Journal of Biotechnology 147 (2010) 88–101

Contents lists available at ScienceDirect

Journal of Biotechnology

journal homepage: www.elsevier.com/locate/jbiotec

CASOP: A Computational Approach for Strain Optimization aiming at high Productivity

Oliver Hädicke[a,b], Steffen Klamt[a,b,*]

[a] Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstrasse 1, D-39106 Magdeburg, Germany
[b] Magdeburg Centre for Systems Biology (MaCS), Sandtorstrasse 1, D-39106 Magdeburg, Germany

ARTICLE INFO

Article history:
Received 23 November 2009
Received in revised form 26 February 2010
Accepted 4 March 2010

Keywords:
Metabolic engineering
Elementary modes
Strain optimization
Succinate overproduction

ABSTRACT

The identification of suitable intervention strategies increasing the productivity of microorganisms is a central issue in metabolic engineering. Here, we introduce a computational framework for strain optimization based on reaction importance measures derived from weighted elementary modes. The objective is to shift the natural flux distribution to synthesis of the desired product with high production rates thereby retaining the ability of the host organism to produce biomass precursors. The stoichiometric approach allows consideration of regulatory/operational constraints and takes product yield and network capacity – the two major determinants of (specific) productivity – explicitly into account. The relative contribution of each reaction to yield and network capacity and thus productivity is estimated by analyzing the spectrum of available conversion routes (elementary modes). A result of our procedure is a reaction ranking suggesting knockout and overexpression candidates. Moreover, we show that the methodology allows for the evaluation of cofactor and co-metabolite requirements in conjunction with product synthesis. We illustrate the proposed method by studying the overproduction of succinate and lactate by *Escherichia coli*. The metabolic engineering strategies identified *in silico* resemble existing mutant strains designed for the synthesis of the respective products. Additionally, some non-intuitive intervention strategies are revealed.

1. Introduction

Metabolic engineering of microbial strains for optimizing the production of chemicals is a fundamental goal in biotechnology (Lee et al., 2008). By natural selection microorganisms are typically evolved for internal cellular objectives, one example is the maximization of biomass yield under certain conditions (Ibarra et al., 2002). Metabolic engineering seeks to prevail against this natural objective and to reshape the metabolism to economically produce the desired substance. Recent developments in molecular biotechnology facilitate the targeted engineering of microbial production systems. As the amount of annotated sequence information and reconstructed genome-scale metabolic networks explodes, there is a growing demand of modeling and computational tools to exploit this information for rational strain design. Several optimization-based methods have been developed in recent years (Kim et al., 2008). The Optknock framework (Burgard et al., 2003) and its extensions OptStrain (Pharkya et al., 2004) and OptReg (Pharkya and Maranas, 2006) rely on a bilevel linear optimization problem solved by mixed integer linear programming (MILP). Optknock identifies a given number of gene (reaction) knockouts that enforce maximal product synthesis if the organism follows its natural objective. For the case of biomass yield maximization as the cellular objective, this enables the use of adaptive evolution for strain design (Fong et al., 2005, 2006; Jantama et al., 2008). OptStrain incorporates non-native reactions into the stoichiometric model of the host organism. Subsequently, the Optknock algorithm is applied to further enhance the product yield. The OptReg procedure extends Optknock and allows up-regulation and down-regulation in addition to gene knockouts as possible genetic modifications. OptGene, another optimization approach (Patil et al., 2005), uses evolutionary programming for the identification of knockout targets after evaluating flux distributions by either flux balance analysis (FBA), minimization of metabolic adjustments (MOMA) or other objective functions.

Several frameworks for computer-aided metabolic engineering are based on elementary modes (EMs). EMs can be interpreted as minimal functional units (pathways) of metabolic networks that allow a balanced operation of the network at steady-state conditions (Schuster et al., 1999). Every possible stationary flux distribution of the network can be described as a weighted combination of EMs. Optimal flux distributions are thus a linear

* Corresponding author at: Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstrasse 1, D-39106 Magdeburg, Germany.
Tel.: +49 391 6110 480; fax: +49 391 6110 509.
E-mail address: klamt@mpi-magdeburg.mpg.de (S. Klamt).

0168-1656/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbiotec.2010.03.006

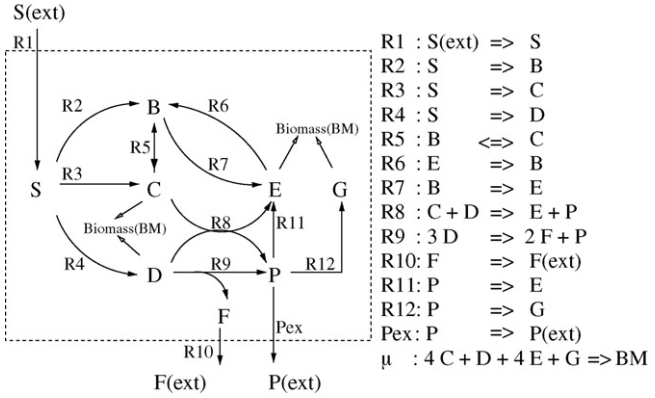

Fig. 1. Example network.

combination of optimal EMs. A theoretical framework for strain optimization relying on EMs was developed by Trinh et al. (2006, 2009). Their approach is targeted on designing organisms with minimal metabolic functionality (MMF) exhibiting highest product yield. For this purpose, reactions are iteratively removed from the network so that only a few optimal EMs with desirable properties remain in the network. Applicability of this approach was, e.g., demonstrated by the development of an ethanol overproducing *Escherichia coli* strain (Trinh et al., 2008).

Another concept for computing knockout strategies is minimal cut sets (MCSs) (Klamt and Gilles, 2004; Klamt, 2006). An MCS is defined as a minimal set of structural interventions repressing certain functionalities. The complete set of MCSs can be computed from the collection of (target) EMs exhibiting the functionality to be repressed.

The approaches mentioned above focus on optimal flux distributions with respect to yield. However, in industrial applications, the productivity of an engineered strain is often the actual performance parameter to be optimized. Productivity (here we will focus on *specific* productivity) is a combined measure of conversion capacity and yield. Increasing only product yield does therefore not necessarily result in sufficient production rates as can be illustrated by the example network in Fig. 1: assume the desired engineering goal is to produce *P* with high rate *Pex*. As one can easily see, the optimal yield is 1 mol *P* per mol substrate *S* and the pathway via R1, R4, and R9 is sub-optimal as it needs 3 mol of *S* to synthesize 1 mol *P*. However, concluding that deleting reaction R9 would contribute to a higher productivity w.r.t. *P* is, in general, not true. Using standard kinetic laws one can easily construct a kinetic model of the network in which the knockout of R9 may even reduce the output rate of *P* and can be one of the most unfavorable interventions (data not shown). Although the result is obviously sensitive to the particular parametrization, it demonstrates that removing pathways being non-optimal w.r.t. yield might be inappropriate for increasing productivity.

Here, we introduce CASOP (Computational Approach for Strain Optimization aiming at high Productivity), a new optimization framework based on EMs taking the relative contribution of a reaction to yield and flux capacity explicitly into account. Applicability of the proposed method will be demonstrated with case studies on overproduction of succinate and lactate in *E. coli*.

2. Materials and methods

2.1. Algorithm

Different criteria for network optimization with respect to product synthesis exist. Maximization of product yield is a common objective function used in computational optimization approaches. The optimization of biomass yield has been shown to be a natural objective function for some microorganisms, though it is not a generally valid principle in biology (Schuster et al., 2008). Another criterion for network evaluation is the specific productivity (specific production rate of the desired metabolite) which can be written as:

$$r_P = Y^{P/S} r_S. \tag{1}$$

(r_P: specific production rate [mol/(gDW h)]; $Y^{P/S}$: yield [mol product per mol substrate]; r_S: specific substrate uptake rate [mol/(gDW h)]). Since this study focuses on stoichiometric modeling approaches (being independent of kinetic information) the production rate r_P cannot directly be used in an objective function for optimization. Eq. (1) suggests that yield and thus the activity of yield-optimal pathways is important for achieving a high production rate. However, Eq. (1) also reveals that not only yield but also the specific substrate uptake rate r_S will determine the production rate. r_S is the amount of substrate that is converted in a given time period and can thus also be interpreted as a measure for network capacity (throughput). Again, without kinetic data we cannot quantify the capacity of a metabolic network under given environmental conditions. As heuristic for a stoichiometric approach we will estimate the relative contribution of a reaction to the network's capacity by its participation in the available conversion routes (elementary modes), hence, by considering its contribution to network flexibility. Note that yield $Y^{P/S}$ and capacity r_S cannot be optimized independently: increasing the yield is usually achieved by knockouts of undesired (low-yield) pathways which may decrease the capacity of the network. On the other hand, keeping a high flexibility in the system implies that sub-optimal pathways can be utilized leading to a lower yield. Thus, optimization of productivity is a trade-off between yield and capacity. For example, allowing utilization of sub-optimal routes whose yield is close to the optimal one might lead to a higher productivity compared to a strain where only optimal pathways are retained. Our method proposed herein considers both yield and capacity and their relative contribution to productivity can be adjusted by a free parameter.

In the following we introduce and illustrate our method by means of the example network in Fig. 1 which consists of 8 internal metabolites and 14 reactions including a biomass (*BM*) synthesis reaction μ. The stoichiometry of the latter is given by $4C + 1D + 4E + 1G \rightarrow 1\,\mathrm{g}\ BM$ (it follows 0.1 g/mol as molar mass for the metabolites). In addition to our desired engineering goal (high rate of *Pex*), we assume that the ability of the organism to grow, i.e. to produce the required biomass precursors *C*, *D*, *E* and *G*, has to be retained. To consider this constraint properly, we rewrite the stoichiometric network as follows: we temporarily consider the external metabolites *P(ext)* and *BM* as internal metabolites and assume that an artificial reaction R_V combines them to an artificial external metabolite *V*:

$$R_V : (1-\gamma)BM + \gamma\,\alpha_P\,P(ext) \rightarrow 1\,[gram]\ V, \quad \gamma \in [0,1]. \tag{2}$$

Parameter γ assigns the mass proportion of *P(ext)* at *V* reflecting in reality the relative production of *P(ext)* with respect to synthesis of biomass. The case of $\gamma = 0$ corresponds to exclusive biomass synthesis without product formation, whereas with $\gamma = 1$ the exclusive production of the chemical is attained. In order to compare flux distributions for different ratios of growth to product synthesis, γ can be varied between 0 and 1 (and each γ represents a "proportion scenario"). In order to maintain stoichiometric consistency, in Eq. (2), *P(ext)* has to be weighted by the reciprocal of its molar mass α_P [mol/g] and for the example network it follows that $\alpha_P = 10$ mol/g. From here we will simply write *P* when referring to *P(ext)*.

