

Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*

Robert Schuetz^{1,2,3}, Lars Kuepfer^{1,3,4} and Uwe Sauer^{1,2,*}

¹ Institute of Molecular Systems Biology, ETH Zurich, Switzerland and ² Life Science Zurich PhD program on Systems Physiology and Metabolic Diseases, Zurich, Switzerland

³ These authors contributed equally to this work

⁴ Present address: Bayer Technology Services, Leverkusen 51368, Germany

* Corresponding author. Institute of Molecular Systems Biology, ETH Zurich, Wolfgang-Pauli-Strasse 16, Zurich 8093, Switzerland. Tel.: + 41 44 633 3672; Fax: + 41 44 633 1051; E-mail: sauer@imsb.biol.ethz.ch

Received 10.1.07; accepted 2.5.07

To which extent can optimality principles describe the operation of metabolic networks? By explicitly considering experimental errors and *in silico* alternate optima in flux balance analysis, we systematically evaluate the capacity of 11 objective functions combined with eight adjustable constraints to predict ¹³C-determined *in vivo* fluxes in *Escherichia coli* under six environmental conditions. While no single objective describes the flux states under all conditions, we identified two sets of objectives for biologically meaningful predictions without the need for further, potentially artificial constraints. Unlimited growth on glucose in oxygen or nitrate respiring batch cultures is best described by nonlinear maximization of the ATP yield per flux unit. Under nutrient scarcity in continuous cultures, in contrast, linear maximization of the overall ATP or biomass yields achieved the highest predictive accuracy. Since these particular objectives predict the system behavior without preconditioning of the network structure, the identified optimality principles reflect, to some extent, the evolutionary selection of metabolic network regulation that realizes the various flux states.

Molecular Systems Biology 10 July 2007; doi:10.1038/msb4100162

Subject Categories: metabolic and regulatory networks; simulation and data analysis

Keywords: ¹³C-flux; evolution; flux balance analysis; metabolic network; network optimality

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits distribution, and reproduction in any medium, provided the original author and source are credited. This license does not permit commercial exploitation or the creation of derivative works without specific permission.

Introduction

The acclaimed goal of systems biology is quantitative understanding of functional interactions between the multiple cellular components to eventually predict network, cell and organism behavior (Aebersold, 2005; Bork and Serrano, 2005). Beyond intuition, quantitative understanding inevitably requires computational models to capture the enormous numbers of molecular components that interact in a highly nonlinear manner within interlinked information and biochemical networks (Kitano, 2002; Cassman, 2005; Kholodenko, 2006). For most cellular networks, such as signaling or protein–protein interaction networks, however, we do not even know all involved components that need to be represented in a model. Hence, much of the current focus is on experimental (Aebersold and Mann, 2003; Sopko *et al.*, 2006) and computational (Kholodenko *et al.*, 2002; Stelling, 2004; Reed *et al.*, 2006; Warner *et al.*, 2006) identification of missing components and their interactions to establish the network topology as a prerequisite for mechanistic modeling.

Metabolic networks are a notable exception because their interaction topology is well established in several cases; that is, we know most reactions, the enzymes that catalyze them, the genes that encode the enzymes and how they interact stoichiometrically within a biochemical network. As incomplete as this knowledge may be, it is currently far beyond that of basically any other cellular network and allows to construct metabolic models that represent almost entire microbial genomes (Price *et al.*, 2004). With up to 1000 biochemical reactions, these genome-scale models allow to predict network capabilities, for example, by using flux balance analysis (FBA) (Fell and Small, 1986). Successful FBA applications include prediction of gene deletion lethality (Edwards and Palsson, 2000a; Forster *et al.*, 2003; Kuepfer *et al.*, 2005), end points of adaptive evolution (Ibarra *et al.*, 2002) and optimal metabolic states (Edwards *et al.*, 2001).

In contrast to dynamic models that require detailed, typically unavailable kinetic parameters, constraint-based modeling with FBA permits steady-state analysis of large-scale networks without large fitted parameter sets (Bailey, 2001; Price *et al.*, 2004; Stelling, 2004). To identify optimal solutions

in the vast solution space, FBA objective functions are defined to solve the system of linear equations that represent the mass balance constraints. While different objectives were proposed for different biological systems (Kacser and Beeby, 1984; Heinrich *et al*, 1997; Ebenhoh and Heinrich, 2001; Holzhütter, 2004; Price *et al*, 2004; Knorr *et al*, 2007), by far the most common assumption is that microbial cells maximize their growth (see below for further explanation). Since the identified optimal solutions are often inconsistent with the biological reality, the solution space is further restricted through additional constraints that reflect thermodynamic, kinetic or biochemical knowledge. Another problem is that, depending on the shape of the solution space, multiple intracellular flux distributions (alternate optima) may underlie the exact same optimal value that is identified by the objective (e.g., the maximum biomass yield) (Lee *et al*, 2000; Mahadevan and Schilling, 2003). This space of steady-state flux solutions has been explored for biological meaning (Mahadevan and Schilling, 2003; Reed and Palsson, 2004; Wiback *et al*, 2004; Bilu *et al*, 2006) and to identify candidate network states (Papin *et al*, 2002; Thiele *et al*, 2005), but was largely ignored in many FBA studies that examined arbitrary, single optimal solutions (Segre *et al*, 2002; Almaas *et al*, 2004; Papp *et al*, 2004).

In a fully complementary approach, ^{13}C -experiments are used to determine intracellular flux states that reveal *in vivo* operation of metabolic networks (Hellerstein, 2003; Sauer, 2004; Wiechert and Nöh, 2005; Sauer, 2006). In some instances, such experimental flux data were used to further restrict the FBA-computed flux solution space. For lack of experimental data, however, only one or two arbitrary flux distributions were considered (Burgard and Maranas, 2003; Wiback *et al*, 2004). Attempts to actually predict intracellular fluxes by FBA methods are few and either unverified (Papp *et al*, 2004) or tested for a single case (Beard *et al*, 2002; Segre *et al*, 2002; Holzhütter, 2004). With the recent availability of large-scale experimental flux data from various microbes (Moreira dos Santos *et al*, 2003; Blank and Sauer, 2004; Fischer and Sauer, 2005; Perrenoud and Sauer, 2005; Blank *et al*, 2005b), a more systematic analysis of the correlation between the *in silico* feasible flux space and the *in vivo* realized fluxes is now possible.

Here we examine the predictive capacity of 11 linear and nonlinear network objectives, by evaluating the accuracy of FBA-based flux predictions through rigorous comparison to ^{13}C -based flux data from *Escherichia coli* grown under six environmental conditions. By systematically testing all permutations of 11 objective functions with or without eight additional constraints, we identify the most appropriate combination(s) to predict *in vivo* fluxes by FBA. More generally, we thus assess whether assumed optimality principles of evolved network operation are generally valid or whether specific objectives are necessary for environmental conditions that require different metabolic activity.

