. Barthelmes, C. Ebeling, A. Chang, I. Schomburg, and D. Schomburg, "BRENDA, AMENDA and FRENDA: the enzyme information system in 2007," *Nucl. Acids Res.*, vol. 35, no. suppl_1, pp. D511–514, 2007. [Online]. Available: http://nar.oxfordjournals.org/cgi/content/abstract/35/suppl_1/D511

[14] I. Schomburg, A. Chang, and D. Schomburg, "BRENDA, enzyme data and metabolic information," *Nucl. Acids Res.*, vol. 30, no. 1, pp. 47–49, 2002. [Online]. Available: http://nar.oxfordjournals.org/cgi/content/abstract/30/1/47

[15] F. Streichert and H. Ulmer, "JavaEvA - A Java Framework for Evolutionary Algorithms," Center for Bioinformatics Tübingen, University of Tübingen, Technical Report WSI-2005-06, 2005. [Online]. Available: http://w210.ub.uni-tuebingen.de/dbt/volltexte/2005/1702/

[16] R. Storn, "On the Usage of Differential Evolution for Function Optimization," in *1996 Biennial Conference of the North American Fuzzy Information Processing Society*. Berkeley: IEEE, New York, USA, 1996, pp. 519–523.

[17] A. Carlisle and G. Dozier, "An off-the-shelf PSO," in *Proceedings of the Workshop on Particle Swarm Optimization*. Indianapolis: Purdue School of Engineering and Technology, Indianapolis, 2001.