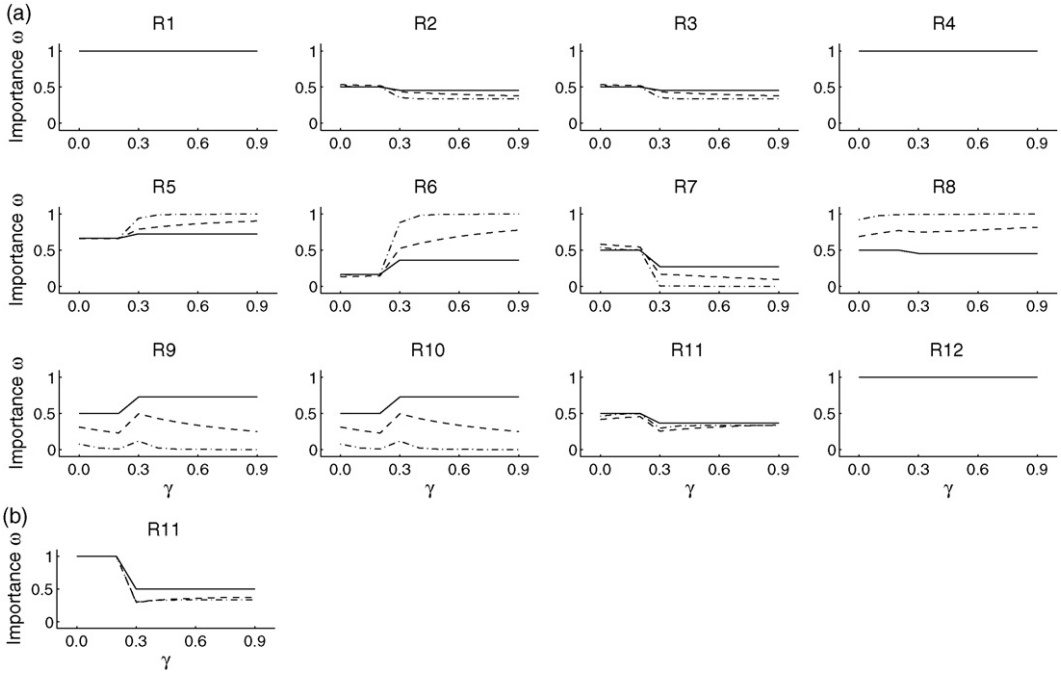

Fig. 2. (a) Reaction importances in the example network for varying γ and different k. (b) Reaction importance for reaction R11 after knockout of R7 in the example network for varying γ and different k. Solid line $k=0$, dashed line $k=2$, dashed dotted line $k=10$.

For illustration purposes, in the following we will increment γ from 0 to 0.9 with a step size of 0.1. In principle, γ can be incremented up to 1, but as we are here interested in coupled product and biomass synthesis we do not consider exclusive production of P. EMs are calculated for each of the resulting ten network scenarios and all EMs with a non-zero V yield $Y^{V/S}$ are selected for further investigations. Recall that product synthesis is implicitly contained in V and that the effective molar yield of synthesized P can be computed from the total V yield by $Y^{P/S} = \gamma \alpha_P Y^{V/S}$. Due to the shift from pure biomass synthesis to coupled product *and* biomass synthesis (with increasing proportion of synthesis of P), the EMs will change their coefficients and certain EMs may even disappear or newly appear (see Supplementary Material). It is this change in the spectrum of all EMs that we will analyze in the following in order to identify suitable interventions. For a given proportion scenario γ, the number of EMs with non-zero yield $Y^{V/S}$ is in the following expressed by $n(\gamma)$. Without kinetic information, we may consider all EMs as equiprobable. However, in order to account for the natural and economic objective, EMs exhibiting a high yield $Y^{V/S}$ should be favored. This can be achieved by assigning a weight to each EM, e.g. by relating the EM's yield to the sum of yields over all EMs. The weight $\nu_{i,\gamma}$ for each $\mathrm{EM}_i^{\gamma}(i = 1, \ldots, n(\gamma))$ in a given proportion scenario γ is thus defined by

$$\nu_{i,\gamma} = \frac{\left(Y_i^{V/S}\right)^k}{\sum_{i=1}^{n(\gamma)} \left(Y_i^{V/S}\right)^k} \tag{3}$$

with $k \geq 0$ as a parameter adjusting the intensity of quantitative weighting. If k is set to zero each EM is equally weighted irrespective of its specific yield. With increasing k the weight accounts more for yield optimality of EMs, whereas lowering k results in a stronger weighting also of non-optimal modes assigning thus a higher emphasis on network flexibility. In a second step, the EM weights are used to assign an importance measure $\omega_\gamma(r_j)$ to each reaction r_j, $j \in 1, \ldots, q$, for each proportion scenario γ. The importance measure estimates the reaction's contribution to productivity and we define it as the sum of all weights of EMs containing reaction r_j (index h runs over all modes EM_h^{γ} in which the rate of reaction r_j is non-zero):

$$\omega_\gamma(r_j) = \sum_{h:\mathrm{EM}_h^{\gamma}(r_j) \neq 0} \nu_{h,\gamma}. \tag{4}$$

A simple transformation leads to an equivalent representation of $\omega_\gamma(r_j)$:

$$\omega_\gamma(r_j) = \sum_{h:\mathrm{EM}_h^{\gamma}(r_j) \neq 0} \nu_{h,\gamma} = \frac{\sum_{h:\mathrm{EM}_h^{\gamma}(r_j) \neq 0} \left(Y_h^{V/S}\right)^k}{\sum_{h=1}^{n(\gamma)} \left(Y_h^{V/S}\right)^k}$$

$$= \frac{m_j(\gamma)}{n(\gamma)} \frac{\overline{\left(Y_j^{V/S}\right)^k}}{\overline{\left(Y^{V/S}\right)^k}}, \tag{5}$$

where $m_j(\gamma)$ expresses the number of EMs for scenario γ in which reaction r_j participates. Eq. (5) reveals how the importance measure captures the reaction's contribution to flexibility and yield. The reaction participation $m_j(\gamma)/n(\gamma)$ is independent of the weighting factor k and represents the contribution of reaction r_j to the flexibility of the network, whereas the second factor accounts for the (average) yield, $\overline{\left(Y_j^{V/S}\right)^k}$, of the pathways containing r_j normalized to the average yield of all pathways in scenario γ.

The values of the importance measure $\omega_\gamma(r_j)$ are in the interval of $[0, 1]$. Essential reactions have always an importance of 1 (independently of k) and reactions which are not included in any V-producing EM have an importance measure of zero. Note that with parameter k set to zero, $\omega_\gamma(r_j)$ is identical to the reaction participation. The approach of assigning weights to elementary modes and importances to reactions as described above is related to the approach of control-effective fluxes (Stelling et al., 2002). In the latter, efficiencies were assigned to EMs and relevances to reactions. For the particular purpose of strain optimization, we adapted name (weights and importances, respectively) and definition (Eqs. (3) and (4)) of both ratings.

In the example network, reaction importances were calculated for $k = 0$, 2 and 10 (Fig. 2a; detailed explanations and a list of EMs are

Table 1
Rating values Z_1 and Z_2 for the reactions in the example network (P is the product of interest) with k set to 0, 2 and 10. Note that Z_1 values reflect reaction importances at $\gamma = 0.9$ (Eq. (6)). Positive Z_2 values indicate overexpression candidates whereas negative values refer to knockout candidates. For example calculations see Supplementary Material.

Reaction	Z_1			Z_2		
	$k=0$	$k=2$	$k=10$	$k=0$	$k=2$	$k=10$
R1	1	1	1	0	0	0
R2	0.45	0.38	0.33	−0.05	−0.16	−0.18
R3	0.45	0.38	0.33	−0.05	−0.16	−0.18
R4	1	1	1	0	0	0
R5	0.73	0.91	≈1	0.06	0.24	0.33
R6	0.36	0.78	≈1	0.20	0.65	0.85
R7	0.27	0.09	≈0	−0.23	−0.49	−0.54
R8	0.45	0.82	≈1	−0.05	0.13	0.07
R9	0.73	0.25	≈0	0.23	−0.06	−0.07
R10	0.73	0.25	≈0	0.23	−0.06	−0.07
R11	0.36	0.34	0.33	−0.14	−0.08	−0.13
R12	1	1	1	0	0	0

given in Supplementary Material). Again, synthesis of metabolite P is implicitly contained in synthesis of V for $\gamma > 0$. The maximally achievable yield of P on substrate S is $Y^{P/S} = 1$ whereas the pathway utilizing reactions R1, R4 and R9 has a low yield of $Y^{P/S} = 1/3$. As can be seen in Fig. 2a, there are four essential reactions for biomass and thus for V synthesis (R1, R4, R12 and, trivially, μ) each having the maximal importance measure of 1 independent of the chosen proportion scenario γ and weighting parameter k. The importance measures of the remaining reactions are more differentiated. R2, R3 and R11 exhibit medium importance with slightly decreasing tendency for larger proportion of P. Apparently, the importance of these three reactions remains almost constant for different values of k. The importance of reaction R8 remains also almost constant with increasing γ, however, for larger k (emphasis on yield-optimal flux distributions) a significantly higher importance can be observed. This can be explained by the fact that all optimal routes for P as well as for the biomass components E and G require R8. A different behavior can be seen for R5, R6 and R7: for increasing γ, reactions R5 and R6 are assigned a higher and R7 a lower importance because a shift in the production ratio between precursor E and product P (towards synthesis P) requires recycling of an increasing amount of E via R5 and R6 in order to drive R8. Reaction R7 becomes for the same reason more and more dispensable. This effect is enhanced with increasing k because recycling of E ensures optimal yield which can only be obtained by the employment of R8. Finally, the importances of R9 and R10 – which are identical as they constitute an enzyme (or reaction) subset (Pfeiffer et al., 1999) – show slightly non-monotone behavior. For $k = 0$ (reaction participation) the importance increases because R9 and R10 help to synthesize P which is required in increasing amounts when shifting γ towards 1. However, as the utilization of R9 and R10 implies a lower yield these reactions are assessed to be of lower importance when using larger k. The case of R9 and R10 also demonstrates, that the importance measure $\omega_\gamma(r_j)$ for a given k may show non-monotone dependence on γ, especially if certain EMs disappear or new EMs appear (e.g. when switching γ from 0.2 to 0.3 in Fig. 2a).