Results

Systematic testing of objective functions and constraints for FBA

To predict intracellular fluxes through the presently known reactions of *E. coli* central carbon metabolism, we constructed

a highly interconnected stoichiometric network model with 98 reactions and 60 metabolites that supports the major carbon flows through the cell (Figure 1 and Supplementary Table I). FBA-based fluxes are typically expressed as relative fluxes that are normalized to the specific glucose uptake rate. Typically, this reference flux is known, hence absolute fluxes can be calculated by re-scaling. Due to linear dependencies in the network, the systemic degree of freedom is restricted to a limited number of reactions that define the entire flux solution. For our network, 10 reactions are sufficient to describe the actual systemic degree of freedom, as identified by calculability analysis (Van der Heijden *et al*, 1994; Klamt *et al*, 2002). These fluxes were expressed as split ratios at pivotal branch points in the network, where each of the 10 reactions that consume a cellular metabolite is divided by the sum of all producing reactions (Figure 1 and Table I). Qualitatively identical results were obtained when repeating all reported simulations directly with the 10 absolute fluxes instead of the 10 split ratios (data not shown).

Dividing a specific consumption flux by all producing fluxes scales to unity, an unbiased comparison of the 10 network fluxes with often-different magnitudes is possible. Moreover, it enhances intuition and biological interpretation because, wherever possible, the ratios were chosen to represent metabolic flux ratios that are obtained from ^{13}C -experiments (Fischer *et al*, 2004) (Table II and Supplementary Table II). For example, split ratio R1 represents the fraction of the intracellular glucose-6-phosphate (G6P) pool that is metabolized through phosphoglucosomerase (Pgi), relative to the summed production via G6P-dehydrogenase (Zwf), glucokinase (Glk) and the phosphotransferase system (Pts) (Figure 1 and Table I). The experimentally determined split ratios (Table II) can be subdivided into three groups: (i) R1, R4, R5, R6 are always active, (ii) R3 is inactive under all considered conditions and (iii) R2, R7–R10 are conditionally active.

Optimal solutions in this underdetermined system of linear equations were identified by FBA with 11 linear and nonlinear objective functions to identify optimal solutions, some of which are combinations of pairs of objectives (Table III). Depending on the shape of the solution space, linear optimization frequently leads to alternate optima; that is, alternate sets of feasible flux distributions with an identical optimal value (Lee *et al*, 2000; Mahadevan and Schilling, 2003). To quantify the overall variance of *in silico* fluxes, we first determined the absolute range of variation for the individual split ratios by maximizing and minimizing each flux separately. For example, maximization of biomass yield (which is synonymous to the frequently used term of maximization of growth rate (Price *et al*, 2004)) results in ranges of the split ratios R1, R4, R6 and R7, but unique values for the remaining six split ratios (Figure 2A). Maximization of ATP yield without further constraints, in contrast, is a much better defined example with unique values for all 10 split ratios (Figure 2B). Beyond objective functions, these flux variabilities are not only dependent on the chosen objective but also the network structure, and were also shown to exist in genome-scale models (Mahadevan and Schilling, 2003; Reed and Palsson, 2004). To further constrain the solution space, we imposed eight additional constraints on network operation (Table IV). The choice of objective functions and constraints

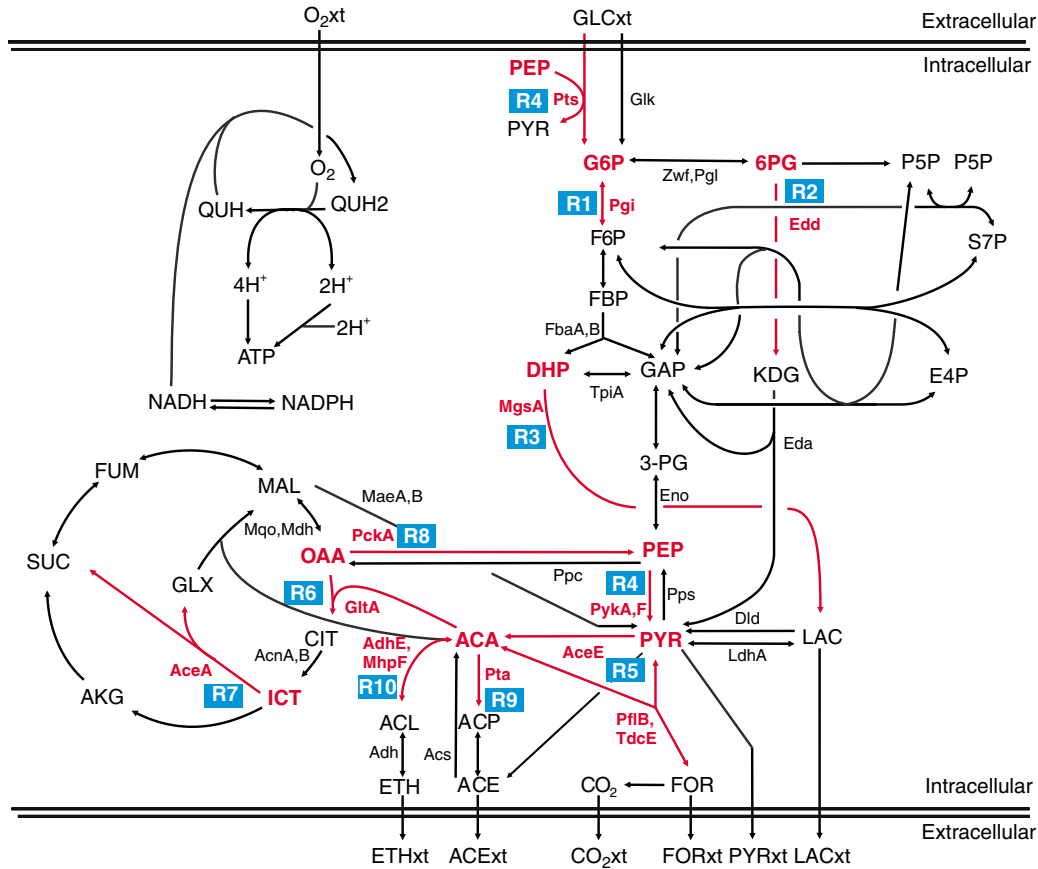


Figure 1 Central carbon metabolism of *E. coli*. The 10 reactions that describe the actual systemic degree of freedom are indicated in red arrows. These 10 reactions are expressed as 10 split ratios, where each of the 10 reactions that consume a cellular metabolite is divided by the sum of all producing reactions. The corresponding metabolites are indicated in red, whereas the 10 split ratios are shown in blue rectangles. Abbreviations: ACA, acetyl-coenzyme A; ACE, acetate; ACL, acetaldehyde; ACP, acetyl-P; AKG, alpha-ketoglutarate; CIT, citrate; DHP, dihydroxyacetone-P; ETH, ethanol; E4P, erythrose-4-P; FBP, fructose-1,6-bi-P; FOR, formate; FUM, fumarate; F6P, fructose-6-P; GAP, glyceraldehyde-3-P; GLX, glyoxylate; G6P, glucose-6-P; ICT, isocitrate; KDG, 2-keto-3-deoxy-6-phosphogluconate; LAC, lactate; MAL, malate; OAA, oxaloacetate; PEP, phosphoenolpyruvate; PYR, pyruvate; 6PG, 6-phosphogluconate; P5P, pentose-5-P; QUH, ubiquinone; QUH2, ubiquinol; S7P, seduheptulose-7-P; SUC, succinate; 3-PG, 3-phosphoglycerate; xt, external.