In the next step we use the reaction importances to compute intervention ratings resulting in a ranking of suitable knockout and overexpression candidates. Several intervention ratings are possible; here we will focus on two which are – in our opinion – the most intuitive ones. For a reaction r_j, the first scoring scheme Z_1 simply takes the absolute importance measure $\omega_\gamma(r_j)$ evaluated at a high value for γ (high proportion of product synthesis), e.g. $\gamma = 0.9$:

$$Z_1 = \omega_{0.9}(r_j). \tag{6}$$

If we assume a yield-driven strategy ($k = 10$), the lowest rating would then follow for R7, R9 and R10 rendering them suitable knockout candidates (Table 1). The essential reactions (R1, R4, and R12) as well as R5, R6 and R8 show high Z_1 scoring and are thus potential overexpression candidates. In contrast, if the emphasis is on network flexibility ($k = 0$), R9 and R10 would even become candidates for overexpression.

Microorganisms grow often optimally with respect to biomass yield or growth rate (Ibarra et al., 2002). Such knowledge of wild-type behavior can be taken into account as reference solution. For example, for biomass-yield optimal strains we may assume that the cell distributes its fluxes (qualitatively) according to the reaction importances at $\gamma = 0$ for high k. If we now want to shift the metabolism towards larger production rates of P (e.g. $\gamma = 0.9$), we should compute the difference between the reaction importances of the desired phenotype ("to be implemented") and of the wild-type. Rating Z_2 thus reads

$$Z_2(r_j) = \omega_{0.9}(r_j) - \omega_0(r_j) \tag{7}$$

and lies in the interval $[-1, 1]$. Negative values indicate potential knockout candidates and positive values point to potential overexpression candidates. Using Z_2 instead of Z_1 results in a similar but partially different ranking of knockout and overexpression candidates in the example network (Table 1; for an example calculation of Z_1 and Z_2 see Supplementary Material). Since R1, R4 and R12 are essential in the wild-type, we can assume that these reactions are already active (with high gene expressions rates for the respective enzymes) and there is thus no need to intervene at these points ($Z_2 = 0$). Similar arguments follow for R8. For R9 and R10, a knockout has only minor relevance when using a yield-driven strategy (large k) as we can then assume that these reactions will also be inactive already in the wild-type. In agreement with Z_1, R5 and R6 would be prominent overexpression candidates and R7 a suitable knockout candidate. As illustrated by this example, Z_1 and Z_2 may differ in the targets suggested; identical targets will be identified if $\omega_{0.9}(r_j)$ is high and $\omega_0(r_j)$ is low or vice versa. As discussed above, ratings Z_1 and Z_2 can directly be used to identify single knockout or overexpression candidates. If multiple knockouts are allowed we may proceed as follows (illustrated with the example network): With Z_2, R7 is identified as best knockout candidate for all k (this even holds for Z_1). Subsequently, EMs containing R7 are excluded and reaction importances and rating Z_2 are recomputed. The reaction importances remain almost unchanged in the second iteration (not shown), except for R11. After knockout of R7, reaction R11 becomes essential for $\gamma < 0.3$ whereas for $\gamma \geq 0.3$ the importance of R11 drops to non-essential values (Fig. 2b). Hence, with additional knockout of R11 no EM for network scenarios with $\gamma < 0.3$ remains. Though not mandatory in our approach (in contrast to several others approaches), this is a desirable result, as it states that biomass and product synthesis are coupled and, if the network operates

Table 2
Published *E. coli* strains for biosynthesis of succinate from glucose under anaerobic and aerobic conditions.

Strain	Modifications	Conditions	Method	Yield (mol/mol)	Reference
KJ060	*Δ(ldhA, adhE, ackA, focA, pflB)*	Anaerobic, glucose medium, CO_2 atmosphere	Genetic modification and adaptive evolution	1.61	Jantama et al. (2008)
KJ073	*Δ(ldhA, adhE, ackA, focA, pflB, mgsA, poxB)*	Anaerobic, glucose medium, CO_2 atmosphere	Genetic modification and adaptive evolution	1.20	Jantama et al. (2008)
SBS110MG (pHL413)	*Δ(ldhA, adhE), +(Ll. pyc)*	Dual phase, anaerobic production, glucose and CO_2	Genetic modification	1.24	Sánchez et al. (2006)
SBS550MG (pHL413)	*Δ(ldhA, adhE, iclR, ackA-pta), +(Ll. pyc)*	Dual phase, anaerobic production, glucose and CO_2	Genetic modification	1.61	Sánchez et al. (2006)
SBS990MG (pHL413)	*Δ(ldhA, adhE, ackA-pta), +(Ll. pyc)*	Dual phase, anaerobic production, glucose and CO_2	Genetic modification	1.70	Sánchez et al. (2006)
HL27659k (pKK313)	*Δ(ackA-pta, iclR, poxB, sdhAB, ptsG), +(S.v. pepc)*	Aerobic, glucose, yeast extract	Genetic modification	0.91	Lin et al. (2005)
HL512769k (pKK313)	*Δ(ackA-pta, iclR , poxB, sdhAB, ptsG icd), +(S.v. pepc)*	Aerobic, glucose, yeast extract	Genetic modification	0.96	Lin et al. (2005)

under growth conditions, at least 30% of the substrate were directed to the production of metabolite *P*. In general, the procedure stops if the desired number of interventions is reached or if no further non-essential reaction exists.

We now summarize our procedure for finding intervention strategies with pseudo-code:

(1) *Include V as artificial external metabolite (container) which is produced from biomass and the desired product P*
(2) *Increment proportion γ of P at V in discrete steps and compute the EMs for each considered scenario γ (in principle, computing EMs only for $\gamma = 0$ and $\gamma = 0.9$ would suffice to compute Z_1 and Z_2)*
(3) *Select EMs according to operational constraints (and interventions performed earlier in step (5))*
(4) *Choose an appropriate k (adjusting the importance of yield optimality vs. network flexibility) and compute the EM weights and then the reaction importances (separately for each scenario γ)*
(5) *Compute intervention rating Z_1 or/and Z_2 and identify knockout (low Z-values) and/or overexpression (high Z-values) candidates*
(6) *If number of interventions is reached → Stop, else go to step 3.*

Extensions or alternative ratings are possible; they can be adjusted to fit specific demands. Furthermore, in the example network we stopped the procedure after identification of two knockouts. The impact of the first knockout R7 for the next iteration was considered by selecting a reduced set of EMs not containing R7. We could have chosen an overexpression candidate in the first iteration as most valuable intervention strategy. The consideration of overexpressions in subsequent iterations is more difficult than for knockouts, as we cannot quantitatively predict how the specific throughputs of the EMs will change (whereas we can safely discard EMs involving removed reactions). One approach accounting for overexpressions is to consider in subsequent iterations only those EMs where the corresponding reaction rate is unequal to zero. If combinations of knockouts and overexpressions are allowed, one may first identify a desired number of knockout candidates before potential overexpression candidates are identified.

2.2. Case study: succinate overproduction with E. coli

Succinate production capabilities of *E. coli* were extensively studied in recent years with different strains (Table 2) for either anaerobic or aerobic fermentation conditions. These findings were reviewed in more detail in Wendisch et al. (2006) and Jantama et al. (2008). For anaerobic fermentation, knockout of *ldhA*, *adhE* and *ackA* were shown to be valuable (Sánchez et al., 2006; Jantama et al., 2008). Overexpression of heterologous *Lactococcus lactis pyc* gene (coding for pyruvate carboxylase) further increased the succinate production capabilities. Deletion of the *iclR* gene led to induction of the glyoxylate cycle, but was not necessary for higher succinate yields, indicating partial activation of the *aceBAK* operon under anaerobic conditions (Sánchez et al., 2005). A combination of metabolic engineering and metabolic evolution led to a further improvement of *E. coli* succinate production strains (Jantama et al., 2008). Knockout of *focA/pflB* (a slash indicates stoichiometrically equivalent (i.e. alternative) knockouts), *ackA* and *poxB* was applied to repress the excretion of formate and acetate, respectively. Aerobic succinate production strains with the highest yields contained knockouts in *pta/ackA*, *sdhAB*, *ptsG* and *iclR*. With these modifications the TCA cycle was interrupted, acetate formation was reduced and the glyoxylate shunt was activated. Overexpression of the heterologous *Sorghum vulgare pepC* gene (a *ppc* homolog) led to further improvements in aerobic succinate yield and productivity.

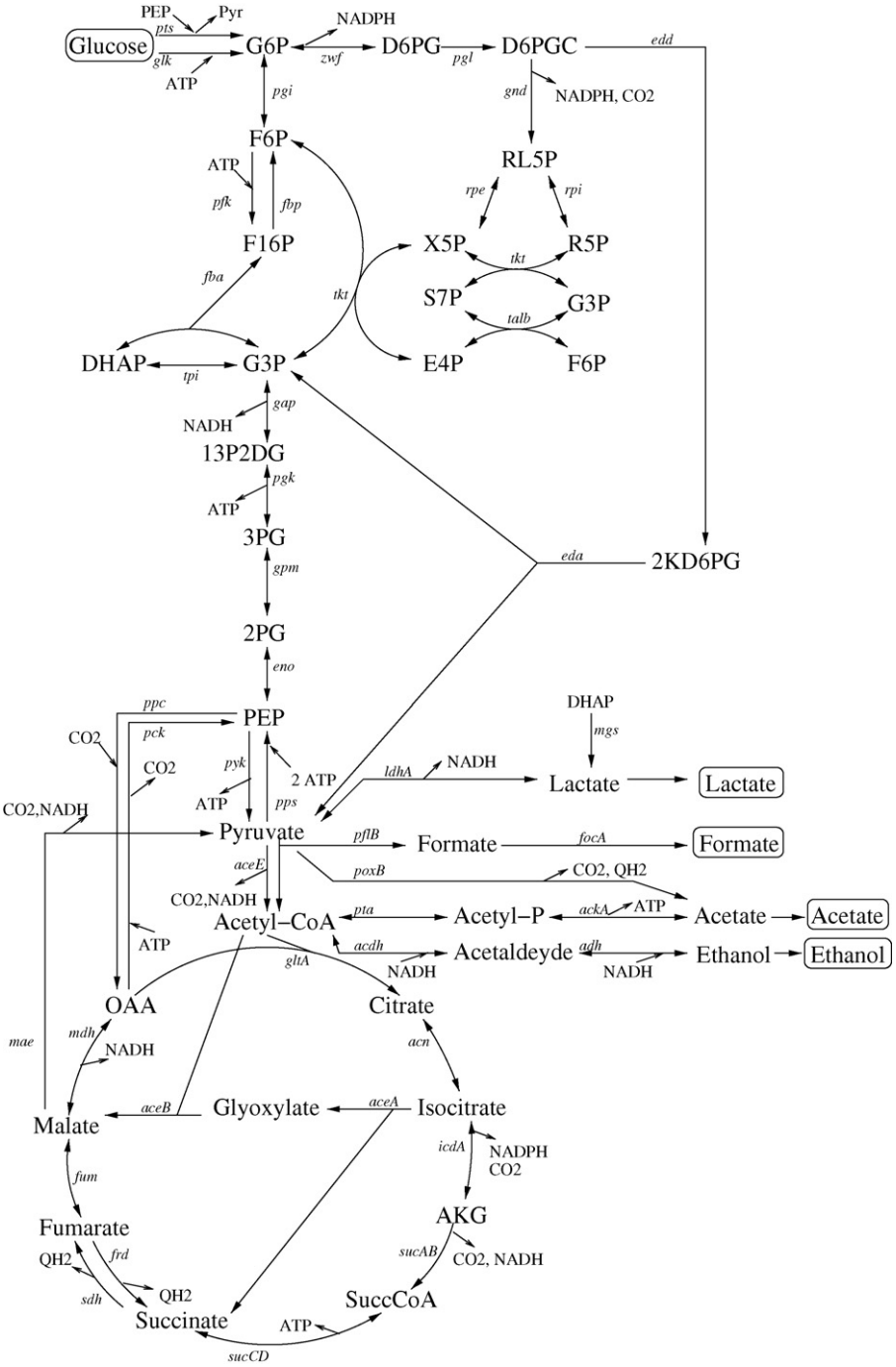

Fig. 3. Central metabolism of *E. coli*. For a detailed definition of the stoichiometric model see Appendix A.

We use the case of succinate overproduction as a proof of principle for our method rather than to discuss particular strain strategies. The described experimentally validated production strains serve for the verification of computed results. We started with setting up a stoichiometric model of the central metabolism of *E. coli* consisting of 89 metabolites and 107 reactions (Appendix A). Glucose was defined as exclusive substrate and all major aerobic and anaerobic pathways were included (Fig. 3). In the following, reactions will be named after the genes encoding the corresponding enzymes. According to the proposed procedure, we inserted an auxiliary external metabolite V which is produced from succinate and biomass. The relative proportion of succinate at V was incremented in steps of 0.1 from $\gamma = 0$ to $\gamma = 0.9$. EMs were computed for each proportion scenario; regulatory/operational constraints were considered by focusing on certain subsets of the full set of EMs (see below). Afterwards, reaction importance measures were calculated for $k = 0$ and $k = 10$ and then the resulting intervention rating Z_2 (Eq. (7); in the discussion we will focus on Z_2 with $k = 10$). With our approach we can easily take into account operational constraints as well as known regulatory constraints allowing the separate consideration of anaerobic or aerobic formation of succinate. Accordingly, we considered four different scenarios:

(1) Anaerobic production conditions (AN): only those EMs were considered where the rate of oxygen uptake reaction is zero.

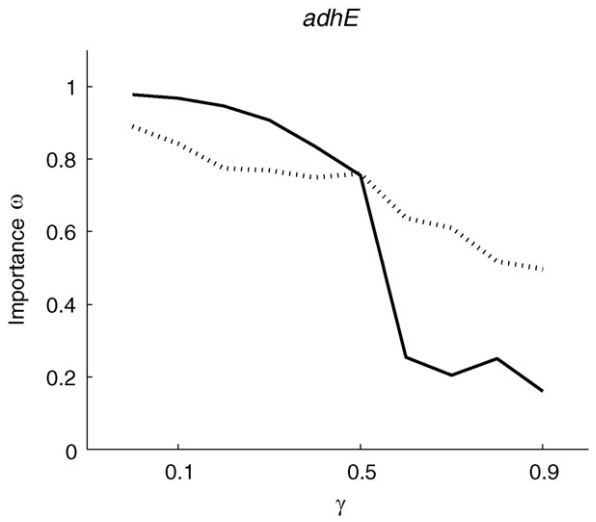

Fig. 4. Reaction importance of *adhE* for varying proportions γ of succinate production (scenario ANr; dotted line: $k = 0$, solid line: $k = 10$).

(2) Anaerobic production conditions with regulatory constraints (ANr): in addition to oxygen uptake the reactions *aceA*, *aceB*, *sucAB*, *aceE*, *sdhAB* and *poxB* were supposed to be inactive (*aceA* and *aceB* because of growth on glucose) and only those EMs were selected for further computations where the corresponding reaction rates were all zero.
(3) Aerobic production conditions (AE): only those EMs were considered where the oxygen uptake reaction has non-zero rate.
(4) Aerobic production conditions with regulatory constraints (AEr): the EMs considered must have non-zero rate for oxygen uptake and a zero rate for reactions *aceA*, *aceB*, *ldhA*, *mgsA*, *adhE* and *pflB* as they are supposed to be inactive under these conditions.

In order to test the robustness of the framework against assumed reaction availabilities derived from regulatory constraints, we compared the results of AE (AN) with those of AEr (ANr). All computations were performed with *CellNetAnalyzer* (Klamt et al., 2007) where the CASOP procedure has been implemented as a new feature.

3. Results

3.1. Succinate overproduction with E. coli

Initially, we discuss the intervention rating (implying the knockout ranking) for anaerobic conditions (Table 3). For both anaerobic scenarios (AN and ANr), as expected, the fermentative ethanol (*adhE*) and the acetate (*pta/ackA + poxB*) excretion pathways are assigned larger negative rating values Z_2 and are thus identified as suitable knockout candidates, in agreement with known mutant strains. These candidates become even more apparent, if emphasis is laid on optimal yield (compare $k = 0$ and $k = 10$ in Table 3). The minimal Z_2 value is assigned to the ethanol excretion pathway (*adhE*) which is therefore ranked as the top knockout candidate. Fig. 4 shows the dependency of *adhE* importance on the chosen proportion scenario. Ethanol excretion is highly important as anaerobic fermentation pathway as long as succinate is produced in small amounts. However, with increasing amounts of succinate to be produced ($\gamma > 0.5$), ethanol excretion becomes highly unfavorable and its importance drops down to values close to zero for larger k. Accordingly, the rating $Z_2(r_{adhE}) = \omega_{0.9}(r_{adhE}) - \omega_0(r_{adhE})$ is close to the minimum value of -1 (for $k = 10$) indicating a strong knockout candidate.

For the scenario without regulatory constraints (AN), other potential knockout targets suggested in Table 3 are formate excretion (*pflB*) and succinate dehydrogenase *sdhAB* (these two only occur because regulatory constraints are not considered in scenario AN), *pykA* (indicating that the flux from glycolysis should be directed to oxaloacetate and not to pyruvate) and *ptsG* (indicating that glucose uptake via PTS is unfavorable).

A reaction with very high Z_2 rating is the CO_2 uptake reaction which is accordingly classified as a top "overexpression" candidate (and CO_2 excretion as a "knockout candidate") suggesting that production conditions with additionally CO_2 supply are favorable. Other suggested overexpression candidates are *mdh*, *ppc* (both contribute to redirection of fluxes from/via PEP to succinate), *aceA* and *aceB* (glyoxylate shunt), *glk* (glucose uptake via hexokinase) and, as expected with very high Z_2 rating, fumarate reductase (*frdA*).

In general, the results for AN and ANr are comparable and do not change top ranked knockouts, e.g. *adhE*, into overexpression candidates or vice versa.

For the identification of multi-target engineering strategies, we concentrated on combinations of knockout targets for $k = 10$. Rating Z_2 was recomputed after each knockout. We stopped the procedure after the identification of five interventions.

As mentioned before, *adhE* was identified as best knockout candidate in the first iteration (for both AN and ANr, Table 4). The lactate excretion pathways (*ldhA + mgsA*) were identified as second and third knockout candidates, again for both scenarios (Table 4). Note that the rating value of lactate excretion changes from "minor knockout candidate" (see Table 3) to a "strong knockout candidate" (e.g. $-0.55/-0.79$ for $k = 0/10$ for ANr) only after removing *adhE*. This example underlines the necessity of recalculating reaction importances and ratings when introducing an intervention. Furthermore, for AN and ANr, knockout of *adhE* and *ldhA + mgsA* leads to a strong coupling of cell growth and succinate production because EMs producing V only exist for $\gamma \geq 0.6$ (i.e. the fraction of succinate at the virtual biomass-succinate container V must be at least 0.6 g/g V meaning in reality that 0.6 g of succinate must be produced per 0.4 g of biomass).