Table I Split ratios of intracellular fluxes that describe the systemic degree of freedom in the network

$R_1 = \frac{v_{Pgi}}{v_{Glk} + v_{Pts} + v_{Zwf} + v_{Pgi}}$	Flux into glycolysis
$R_2 = \frac{v_{Edd}}{v_{Pgi}}$	Flux into Entner–Doudoroff pathway
$R_3 = \frac{v_{MgsA}}{v_{FbaA} + v_{FbaB} + v_{TpiA}}$	Flux into methylglyoxal pathway
$R_4 = \frac{v_{PykA} + v_{PykF} + v_{Pts}}{v_{Eno} + v_{Pps} + v_{PckA}}$	PEP to pyruvate flux
$R_5 = \frac{v_{AceE} + v_{PflB} + v_{TdcE}}{v_{PykA} + v_{PykF} + v_{MaeA} + v_{MaeB} + v_{Dld} + v_{LdhA} + v_{Eda} + v_{Pts} + v_{PflB} + v_{TdcE}}$	Pyruvate to acetylcoenzyme A flux
$R_6 = \frac{v_{GltA} + v_{PrpC}}{v_{AceE} + v_{PflB} + v_{TdcE} + v_{Acs} + v_{AdhE} + v_{MhpF}}$	Flux into TCA cycle
$R_7 = \frac{v_{AceA}}{v_{AcnA} + v_{AcnB}}$	Flux into glyoxylate shunt
$R_8 = \frac{v_{PckA}}{v_{Mdh} + v_{Mqo} + v_{Ppc}}$	Oxaloacetate to PEP flux
$R_9 = \frac{v_{Pta}}{v_{AceE} + v_{PflB} + v_{TdcE} + v_{Acs} + v_{AdhE} + v_{MhpF}}$	Acetate secretion
$R_{10} = \frac{v_{AdhE} + v_{MhpF}}{v_{AceE} + v_{PflB} + v_{TdcE} + v_{Acs} + v_{AdhE} + v_{MhpF}}$	Ethanol secretion

Table II Experimentally determined flux split ratios under the six conditions considered

Growth rate Split ratio ^a	Batch			Chemostat		
	0.6 h ⁻¹ Aerobe ^b	0.3 h ⁻¹ Anaerobe NO ₃ ^b	0.2 h ⁻¹ Anaerobe ^b	0.1 h ⁻¹ C-limited ^c	0.4 h ⁻¹ C-limited ^c	0.4 h ⁻¹ N-limited ^d
R1 (flux into glycolysis)	0.70 ± 0.02	0.82 ± 0.02	0.90 ± 0.05	0.69 ± 0.12	0.64 ± 0.05	0.96 ± 0.14
R2 (flux into Entner-Doudoroff pathway)	0.13 ± 0.06	0.00 ± 0.05	0.86 ± 0.94	0.23 ± 0.20	0.19 ± 0.11	0.00 ± 0.05
R3 (flux into methylglyoxal pathway)	0.00 ± 0.05	0.00 ± 0.05	0.00 ± 0.05	0.00 ± 0.05	0.00 ± 0.05	0.00 ± 0.05
R4 (PEP to pyruvate flux)	0.78 ± 0.02	0.96 ± 0.02	0.94 ± 0.05	0.84 ± 0.14	0.70 ± 0.06	0.72 ± 0.10
R5 (pyruvate to acetylcoenzyme A flux)	0.81 ± 0.03	0.96 ± 0.02	0.96 ± 0.1	0.91 ± 0.21	0.84 ± 0.10	0.90 ± 0.15
R6 (flux into TCA cycle)	0.24 ± 0.02	0.02 ± 0.01	0.02 ± 0.01	0.64 ± 0.13	0.85 ± 0.09	0.50 ± 0.06
R7 (flux into glyoxylate shunt)	0.00 ± 0.05	0.00 ± 0.05	0.00 ± 0.05	0.46 ± 0.13	0.00 ± 0.05	0.00 ± 0.05
R8 (oxaloacetate to PEP flux)	0.00 ± 0.05	0.00 ± 0.05	0.00 ± 0.05	0.35 ± 0.08	0.12 ± 0.03	0.01 ± 0.01
R9 (acetate secretion)	0.58 ± 0.03	0.65 ± 0.01	0.34 ± 0.01	0.00 ± 0.05	0.00 ± 0.05	0.04 ± 0.01
R10 (ethanol secretion)	0.00 ± 0.05	0.30 ± 0.02	0.61 ± 0.04	0.00 ± 0.05	0.00 ± 0.05	0.00 ± 0.05

^aSplit ratios were calculated with the equations given in Table I from published absolute fluxes (Supplementary Table II).

^bExperimental data was taken from (Perrenoud and Sauer, 2005).

^cExperimental data was taken from (Nanchen *et al*, 2006).

^dExperimental data was taken from (Emmerling *et al*, 2002).

widely predefine the degree of freedom in terms of specific pathway usage (data not shown), hence appropriate objective/constraint combinations can potentially be used to approximate metabolic behavior.

We systematically assessed the predictive capability of FBA by comparing all objective/constraint permutations to ¹³C-detected *in vivo* flux distributions from six growth conditions, including glucose- and ammonium-limited chemostat cultures and batch cultures with excess nutrient supply (Table II and Supplementary Table II). For each of the 99 different optimization problems, the maximum and minimum Euclidean distance between *in silico* and *in vivo* flux solutions was evaluated by simultaneously considering the 10 split ratios. Confidence intervals of the experimental flux ratios were considered by the standardized Euclidean distance, which weights the distance between prediction and data by the square of the corresponding standard deviation δ^{exp} . The resulting value describes the overall deviation of the *in silico* predicted flux distribution (or range of flux distributions) with respect to the corresponding experimental reference solution, and is henceforth referred to as predictive fidelity (Supplementary Figure 1). This predictive fidelity depends on two factors: (i) the minimal possible standardized Euclidean distance to the *in vivo* results and (ii) the potential variance of the *in silico* fluxes that arises from alternate FBA optima.