For scenario ANr, with knockouts of *adhE*, *ldhA* and *mgsA*, the acetate secretion pathways (*pta/ackA + poxB*) were identified in the third iteration as fourth and fifth knockout candidates. After incorporation of these five knockouts, only 10 EMs remained for growth and the coupling increased to 0.9. When omitting regulatory constraints (scenario AN), formation of formate (*pflB/focA*) is selected as knockout before acetate excretion (*pta/ackA + poxB*) (Table 4). Normally, *pflB* is essential for anaerobic growth with glucose. By adaptive evolution this limitation can be overcome (Jantama et al., 2008). Considering the network without anaerobic regulatory constraints mimics the potential of the organism to activate latent pathways and, indeed, *pflB* was identified as knockout target for this unconstrained network. Generally, the employed regulatory constraints in ANr had no significant impact on the identified engineering strategies for anaerobic conditions.

For aerobic conditions with (AEr) and without (AE) regulatory constraints, the succinate dehydrogenase reaction (*sdhAB*) was identified as the top knockout candidate as it was assigned the minimal Z_2 value (besides CO_2 excretion as already mentioned, but this reaction is not accessible as knockout (Table 3)). Acetate excretion pathways were assigned negative values in both scenarios which is in good agreement with mutants designed for aerobic succinate production (Table 2). Surprisingly, as for the anaerobic case, the ratings for uptake and excretion of CO_2 suggest CO_2 supply also for aerobic operational conditions. This can be explained by a strong weighting of pathways using the CO_2 fixing anaplerotic reactions for succinate production. For example, *ppc* occurs as overexpression candidate in the AE scenario. Since the implications of effectively applied overexpression of *ppc* are similar to overexpression of het-

Table 3
Rating values for reactions in the central metabolism of *E. coli* for succinate production ($k = 0/10$). Many zeros in ANr/AEr columns occur because the corresponding reactions become essential in these scenarios, implying a zero for Z_2 (the importance itself is 1).

Reaction	Gene	Anaerobic				Aerobic			
		Z_2 (AN)		Z_2 (ANr)		Z_2 (AE)		Z_2 (AEr)	
		$k=0$	$k=10$	$k=0$	$k=10$	$k=0$	$k=10$	$k=0$	$k=10$
Glc (external) + PEP → G6P + Pyr	*ptsG*	−0.15	−0.22	−0.29	−0.47	−0.08	−0.15	−0.08	−0.12
Glc (external) + ATP + H_ex → G6P	*glk*	0.14	0.58	0.15	0.64	0.04	0.04	0.01	−0.02
G6P ↔ F6P	*pgi*	0.07	0.02	0.02	−0.11	0.07	0.12	0.01	0.01
F6P + ATP → F16P	*pfkA*	0.04	0.04	0.05	0.04	0.08	0.08	0.04	0.06
F16P ↔ DHAP + G3P	*fbaA*	0.04	0.04	0.05	0.04	0.08	0.08	0.04	0.06
DHAP ↔ G3P	*tpiA*	0.09	0.06	0.10	0.06	0.01	0.00	0.05	0.07
G3P ↔ DPG + NADH	*gapA*	0.00	0.00	0.00	0.00	−0.02	0.00	−0.03	0.00
DPG ↔ 3PG + ATP	*pgk*	0.00	0.00	0.00	0.00	−0.02	0.00	−0.03	0.00
3PG ↔ 2PG	*gpmA*	0.00	0.00	0.00	0.00	−0.02	0.01	−0.03	0.00
2PG ↔ PEP	*eno*	0.00	0.00	0.00	0.00	−0.02	0.01	−0.03	0.00
PEP → Pyr + ATP	*pykA*	0.08	−0.29	−0.15	−0.30	−0.01	−0.06	−0.04	−0.05
Pyr + 2ATP → PEP	*pps*	0.07	0.19	0.16	0.16	0.14	0.22	0.17	0.32
Pyr → AcCoA + NADH + CO_2	*aceE*	−0.01	0.41	0.00	0.00	−0.05	−0.13	−0.06	−0.17
AcCoA + OxA → Cit	*gltA*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cit ↔ ICit	*acnA*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ICit ↔ alKG + NADPH + CO_2	*icd*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
alKG → SuccCoA + NADH + CO_2	*sucAB*	−0.05	0.05	0.00	0.00	−0.10	−0.28	0.08	−0.11
SuccCoA ↔ Succ + ATP	*sucCD*	0.17	0.32	0.00	0.00	0.15	0.23	0.18	0.06
Succ → Fum + $QuiH_2$	*sdhAB*	−0.72	−0.60	0.00	0.00	−0.74	−0.93	−0.68	−0.86
Fum + $QuiH_2$ → Succ	*frdA*	0.77	1.00	1.00	1.00	0.58	0.94	0.70	0.98
Fum ↔ Mal	*fumA*	−0.27	0.00	0.00	0.00	−0.28	−0.05	−0.16	−0.01
Mal ↔ OxA + NADH	*mdh*	−0.06	0.36	0.46	0.47	−0.11	−0.03	0.17	0.33
ICit → Succ + Glyox	*aceA*	0.20	0.30	0.00	0.00	0.20	0.05	0.00	0.00
AcCoA + Glyox → Mal	*aceB*	0.20	0.30	0.00	0.00	0.20	0.05	0.00	0.00
G6P ↔ PGlac + NADPH	*zwf*	−0.03	0.01	−0.02	0.08	−0.05	−0.05	−0.02	−0.05
AcCoA + NADH ↔ Adh	*adhE*	−0.01	−0.84	−0.39	−0.82	−0.03	0.00	0.00	0.00
NADH + Adh ↔ Eth	*adhE*	−0.01	−0.84	−0.39	−0.82	−0.03	0.00	0.00	0.00
PGlac → PGluc	*pgl*	−0.03	0.01	−0.02	0.08	−0.05	−0.05	−0.02	−0.05
PGluc → Rl5P + NADPH + CO_2	*gnd*	0.00	0.19	−0.04	0.18	−0.03	−0.08	−0.01	−0.05
Rl5P ↔ X5P	*rpe*	−0.03	−0.09	−0.01	0.13	−0.01	−0.06	−0.02	−0.04
Rl5P ↔ R5P	*rpiA*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
R5P + X5P ↔ G3P + S7P	*tktA*	−0.04	−0.07	−0.01	0.17	−0.03	−0.07	−0.01	−0.05
G3P + S7P ↔ F6P + E4P	*talA*	−0.04	−0.07	−0.01	0.17	−0.03	−0.07	−0.01	−0.05
E4P + X5P ↔ F6P + G3P	*tktA*	−0.01	−0.08	0.01	0.14	0.02	−0.02	−0.01	0.00
OxA + ATP → PEP + CO_2	*pck*	0.04	0.07	0.18	0.13	−0.01	0.04	0.11	0.13
PEP + CO_2 → OxA	*ppc*	0.35	0.18	0.00	0.00	0.32	0.44	0.00	0.00
AcCoA ↔ AcP	*pta*	−0.01	−0.27	0.11	−0.37	0.00	0.02	0.02	0.03
AcP ↔ ATP + Ac	*ackA*	−0.01	−0.27	0.11	−0.37	0.00	0.02	0.02	0.03
Pyr → AcCoA + Form	*pflB*	−0.11	−0.51	0.00	0.00	0.03	0.09	0.00	0.00
Pyr + NADH ↔ Lac	*ldhA*	−0.05	0.04	−0.15	0.08	−0.01	0.09	0.00	0.00
Mal → Pyr + NADH + CO_2	*maeA*	−0.10	−0.29	−0.30	−0.38	−0.07	−0.19	−0.24	−0.27
Pyr → $QuiH_2$ + CO_2 + Ac	*poxB*	0.22	0.03	0.00	0.00	−0.01	0.10	−0.07	0.12
DHAP → Lac	*mgsA*	0.18	0.13	0.17	0.18	0.17	0.08	0.08	−0.03
2KD6PG → G3P + Pyr	*eda*	−0.14	−0.27	−0.10	−0.10	−0.16	−0.10	−0.09	−0.10
PGluc → 2KD6PG	*edd*	−0.14	−0.27	−0.10	−0.10	−0.16	−0.10	−0.09	−0.10
NADH + H_ex ↔ NADPH	*pntA*	−0.05	−0.16	−0.04	0.27	0.27	0.14	0.45	0.39
3H_ex ↔ ATP	*atp*	0.01	0.27	0.04	0.31	0.00	0.00	0.00	0.00
Lac → Lac (external)	*ldhA + mgsA*	−0.05	0.02	−0.15	0.07	0.03	−0.02	0.08	−0.03
Ac → Ac (external)	*ackA/pta + poxB*	−0.06	−0.31	0.09	−0.39	−0.13	−0.07	−0.20	−0.11
CO_2 → CO_2(external)		−0.38	−0.66	−0.71	−0.61	−0.25	−0.71	−0.31	−0.84
CO_2 (external) → CO_2		0.13	0.89	0.64	0.79	0.04	0.24	0.05	0.34

erologous *pyc* (CO_2 fixation and anaplerosis) this classification is in accordance with practically applied mutants.

We also computed multi-gene knockout strategies for the aerobic scenarios AEr and AE (again with $k = 10$; Table 4). *sdhAB* was identified in both cases as the first knockout candidate as discussed before. If regulatory constraints were taken into account (AEr), *maeA* was identified as second knockout candidate followed by knockout of the acetate excretion pathways (*pta/ackA + poxB*). If more than 4 knockouts are allowed, obligatory coupling occurs after additional knockouts of *pgi* and *sucAB* with $\gamma \geq 0.1$ and $\gamma \geq 0.3$, respectively (Table 4). Interestingly, these two interventions were not discussed in the literature so far.

If the regulatory constraints were omitted (AE), *maeA* was identified as second knockout candidate followed by *ptsG*; both are known from existing mutants. Candidates *pykA* and *mgsA* were selected in iterations four and five leading to a final coupling of $\gamma \geq 0.3$. AEr led partially to different knockout candidates demonstrating the necessity of considering regulatory constraints. However, both suggested strategies would end up with an obligatory coupling of 0.3, but only if at least 5 knockouts are allowed. The higher obligatory couplings under anaerobic conditions are in agreement with experimentally observed yields which are generally higher for anaerobic succinate production. In general, coupling is not a primary objective of CASOP but may occur as a useful side effect.

We investigated the implications of the iterative steps of the identified engineering strategies on the relative number of remaining EMs. For scenario ANr (Fig. 5a), after knockout of *adhE* 50% of EMs for proportion scenarios with high amounts of succinate at *V* was retained, whereas less than 10% of EMs retained for the net-

Table 4
Multiple knockout strategies identified *in silico* for succinate production with *E. coli* for different scenarios with $k = 10$ and rating Z_2 (stopped after 5 identified knockouts).

Iteration	Reaction	Gene	Minimal coupling
Anaerobic with regulatory constraints (ANr)			
1	AcCoA + NADH ↔ Adh	*adhE*	0.0
2	Lactate → Lactate (external)	*ldhA* + *mgsA*	0.6
3	Acetate → Acetate (external)	*pta/ackA* + *poxB*	0.9
Anaerobic (AN)			
1	AcCoA + NADH ↔ Adh	*adhE*	0.0
2	Lactate → Lactate (external)	*ldhA* + *mgsA*	0.6
3	Pyr → AcCoA + Form	*pflB/focA*	0.8
4	Acetate → Acetate (external)	*pta/ackA* + *poxB*	0.9
Aerobic with regulatory constraints (AEr)			
1	Succinate → Fumarate + QH_2	*sdhAB*	0.0
2	Malate → Pyruvate + CO_2 + NADH	*maeA*	0.0
3	Acetate → Acetate (external)	*pta/ackA* + *poxB*	0.0
4	G6P ↔ F6P	*pgi*	0.1
5	alKG → SuccCoA + NADH + CO_2	*sucAB*	0.3
Aerobic (AE)			
1	Succinate → Fumarate + QH_2	*sdhAB*	0.0
2	Malate → Pyruvate + CO_2 + NADH	*maeA*	0.0
3	Glucose (external) + PEP → G6P + Pyr	*ptsG*	0.0
4	PEP → Pyr + ATP	*pykA*	0.0
5	DHAP → Lac	*mgsA*	0.3

work not producing succinate ($\gamma = 0$). With the additional knockout of lactate excretion pathways, there exist only EMs for networks with a proportion of at least 60% succinate at V and 20% of all EMs for the scenario with $\gamma = 0.9$ were retained. The complete knockout strategy, consisting of *adhE*, *ldhA*, *mgsA*, *pta/ackA* and *poxB* results in the highest coupling of 0.9 but only 10 EMs or 0.1% of all EMs (summed over all proportion scenarios) were retained for this scenario. Although this strain will produce succinate with high yield, one may stop before because this strain might be vulnerable and have a low productivity, since there is almost no flexibility left in the network.

For the aerobic scenario AE, after knockout of *sdhAB* only about 10% of EMs for $\gamma = 0$ remained whereas more than 80% for $\gamma = 0.9$ were preserved (Fig. 5b). Additional knockout of *maeA* further reduced the amount of EMs down to 6% for $\gamma = 0$ but had the disadvantage of reducing the network flexibility and thus potentially the capacity of the mutant, since it excluded a high fraction of EMs also for proportion scenarios with high production of succinate (Fig. 5b). Knockout of *ptsG* in third iteration had a similar effect. The number of EMs were further reduced, but since there were still some EMs

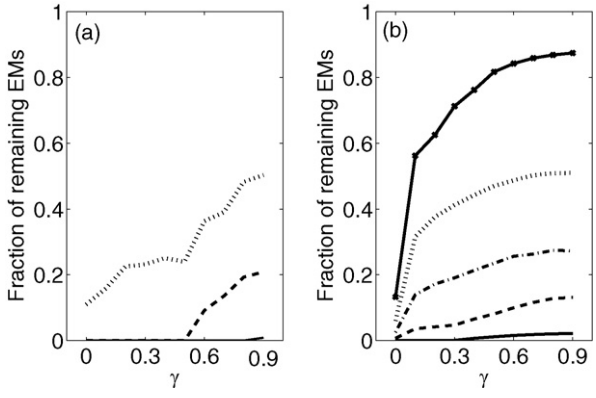

Fig. 5. Distribution of remaining EMs with multiple knockout strategies. (a) Succinate production with scenario ANr: dotted line – *adhE* knockout; dashed line – *adhE*, *ldhA*, *mgsA* triple knockout; solid line – *adhE*, *ldhA*, *mgsA*, *pta/ackA*, *poxB*. (b) Succinate production with scenario AE: solid-asterisked line – *sdhAB* knockout; dotted line – *sdhAB*, *maeA* double knockout; dashed dotted line – *sdhAB*, *maeA*, *ptsG* triple knockout; dashed line – *sdhAB*, *maeA*, *ptsG*, *pykA* quadruple knockout; solid line – *sdhAB*, *maeA*, *ptsG*, *pykA*, *mgsA* knockout strategy.

for $\gamma = 0$ no obligatory coupling occurs. In comparison, the number of EMs for high proportions was reduced by 50%. The same arguments hold for the fourth knockout candidate *pykA*. Finally, with knockout of *mgsA* a coupling of $\gamma \geq 0.3$ can be achieved.

These considerations suggest carefully checking at which iteration the introduction of new interventions should be stopped. In general, our approach does not result in *one* optimal solution because there are two degrees of freedom: one by the free parameter k (adjusting the relative importance of yield optimality vs. network flexibility for productivity) and one regarding how many interventions are to be performed. For the latter it is useful to analyze the distribution of (remaining) EMs as in Fig. 5.

Analogous to succinate overproduction we also used CASOP for identifying suitable interventions for overproduction of lactate with *E. coli* (Supplementary Material). Again, most of the identified metabolic engineering strategies resemble existing mutant strains.

3.2. Identifying potential undersupply or excess of cofactors and co-metabolites

A further useful application of our framework not discussed so far is to investigate how cofactor (e.g. NAD/NADH, ADP/ATP) or co-metabolite (e.g. CO_2) requirements change when product synthesis is to be increased. For this purpose one may temporarily include artificial reactions, for example, mimicking a sink for electrons (NADH → NAD) or for high-energy phosphate (ATP → ADP) or, conversely, mimicking a "free" source of electrons (NAD → NADH) or high-energy phosphate (ADP → ATP). For co-metabolites of interest one may insert artificial exchange reactions for uptake and excretion (if they are not yet part of the original model as is, for example, CO_2 uptake and excretion). The importances of these reactions can then be evaluated in the same manner as described in the previous sections. In this way, we investigated (separately) whether there is a potential excess or undersupply of ATP, NADH and CO_2 for succinate production under anaerobic conditions (scenario AN; Fig. 6). For high succinate production rates, this analysis suggests that the availability of NADH and CO_2 may become limiting factors as the importance of CO_2 uptake and of the artificial reaction converting NAD to NADH increases for higher succinate proportions (symmetrically, the relevance of the corresponding reactions in reverse direction decreases). The possible limitation of NADH is non-intuitive as under anaerobic growth conditions an

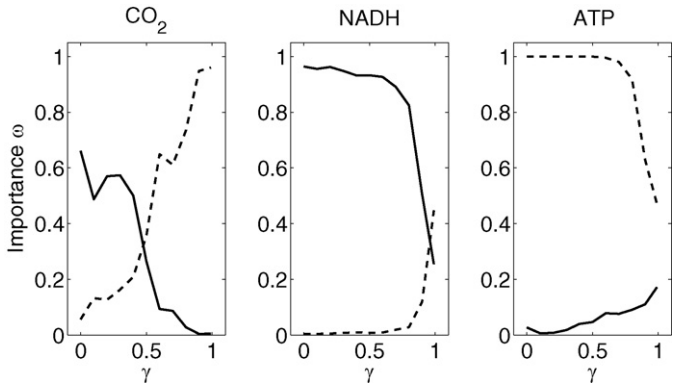

Fig. 6. Identification of potential undersupply or excess of cofactors or co-metabolites for anaerobic succinate production. For this purpose, the importances of the following (partially artificial) reactions were computed and evaluated (scenario AN, $k = 10$). Solid lines: CO_2 excretion, NADH → NAD and ATP → ADP, respectively. Dashed lines: CO_2 uptake, NAD → NADH and ADP → ATP, respectively. See main text for further explanations.

excess of NADH is generally assumed and actually correctly identified by the high importance of the reaction converting NADH to NAD for low succinate production ($\gamma < 0.7$). The favored supply of NADH for succinate overproduction can be explained by the demand of 2 mol NADH for the production of 1 mol succinate when utilizing the yield-optimal pathway via reductive TCA. Regarding CO_2, we already identified in the previous section the high importance of CO_2 uptake suggesting that providing CO_2 could be highly advantageous. Moreover, after knockout of the ethanol excretion pathway, the availability of CO_2 becomes even essential for high succinate yields under anaerobic conditions (results not shown). This finding explains why the experimentally applied growth condition with an atmosphere of 100% CO_2 is advantageous (Sánchez et al., 2006; Jantama et al., 2008). A recent work of Lu et al. (2009) confirms both predictions made for CO_2 and NADH: an *E. coli* strain (containing mutations in *pfl*, *ptsG* and *ldh*) showed under increased CO_2 concentration a dramatic increase in succinate production during anaerobic growth on glucose (productivity increased from 1.9 mg/(g h) to 225 mg/(g h)) and this behavior is coupled to a higher activity of NAD(P)H producing pathways.

Contrary, ATP might become a cofactor in excess (Fig. 6) if the cell produces large amounts of succinate. As expected, the free production of ATP from ADP is favored if the substrate is mainly converted into biomass but its importance drops significantly when fluxes are redirected to succinate production.

4. Discussion

Several computational methods have been developed to identify genetic targets for metabolic engineering. The common approach is to focus on yield-optimal flux distributions. However, eliminating all conversion routes being not optimal with respect to yield, implies the risk of unstable strains or those that have a low specific productivity. In this work we have introduced a new stoichiometric method for identifying targets for metabolic engineering. Our framework CASOP is in some aspects similar to Optknock (Burgard et al., 2003) and to the approach of minimal metabolic functionality introduced by Trinh et al. (2006). But, as demonstrated in this study, it has major differences and addresses new aspects for strain optimization:

(i) CASOP seeks to find a trade-off between high yield and high flexibility (capacity) with the goal to achieve maximal production rates. Obviously, using a stoichiometric modeling framework, we are not able to predict quantitatively the productivity of wild-type or mutant strains. However, our approach seeks to assess structural properties of metabolic networks that have an influence on the productivity and to exploit this information for strain optimization.
(ii) An advantage of CASOP is the identification of overexpression candidates what is often not possible (or has not been considered so far) in other approaches.
(iii) A further useful application of our methodology not provided by others is the assessment of cofactor/co-metabolite requirements in conjunction with product synthesis which may help to identify non-intuitive metabolic limitations.