FBA-based flux prediction for batch cultures

First, we determined the predictive fidelity of the 99 objective/constraint combinations for unlimited batch growth on glucose under aerobic, anaerobic and nitrate-respiring conditions (Sauer *et al*, 2004; Perrenoud and Sauer, 2005). Obviously, the agreement is specific for each case, since it depends (i) on the particular objective/constraint combination that defines the shape of the solution space and (ii) the experimental reference flux distribution (Figure 3, Supplementary Figure 1 and Supplementary Table III). Since minimization of glucose consumption and maximization of ATP yield per reaction step gave almost identical results as maximization of biomass yield and maximization of ATP yield, respectively, only the latter two are discussed (Supplementary Table III). The results obtained by minimization

of reaction steps, minimization of the redox potential and minimization of ATP producing fluxes were considerably worse than those obtained with the other objectives, hence are not discussed in the following (Supplementary Table III).

Without requiring additional constraints, the highest predictive fidelity for aerobic batch cultures was obtained by maximizing ATP yield per flux unit, yielding a unique flux prediction that is closest to the experimental data (Figure 3). Since this nonlinear optimization function is non-convex, it bears the danger of identifying only local optima. To confirm that indeed a global optimum was identified, we first reformulated the original objective function as a convex function that contains a linear and a nonlinear, but convex term (see Materials and methods for details). This new nonlinear but convex function cannot be optimized *per se*, since it needs *a priori* weighting of both function terms. Thus, in a second step, we performed a sensitivity analysis around the previously identified optimal solutions. Since no iteration resulted in a solution with a higher optimal value, we have strong indication that indeed global optima were identified (Supplementary Figure 7). Of the remaining objective functions, only the maximization of the ATP yield objective achieved similar predictive fidelities when combined with particular constraints. Maximization of biomass yield, in contrast, suffered from alternate optima over a wide range of constraints (Figure 3). Although unique results are feasible by invoking a P-to-O-ratio (moles of ATP produced per mole of oxygen) of unity, the predictive fidelity is still inferior to the one obtained with the maximization of ATP objectives.

The predictive fidelity is a general criterion for the predictive accuracy. It cannot, however, identify the individual metabolic functions that are responsible for the deviations. To elucidate whether these are based on large errors in single ratios or on small errors in many ratios, we plotted *in silico* and *in vivo* ratios as scatter plots where perfect predictions fall on the bisecting diagonal (Figure 4). For aerobic batch cultures, acetate secretion (R9) in combination with a sound predictive fidelity is one main discriminating variable that was only captured by the maximization of ATP yield per flux unit (Figure 4A and C). In combination with oxygen constraints, the maximization of ATP yield mimics the maximization of

Objective function ^a	Mathematical definition	Explanation	Rationale	Reference
Max biomass ^b	$\max \frac{v_{\text{biomass}}}{v_{\text{glucose}}}$	Maximization of biomass yield	Evolution drives selection for maximal biomass yield ($Y_{X/S}$)	(van Gulik and Heijnen, 1995; Edwards and Palsson, 2000b; Price <i>et al.</i> , 2004)
Max ATP	$\max \frac{v_{\text{ATP}}}{v_{\text{glucose}}}$	Maximization of ATP yield	Evolution drives maximal energetic efficiency ($Y_{\text{ATP}/S}$)	(van Gulik and Heijnen, 1995; Ramakrishna <i>et al.</i> , 2001)
$\text{Min} \sum v_i^{2c}$	$\min \sum_{i=1}^n v_i^2$	Minimization of the overall intracellular flux	Postulates maximal enzymatic efficiency for cellular growth (analogous to minimization of the Euclidean norm)	(Bonarius <i>et al.</i> , 1996; Blank <i>et al.</i> , 2005a)
Max ATP per flux unit ^c	$\max \frac{v_{\text{ATP}}}{n \sum_{i=1}^n v_i^2}$	Maximization of ATP yield per flux unit	Cells operate to maximize ATP yield while minimizing enzyme usage	(Dauner and Sauer, 2001)
Max biomass per flux unit ^c	$\max \frac{v_{\text{biomass}}}{n \sum_{i=1}^n v_i^2}$	Maximization of biomass yield per flux unit	Cells operate to maximize biomass yield while minimizing enzyme usage	
Min glucose	$\min \frac{v_{\text{glucose}}}{v_{\text{biomass}}}$	Minimization of glucose consumption	Evolution drives selection for most efficient usage of substrate	(Oliveira <i>et al.</i> , 2005)
Min reaction steps ^c	$\min \sum_{i=1}^n y_i^2, y_i \in \{0, 1\}$	Minimization of reaction steps	Cells minimize number of reaction steps to produce biomass	(Melendez-Hevia and Isidoro, 1985)
Max ATP per reaction step ^c	$\min \frac{v_{\text{ATP}}}{\sum_{i=1}^n y_i^2}, y_i \in \{0, 1\}$	Maximization of ATP yield per reaction step	Cells operate to maximize ATP yield per reaction step	
Min redox potential ^{d,e}	$\min \frac{\sum v_{\text{NADH}}}{n \frac{v_{\text{glucose}}}{v_{\text{biomass}}}}$	Minimization of redox potential ^f	Cells decrease number of oxidizing reactions thus conserving their energy or using their energy in the most efficient way possible	(Knorr <i>et al.</i> , 2007)
Min ATP production ^{d,e}	$\min \frac{v_{\text{ATP}}}{n \frac{v_{\text{glucose}}}{v_{\text{biomass}}}}$	Minimization of ATP producing fluxes ^g	Cells grow while using the minimal amount of energy, thus conserving energy	(Knorr <i>et al.</i> , 2007)
Max ATP production ^{d,e}	$\max \frac{v_{\text{ATP}}}{n \frac{v_{\text{glucose}}}{v_{\text{biomass}}}}$	Maximization of ATP producing fluxes ^h	Cells produce as much ATP as possible	(Heinrich <i>et al.</i> , 1997; Ebenhoh and Heinrich, 2001; Knorr <i>et al.</i> , 2007)

Both maximization of biomass objectives (absolute and per flux unit) require no *a priori* assumptions. For all other objectives the specific growth rate was set to the experimentally determined value under each condition.

Often also referred to as optimization of growth rate (Price *et al.*, 2004).

n refers to the overall number of reactions in the network, that is 98 in the present case.

[†]Reaction name is that specified in Supplementary Table 1; ‘_R’ refers to the reverse reaction.

All reversible reactions in Supplementary Table I were converted to two irreversible reactions resulting in a final stoichiometric model of 60 metabolites and 151 reactions.

Reactions: *gapA*, *aceE/F*, *maeA*, *sucAB*, *sucAB*, *mdh*, *udhA*, *fdhF*, *fdoGHI*, *fdn*

⁵Reactions: *pgk*, *pykA*, *pykF*, *sucCD*, *atpA-H*, *ackA*, *ackB*, *tdcD*, *purT*.