Although the computation of knockout candidates is different, our approach resembles the approach of Trinh et al. (2006) if one chooses a large k. Furthermore, albeit not a primary objective of CASOP, our procedure may deliver intervention strategies by which product synthesis becomes more likely if the cell produces precursor for biomass formation, similar as in the OptKnock procedure. An advantage of optimization-based methods such as flux balance analysis and OptKnock is their applicability to genome scale models, whereas methods based on EM analysis can only be applied to reduced models. However, promising engineering targets identified with genome-scale models appear typically in the central metabolism (see, e.g., Burgard et al., 2003) and it seems therefore reasonable to confine EM analysis to the central metabolism. The advantage of EM analysis is that it accounts for multiple optimal and also sub-optimal solutions and that it reveals insights in aspects of network flexibility. Moreover, the proposed framework can easily be extended by the preprocessing steps performed by OptStrain if more knowledge is available (Pharkya et al., 2004). Operational constraints as well as any other limiting constraints (e.g. regulatory constraints) can easily be incorporated into the presented framework by reducing the set of active reactions or EMs (Cox et al., 2006).

The presented case studies for *E. coli* demonstrate the validity of the proposed method. Many of the computationally predicted knockout targets reproduced experimentally verified engineering strategies. In addition, several new intervention candidates were proposed that await experimental verification. The choice of ranking scheme and weighting parameter k depends on the particular situation and modelers preferences. However, our results indicate that most intervention candidates are identified in different rankings (Z_1, Z_2) and for a certain range of parameter k. If different k's change the importance and thus ranking of a reaction significantly, it is nevertheless useful to realize that a reaction plays an ambivalent role with respect to productivity (e.g. involvement in many but sub-optimal pathways for synthesizing a product). Furthermore, if the contribution of each pathway (EM) to the overall productivity could be estimated in a more mechanistic way (e.g. by kinetic data or thermodynamic considerations) the weights in Eq. (3) could be adapted accordingly and the free parameter k would be dispensable.

Although our procedure did not incorporate kinetic information, the agreement between suggested modifications and experimentally successfully applied engineering strategies indicate that a network model of the central carbon metabolism can be sufficient to identify meaningful intervention targets. The correct predictions on the increased demand of the cofactors CO_2 and NADH for high succinate specific productivity is another example of the added value of our method since such predictions cannot directly be made by existing approaches. Herein we concentrated on succinate and lactate overproduction with *E. coli*, but our method is applicable to arbitrary stoichiometric networks and products. Therefore, the presented framework CASOP may become a valuable tool for

metabolic engineering and strain optimization in a wide variety of applications.

Acknowledgements

We are grateful to Katja Bettenbrock and Axel von Kamp for helpful comments during the preparation of the manuscript. This work was supported by the German Federal Ministry of Education and Research ("FORSYS"), MaCS (Magdeburg Centre for Systems Biology) and the Ministry of Education and Research of Saxony-Anhalt (Research Center "Dynamic Systems").

Appendix A.

Stoichiometric network model of the central metabolism in *E. coli*.

A.1. Reactions

Reactions marked with an asterisk are considered to be accessible for overexpression or knockout. Several cofactors and co-metabolites (such as NAD, NADP, and ADP) are not considered in the reaction stoichiometires as these metabolites would occur in conservation relations (with ATP, NADH, NADPH, etc). Gene names are given for reactions of central pathways (which also serve as abbreviations for the reactions in the main text). For several reactions we considered only one representative gene (and neglected, for example, isozymes). Repressing lactate excretion requires interventions at two reactions/genes indicated by a '+'; the same holds for acetate excretion (formate and ethanol excretion can be repressed by removing the (unique) reaction producing the respective metabolite).

Number	Reaction	Gene(s)
1*	Glucose (external) + PEP → G6P + Pyr	*ptsG*
2*	Glucose (external) + ATP + H_ex → G6P	*glk*
3*	G6P ↔ F6P	*pgi*
4*	F6P + ATP → F16P	*pfkA*
5*	F16P ↔ DHAP + G3P	*fbaA*
6*	DHAP ↔ G3P	*tpiA*
7*	G3P ↔ DPG + NADH	*gapA*
8*	DPG ↔ 3PG + ATP	*pgk*
9*	3PG ↔ 2PG	*gpmA*
10*	2PG ↔ PEP	*eno*
11*	PEP → Pyr + ATP	*pykA*
12*	Pyr + 2 ATP → PEP	*pps*
13*	Pyr → AcCoA + NADH + CO_2	*aceE*
14*	AcCoA + OxA → Cit	*gltA*
15*	Cit ↔ ICit	*acnA*
16*	ICit ↔ alKG + NADPH + CO_2	*icd*
17*	alKG → SuccCoA + NADH + CO_2	*sucAB*
18*	SuccCoA ↔ Succ + ATP	*sucCD*
19*	Succ → Fum + $QuiH_2$	*sdhAB*
20*	Fum + $QuiH_2$ → Succ	*frdA*
21*	Fum ↔ Mal	*fumA*
22*	Mal ↔ OxA + NADH	*mdh*
23*	ICit → Succ + Glyox	*aceA*
24*	AcCoA + Glyox → Mal	*aceB*
25*	G6P ↔ PGlac + NADPH	*zwf*
26*	AcCoA + NADH ↔ Adh	*adhE*
27*	NADH + Adh ↔ Eth	*adhE*
28*	PGlac → PGluc	*pgl*
29*	PGluc → Rl5P + NADPH + CO_2	*gnd*
30*	Rl5P ↔ X5P	*rpe*
31*	Rl5P ↔ R5P	*rpiA*
32*	R5P + X5P ↔ G3P + S7P	*tktA*
33*	G3P + S7P ↔ F6P + E4P	*talA*
34*	E4P + X5P ↔ F6P + G3P	*tktA*
35*	OxA + ATP → PEP + CO_2	*pck*
36*	PEP + CO_2 → OxA	*ppc*
37*	AcCoA ↔ AcP	*pta*
38*	AcP ↔ ATP + Ac	*ackA*
39*	Pyr → AcCoA + Form	*pflB*
40*	Pyr + NADH ↔ Lac	*ldhA*
41*	Mal → Pyr + NADH + CO_2	*maeA*
42*	Pyr → $QuiH_2$ + CO_2 + Ac	*poxB*
43*	DHAP → Lac	*mgsA*
44*	2KD6PG → G3P + Pyr	*eda*
45*	PGluc → 2KD6PG	*edd*
46*	NADH + H_ex ↔ NADPH	*pntA*
47*	3 H_ex ↔ ATP	*atp*
48	2 PEP + E4P + ATP + NADPH → Chor	
49	R5P + 2 ATP → PRPP	
50	Gly ↔ MTHF + CO_2 + NADH + N	
51	Pyr + Glu → alKG + Ala	
52	2 Pyr + NADPH + Glu → alKG + CO_2 + Val	
53	2 Pyr + AcCoA + NADPH + Glu → alKG + NADH + 2 CO_2 + Leu	
54	2 ATP + N + Asp → Asn	
55	OxA + Glu → alKG + Asp	
56	ATP + NADPH + Asp → AspSAld	

Number	Reaction	Gene(s)
57	NADPH + AspSAld → HSer	
58	di_am_pim → CO_2 + Lys	
59	SuccCoA + MTHF + HSer + Cys → Pyr + Succ + N + Met	
60	ATP + HSer → Thr	
61	Pyr + NADPH + Glu + Thr → alKG + CO_2 + N + Ile	
62	ATP + PRPP + Gln → alKG + 2 NADH + His	
63	alKG + NADPH + N → Glu	
64	ATP + N + Glu → Gln	
65	ATP + 2 NADPH + Glu → Pro	
66	AcCoA + 4 ATP + NADPH + CO_2 + N + Asp + 2 Glu → alKG + Fum + Ac + Arg	
67	Chor + PRPP + Gln + Ser → G3P + Pyr + CO_2 + Glu + Trp	
68	Chor + Glu → alKG + NADH + CO_2 + Tyr	
69	Chor + Glu → alKG + CO_2 + Phe	
70	3PG + Glu → alKG + NADH + Ser	
71	Ser → MTHF + Gly	
72	AcCoA + S + Ser → Ac + Cys	
73	5 ATP + CO_2 + PRPP + 2 MTHF + 2 Asp + Gly + 2 Gln → 2 Fum + NADPH + 2 Glu + rATP	
74	6 ATP + CO_2 + PRPP + 2 MTHF + Asp + Gly + 3 Gln → Fum + NADH + NADPH + 3 Glu + rGTP	
75	ATP + Gln + rUTP → Glu + rCTP	
76	4 ATP + N + PRPP + Asp → NADH + rUTP	
77	NADPH + rATP → dATP	
78	NADPH + rGTP → dGTP	
79	NADPH + rCTP → dCTP	
80	2 NADPH + MTHF + rUTP → dTTP	
81	8.24 AcCoA + 7.24 ATP + 13.91 NADPH → mit_FS	
82	G6P + ATP → UDPGlc	
83	3PG + 3 ATP + NADPH + N → NADH + CDPEth	
84	7 AcCoA + 6 ATP + 11 NADPH → OH_myr_ac	
85	7 AcCoA + 6 ATP + 12 NADPH → C14_0_FS	
86	PEP + R5P + 2 ATP → CMP_KDO	
87	1.5 G6P + ATP → 4 NADPH + NDPHep	
88	F6P + 2 ATP + N → TDPGlcs	
89	F6P + AcCoA + ATP + Gln → Glu + UDP_NAG	
90	PEP + NADPH + UDP_NAG → UDP_NAM	
91	Pyr + SuccCoA + NADPH + AspSAld + Glu → alKG + Succ + di_am_pim	
92	G6P + ATP → ADPGlc	
93	DHAP + NADH → Glyc3P	
94	0.14176 Glyc3P + 26.2949 ATP + 0.60097 Ala + 0.10124 Cys + 0.26647 Asp + 0.30747 Glu + 0.2048 Phe + 0.67725 Gly + 0.10473 His + 0.32116 Ile + 0.37935 Lys + 0.49804 Leu + 0.16989 Met + 0.26647 Asn + 0.24436 Pro + 0.29091 Gln + 0.32698 Arg + 0.38031 Ser + 0.28044 Thr + 0.46778 Val + 0.062835 Trp + 0.15244 Tyr + 0.1489 rATP + 0.18319 rGTP + 0.11366 rCTP + 0.12273 rUTP + 0.023904 dATP + 0.024582 dGTP + 0.024582 dCTP + 0.023904 dTTP + 0.28352 av_FA + 0.0069264 UDPGlc + 0.010368 CDPEth + 0.010368 OH_myr_ac + 0.010368 C14_0_FS + 0.010368 CMP_KDO + 0.010368 NDPHep + 0.0069264 TDPGlcs + 0.01656 UDP_NAG + 0.01656 UDP_NAM + 0.01656 di_am_pim + 0.0924 ADPGlc → Biomass	
95	$QuiH_2$ + 0.5 O_2 → 2 H_ex	
96	NADH ↔ $QuiH_2$ + 2 H_ex	
97	ATP → (maintenance)	
98	O_2 (external) → O_2	
99	N (external) → N	
100	S (external) + 4 ATP + 4 NADPH → S	
101*	CO_2 (external) → CO_2	
102*	CO_2 → CO_2 (external)	
103*	Lac → Lac (external)	*ldhA* + *mgsA*
104	Eth → Eth (external)	
105*	Ac → Ac (external)	*pta/ackA* + *poxB*
106	Form → Form (external)	*focA*
107	Succ → Succ (external)	
(CASOP)	$(1-\gamma)$ * Biomass + γ * 8.4631 Succ → *V*	