¹Reactions: *pgk*, *pykA*, *pykF*, *sucCD*, *atpA-H*, *ackA*, *ackB*, *tdcD*, *purT*.

ATP yield per flux unit objective (Figures 3, 4B and C). Minimizing the overall intracellular fluxes inherently leads to acetate secretion, however, at the cost of deviations in other ratios, in particular for Entner–Doudoroff activity (R2) (data not shown).

Akin to aerobic cultures, the maximization of ATP yield per flux unit was the only objective that achieved reasonable predictions without further constraints for anaerobic nitrate-respiring batch cultures (Figures 3 and 4E). Improved predictions are possible, however, by setting a particular constraint on the nitrate respiration rate for the maximization of biomass yield objective (Figures 3 and 4D). Largely

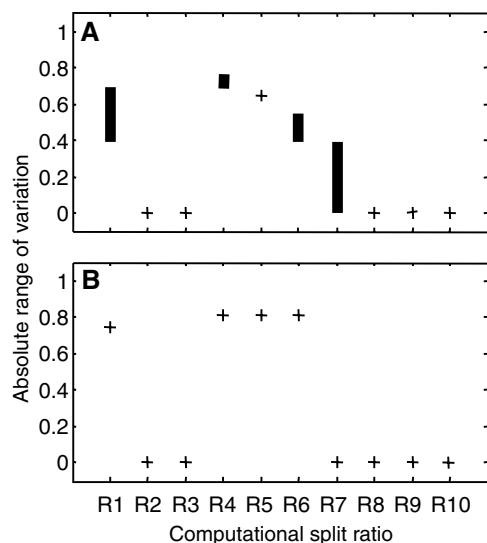


Figure 2 Absolute range of *in silico* variation in individual split ratios due to alternate optima for the maximization of biomass yield (A) and maximization of ATP yield (B) objectives without additional constraints.

independent on the invoked constraints the anaerobic batch culture was well predicted by all considered objective functions (Figures 3 and 4F). This behavior is due to the low degree of freedom in the absence of an external electron acceptor and to the experimental uncertainty of some of the fluxes (see Supplementary Table II), which allow for a relatively high predictive fidelity even if the actual agreement is low.

FBA-based flux prediction for chemostat cultures

In contrast to unrestricted nutrient supply in batch cultures, a single, defined nutrient limits the rate of growth in continuous chemostat cultures (Russell and Cook, 1995; Hoskisson and Hobbs, 2005). To evaluate predictions for the rather different metabolic behavior under such conditions, we used experimental flux data from slowly (0.1 h^{-1}) and rapidly (0.4 h^{-1}) growing chemostat cultures under glucose- (Nanchen *et al*, 2006) and ammonium limitation (Emmerling *et al*, 2002) (Table II and Supplementary Table II).

Largely independent of additional constraints, the maximization of ATP or biomass yield objectives approximated all chemostat cultures best (Figures 5 and 6). Maximization of ATP producing fluxes resulted in similar predictive fidelities as maximization of ATP yield (Supplementary Table III). Although mathematically distinct, both objectives maximize ATP production and thus lead to similar *in silico* flux predictions. Alternate optima occurred only in one case for the biomass objective and can be overcome, as for aerobic batch cultures, by constraining the P-to-O-ratio to unity. The relative independence of the predictive fidelity on constraints demonstrates that these objectives provide somewhat robust predictions for metabolism under nutrient limitation. Nevertheless, various specific objective/constraint combinations

Table IV Additional constraints implemented in constraint-based FBA

Constraint	Explanation	Reference
$P/O = 1^a$	The P-to-O ratio is equal to 1; that is, one ATP is generated per NADH in oxidative phosphorylation. The value is based on known coupling efficiencies and expression levels of respiratory chain components	(Calhoun <i>et al</i> , 1993; Gennis and Stewart, 1996)
$q(\text{glc})/q(\text{O}_2) = 2/3$	Oxygen uptake is coupled to the glucose uptake rate at a stoichiometry of 3:2. Reflects enzymatic limitation for oxygen utilization	(Varma and Palsson, 1994)
$q\text{O}_2 \text{ max} \leq 11.5$	An upper limit on oxygen uptake was set to $11.5 \text{ mmol}/(\text{g h})$. Reflects the lower end of the range of the measured enzymatic limitation for oxygen utilization	(Xu <i>et al</i> , 1999)
$q\text{O}_2 \text{ max} \leq 14.75$	An upper limit on oxygen uptake was set to $14.75 \text{ mmol}/(\text{g h})$. Reflects the upper end of the range of the measured enzymatic limitation for oxygen utilization	(Varma <i>et al</i> , 1993)
Maintenance	Including the growth-independent maintenance requirements of the cell, which was experimentally determined to be equal to $7.6 \text{ mmolATP}/(\text{g h})$	(Varma and Palsson, 1994; Nanchen <i>et al</i> , 2006)
Bounds	Constraining intracellular fluxes to a maximum of 200% of the glucose uptake rate because higher values were never found experimentally	(Emmerling <i>et al</i> , 2002; Flores <i>et al</i> , 2002; Fischer and Sauer, 2003; Zhao and Shimizu, 2003)
NADPH	NADPH overproduction of 35% relative to the NADPH requirements for biomass production by forcing UdhA-mediated re-oxidation of excess NADPH	(Nanchen <i>et al</i> , 2006)
All constraints	Simultaneous implementation of all constraints (oxygen uptake constraint of $11.5 \text{ mmol}/(\text{g h})$)	

^aFor a ratio of unity, cytochrome oxidase *bd* (Cyd) and NADH dehydrogenase I (Nuo) were set equal to zero.

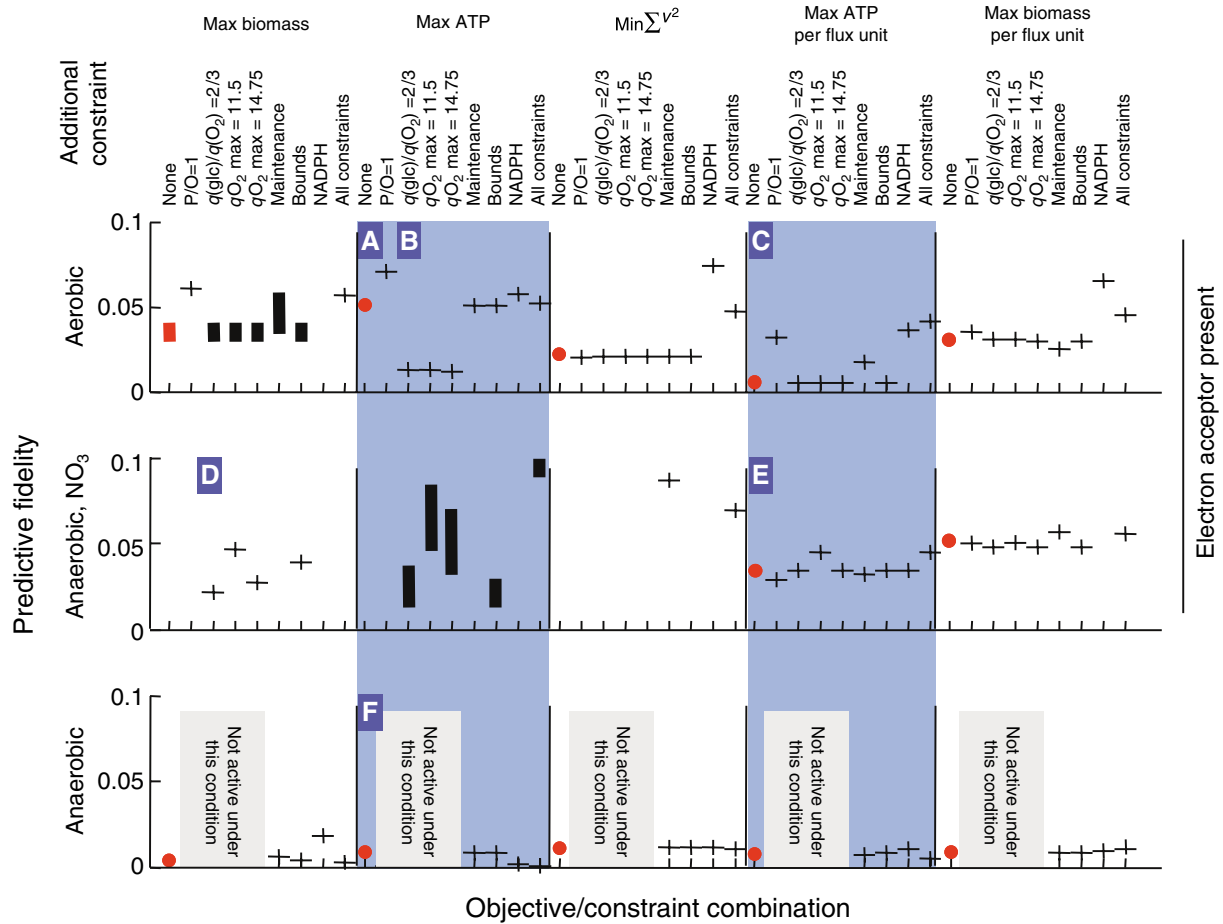


Figure 3 Predictive fidelities for aerobic and anaerobic batch cultures in minimal medium with glucose (arbitrary units). The results were obtained by minimization and maximization of the standardized Euclidean distance of the 10 split ratios for the reference flux solutions (Table II). The four oxygen constraints were not implemented for anaerobic batch cultures. Predictive fidelities above 0.1 are not shown. Crosses and red dots signify that the range of the predictive fidelity is less than 1%. Red dots and bars highlight predictive fidelities without additional constraints. Bars signify the predictive fidelity range for that particular combination of objective function, constraint and environmental condition. For the sole case of nitrate respiration, the upper oxygen uptake rates were translated into corresponding upper bounds for nitrate uptake. Objective functions and constraints are defined in Tables III and IV.

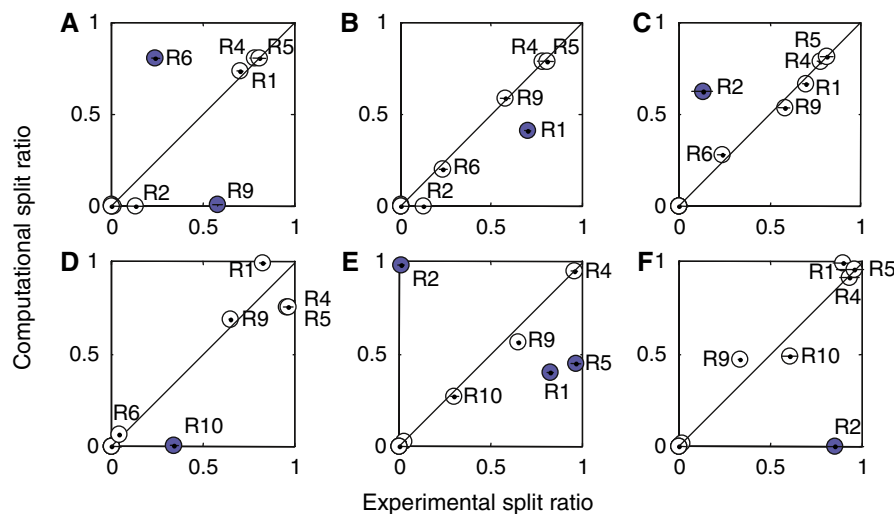


Figure 4 Scatter plots of *in vivo* and *in silico* split ratios in selected batch cultures. Letters A–F refer to the cases indicated in Figure 3. Horizontal and vertical error bars indicate experimental and computational variances, respectively. Blue circles indicate that the predicted value is more than 0.25 units different from the experimental reference value, hence were poorly predicted. Split ratios that contain zero fluxes cannot be calculated due to singularities in the error propagation, hence a default accuracy of 5% was assumed. Split ratios with a value above 0.05 are indicated.