A.2. Metabolites

2KD6PG	2-Keto-3-deoxy-6-gluconate
2PG	2-Phosphoglycerate
3PG	3-Phosphoglycerate
Ac	Acetate
AcCoA	Acetyl-CoA
AcP	Acetyl-P
Adh	Acetaldehyde
ADPGlc	ADP-glucose
Ala	Alanine
alKG	Alpha-ketoglutarate (=oxoglutarate)
Arg	Arginine
Asn	Asparagine
Asp	Aspartate
AspSAld	Aspartate-4-semialdehyde

ATP	Adenosintriphosphate
C14_0_FS	14:0 Fatty-acid
CDPEth	CDP-ethanolamine
Chor	Chorismate
Cit	Citrate
CMP_KDO	CMP-KDO
CO_2	Carbon dioxide
Cys	Cysteine
dATP	ATP for DNA synthesis
dCTP	CTP for DNA synthesis
dGTP	GTP for DNA synthesis
DHAP	Dihydroxyacetone-phosphate
di_am_pim	Diaminopimelate
DPG	Diphosphoglycerate
dTTP	TTP for DNA synthesis
E4P	Erythrose-4-phosphate
Eth	Ethanol
F16P	Fructose-1,6-bisphosphate
F6P	Fructose-6-phosphate
Form	Formate
Fum	Fumarate
G3P	Glyceraldehyde-3-phosphate
G6P	Glucose-6-phosphate
Gln	Glutamine
Glu	Glutamate
Gly	Glycine
Glyc3P	Glycerol-3-phosphate
Glyox	Glyoxylate
H_ex	Periplasmatic protons
His	Histidine
HSer	Homoserine
ICit	Iso-citrate
Ile	Isoleucine
Lac	Lactate
Leu	Leucine
Lys	Lysine
Mal	Malate
Met	Methionine
av_FA	Average fatty acid
MTHF	Methylene-tetrahydrofolate
N	NH_4
NADH	NADH
NADPH	NADPH
NDPHep	NDP-heptose
O_2	Oxygen
OH_myr_ac	OH myristic acid
OxA	Oxalacetate
PEP	Phosphoenolpyruvate
PGlac	6-Phospho-gluconolactone
PGluc	6-Phospho-gluconate
Phe	Phenylalanine
Pro	Proline
PRPP	5-Phosphoribosyl-1-pyrophosphate
Pyr	Pyruvate
$QuiH_2$	Ubiquinone
R5P	Ribose-5-phosphate
rATP	ATP for RNA synthesis
rCTP	CTP for RNA synthesis
rGTP	GTP for RNA synthesis
Rl5P	Ribulose-5-phosphate
rUTP	UTP for RNA synthesis
S	Sulfur (SO_4)
S7P	Seduheptulose-7-phosphate
Ser	Serine
Succ	Succinate
SuccCoA	Succinyl-CoA
TDPGlcs	TDP-glucosamine
Thr	Threonine
Trp	Tryptophane
Tyr	Tyrosine
UDP_NAG	UDP-acetylglucosamine
UDP_NAM	UDP-N-acetylmuramic acid
UDPGlc	UDP-glucose
Val	Valine
X5P	Xylulose-5-phosphate

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jbiotec.2010.03.006.

References

Burgard, A.P., Pharkya, P., Maranas, C.D., 2003. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol. Bioeng. 84, 647–657.

Cox, S.J., Levanon, S.S., Sanchez, A., Lin, H., Peercy, B., Bennett, G.N., San, K.-Y., 2006. Development of a metabolic network design and optimization framework incorporating implementation constraints: a succinate production case study. Metab. Eng. 8, 46–57.

Fong, S.S., Burgard, A.P., Herring, C.D., Knight, E.M., Blattner, F.R., Maranas, C.D., Palsson, B.O., 2005. In silico design and adaptive evolution of Escherichia coli for production of lactic acid. Biotechnol. Bioeng. 91, 643–648.

Fong, S.S., Nanchen, A., Palsson, B.O., Sauer, U., 2006. Latent pathway activation and increased pathway capacity enable Escherichia coli adaptation to loss of key metabolic enzymes. J. Biol. Chem. 281, 8024–8033.

Ibarra, R.U., Edwards, J.S., Palsson, B.O., 2002. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature 420, 186–189.

Jantama, K., Haupt, M.J., Svoronos, S.A., Zhang, X., Moore, J.C., Shanmugam, K.T., Ingram, L.O., 2008. Combining metabolic engineering and metabolic evolution to develop nonrecombinant strains of Escherichia coli C that produce succinate and malate. Biotechnol. Bioeng. 99, 1140–1153.

Kim, H.U., Kim, T.Y., Lee, S.Y., 2008. Metabolic flux analysis and metabolic engineering of microorganisms. Mol. Biosyst. 4, 113–120.

Klamt, S., 2006. Generalized concept of minimal cut sets in biochemical networks. BioSystems 83, 233–247.

Klamt, S., Gilles, E.D., 2004. Minimal cut sets in biochemical reaction networks. Bioinformatics 20, 226–234.

Klamt, S., Saez-Rodriguez, J., Gilles, E.D., 2007. Structural and functional analysis of cellular networks with CellNetAnalyzer. BMC Syst. Biol. 1, 2.

Lee, S.Y., Kim, H.U., Park, J.H., Park, J.M., Kim, T.Y., 2008. Metabolic engineering of microorganisms: general strategies and drug production. Drug Discov. Today 14 (1–2), 78–88.

Lin, H., Bennett, G.N., San, K.-Y., 2005. Metabolic engineering of aerobic succinate production systems in Escherichia coli to improve process productivity and achieve the maximum theoretical succinate yield. Metab. Eng. 7, 116–127.

Lu, S., Eiteman, M.A., Altman, E., 2009. Effect of CO_2 on succinate production in dual-phase Escherichia coli fermentations. J. Biotechnol. 143, 213–223.

Patil, K.R., Rocha, I., Förster, J., Nielsen, J., 2005. Evolutionary programming as a platform for in silico metabolic engineering. BMC Bioinformatics 6, 308.

Pfeiffer, T., Sanchez-Valdenebro, I., Nuno, J.C., Montero, F., Schuster, S., 1999. METATOOL: for studying metabolic networks. Bioinformatics 15, 251–257.

Pharkya, P., Burgard, A.P., Maranas, C.D., 2004. OptStrain: a computational framework for redesign of microbial production systems. Genome Res. 14, 2367–2376.

Pharkya, P., Maranas, C.D., 2006. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. Metab. Eng. 8, 1–13.

Sánchez, A.M., Bennett, G.N., San, K.-Y., 2005. Novel pathway engineering design of the anaerobic central metabolic pathway in Escherichia coli to increase succinate yield and productivity. Metab. Eng. 7, 229–239.

Sánchez, A.M., Bennett, G.N., San, K.-Y., 2006. Batch culture characterization and metabolic flux analysis of succinate-producing Escherichia coli strains. Metab. Eng. 8, 209–226.

Schuster, S., Dandekar, T., Fell, D.A., 1999. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. Trends Biotechnol. 17, 53–60.

Schuster, S., Pfeiffer, T., Fell, D.A., 2008. Is maximization of molar yield in metabolic networks favoured by evolution? J. Theor. Biol. 252, 497–504.

Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., Gilles, E.D., 2002. Metabolic network structure determines key aspects of functionality and regulation. Nature 420, 190–193.

Trinh, C.T., Carlson, R., Wlaschin, A., Srienc, F., 2006. Design, construction and performance of the most efficient biomass producing E. coli bacterium. Metab. Eng. 8, 628–638.

Trinh, C.T., Unrean, P., Srienc, F., 2008. Minimal Escherichia coli cell for the most efficient production of ethanol from hexoses and pentoses. Appl. Environ. Microbiol. 74, 3634–3643.

Trinh, C.T., Wlaschin, A., Srienc, F., 2009. Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. Appl. Microbiol. Biotechnol. 81, 813–826.

Wendisch, V.F., Bott, M., Eikmanns, B.J., 2006. Metabolic engineering of Escherichia coli and Corynebacterium glutamicum for biotechnological production of organic acids and amino acids. Curr. Opin. Microbiol. 9, 268–274.