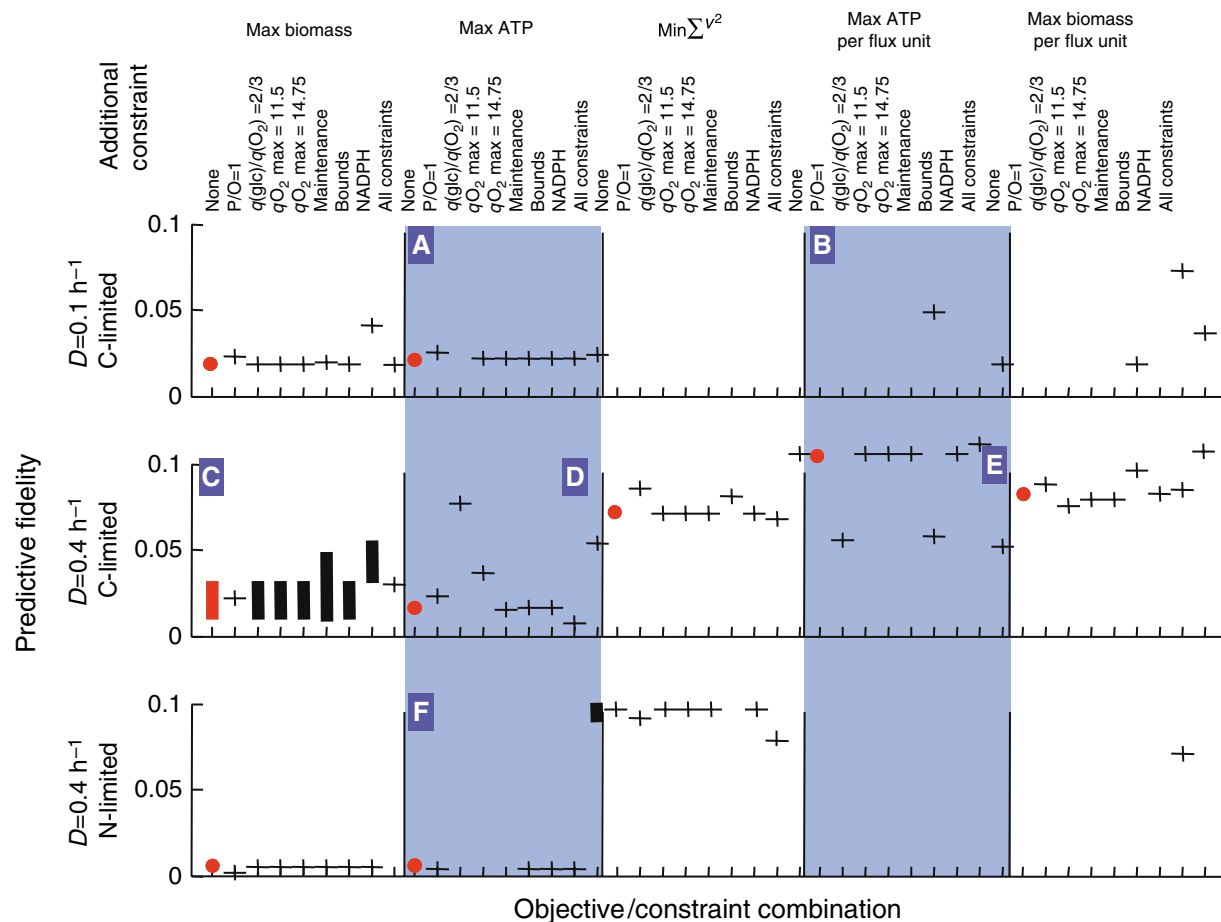


Figure 5 Predictive fidelities for glucose- and ammonium-limited chemostat cultures (arbitrary units). The results were obtained by minimization and maximization of the standardized Euclidean distance of the 10 split ratios for the reference flux solutions (Table II). Predictive fidelities above 0.1 are not shown. Crosses and red dots signify that the range of the predictive fidelity is less than 1%. Red dots and bars highlight predictive fidelities without additional constraints. Bars signify the predictive fidelity range for that particular combination of objective function, constraint and environmental condition. Objective functions and constraints are defined in Tables III and IV.

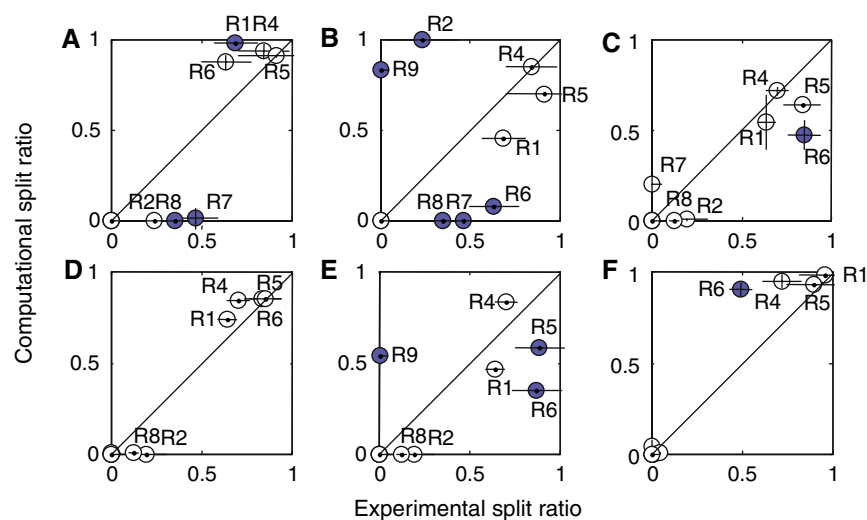


Figure 6 Scatter plots of *in vivo* and *in silico* split ratios in selected chemostat cultures. Letters **A–F** refer to the cases indicated in Figure 5. Horizontal and vertical error bars indicate experimental and computational variances, respectively. Blue circles indicate that the predicted value is more than 0.25 units different from the experimental reference value, hence were poorly predicted. Split ratios that contain zero fluxes cannot be calculated due to singularities in the error propagation, hence a default accuracy of 5% was assumed. Split ratios with a value above 0.05 are indicated.

were also capable of describing the different conditions, in particular the maximization of ATP yield per flux unit, in combination with all eight constraints for both carbon-limited chemostat cultures (Figures 5 and 6E). In sharp contrast to batch cultures, however, the maximization of ATP yield per flux unit objective was basically useless without further constraints (Figure 6B).

Robustness of predicted flux solutions

As a main discriminating variable of good objectives for aerobic batch cultures, the well-known phenomenon of acetate overflow was only captured when maximizing the ATP yield per flux unit (Figure 3). Maximizing the overall ATP yield mimicked this behavior only when combined with oxygen uptake constraints. Since particular combinations of oxygen uptake and P-to-O ratio constraints and the frequently used maximization of biomass yield objective should achieve the same effect (Varma *et al*, 1993; Varma and Palsson, 1994), we performed a sensitivity analysis by determining the predictive fidelity and acetate production for step-wise increases of the oxygen uptake constraint for four P-to-O ratios (Figure 7A and B and Supplementary Figures 2 and 3).

As for maximization of the overall ATP yield, only fine-tuning of the network by invoking particular combinations of P-to-O ratio and oxygen constraints resulted in reasonable flux predictions and acetate secretion rates for the maximization of biomass yield objective function. However, the predictive fidelity was very sensitive to changes in the parameters, such that only a narrow range of oxygen uptake constraints enforced a good fit for a given P-to-O ratio, for example unrealistically low oxygen uptake rates of 5–7 mmol/g h at a P-to-O ratio of 2 (Supplementary Figures 2 and 3). For maximization of the ATP yield objective, in contrast, the predictive fidelity was relatively insensitive to the exact value of the oxygen uptake and the P-to-O ratio constraint, with a critical threshold for the oxygen uptake constraint of around 15 mmol/g h; that is, the upper bound of experimentally observed values in glucose batch cultures (Varma *et al*, 1993; Varma and Palsson, 1994; Xu *et al*, 1999). Since such metabolic parameters often vary between strains or with small environ-

mental differences, a certain robustness of predicted flux solutions with respect to the chosen constraints is a highly desirable property for constraint-based modeling. Hence, both ATP objectives are clearly of superior robustness for the prediction of fluxes in aerobic batch cultures.

Are all fluxes equally difficult to predict?

To systematically analyze the predictive capability for individual flux ratios, we defined the specific agreement ρ_i (Supplementary Figure 4). This value considers the deviation between each single pair of *in silico* and *in vivo* values for all 10 split ratios and is weighted by the ratio of the corresponding accuracies, that is the possible range of the computational split ratio divided by the experimental standard deviation.

To obtain an overview of the most difficult to predict fluxes, independent of the specific objective and constraints, we applied cluster analysis to the Euclidean distance among specific agreements ρ (Figure 8 and Supplementary Figures 5 and 6). The 2700 considered data points represent the 10 split ratios for all objective/constraint combinations considered in Figures 3 and 5 under each environmental condition. Most difficult to predict in four out of the six conditions were the fluxes into glycolysis (R1) and acetate secretion (R9), although sometimes in different combinations (Figure 8A–C and F). Exceptions were the C-limited chemostats, where fluxes through the glyoxylate shunt (R7) and the TCA cycle (R6) were most difficult to predict (Figure 8D and E). Thus, some fluxes are clearly more difficult to predict than most others, but those problematic ones often change with the environmental conditions.

Discussion

The key question addressed here is whether intracellular fluxes in metabolic steady state can be predicted from network stoichiometry alone by invoking optimality principles. Our systematic and statistically rigorous comparison of FBA-based *in silico* flux predictions from 99 objective/constraint combinations to *in vivo* fluxes from ^{13}C -experiments demonstrated that prediction of relative flux distributions is, within limits,

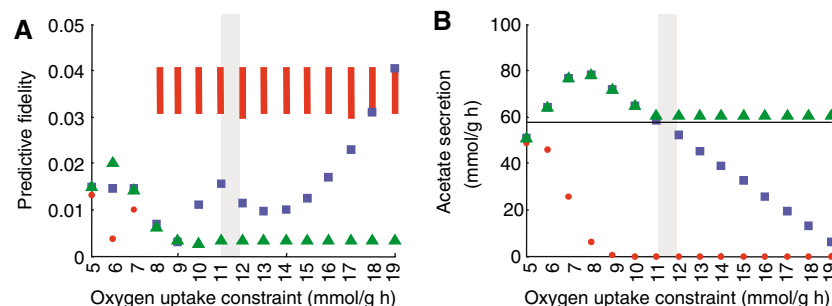


Figure 7 Sensitivity analysis of (A) the predictive fidelity (arbitrary units) and (B) acetate secretion on the oxygen uptake constraint in aerobic batch cultures given no assumption on the P-to-O ratio constraint (i.e., in practice a P-to-O ratio of 2 is chosen). The upper bound for the maximal oxygen uptake rate (5–19 mmol/g h) was varied, whereas the minimal experimentally reported maximal oxygen uptake rate (Varma *et al*, 1993; Varma and Palsson, 1994; Xu *et al*, 1999) is indicated in gray. Blue rectangles are for the maximization of ATP yield, red dots and bars for the maximization of biomass yield and green triangles for the maximization of ATP yield per flux unit. In (A), bars indicate the range of the predictive fidelities, whereas a dot, rectangle or triangle indicates a unique solution. Predictive fidelities above 0.05 are not shown. In (B), the line indicates the experimentally determined value of acetate secretion of 58 mmol/g h (Perrenoud and Sauer, 2005).

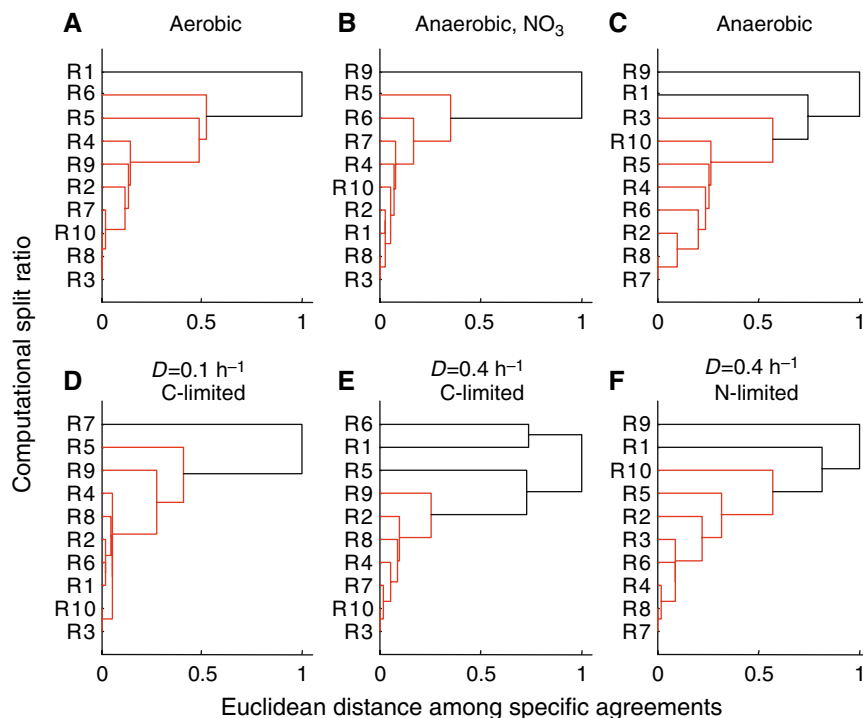


Figure 8 Hierarchical cluster trees based on the Euclidean distance among specific agreements ρ for the five objective functions considered in Figures 3 and 5, including all constraint combinations under the six conditions. Difficult to predict groups of split ratios are highlighted by black lines. Groups of nodes were assigned where the linkage among the nodes was less than 0.7, when the linkage was normalized to values between 0 and 1.

possible. Since no single objective predicted the experimental data for wild-type *E. coli* under all conditions, the pivotal element is to identify the most relevant objective for each condition.

For unlimited growth on glucose in oxygen or nitrate respiring batch cultures, by far the best objective function was nonlinear maximization of the ATP yield per unit of flux, which is a combination of the linear maximization of overall ATP yield and minimization of the overall flux. In some cases, similar predictions could be achieved by combining the overall maximization of ATP and biomass yield objectives with particular oxygen uptake constraints. Under nutrient scarcity in nutrient-limited continuous cultures, in contrast, the linear maximization of ATP or biomass yield were clearly superior. As a result of the low degree of freedom in non-respiring batch cultures, all objective functions lead to equally well flux predictions.

As an unexpected key result, model preconditioning through additional and potentially artificial constraints is not necessary if the appropriate objective function is chosen for a given condition. Invoking additional constraints for suitable objectives achieved only subtle improvements or avoided alternative optima in few cases. When combined with particular constraints, even suboptimal objectives could be forced to yield equally accurate predictions for some conditions; in the sole case of nitrate respiration even better predictions. Setting of these additional constraints, however, is condition- and objective specific, thus requires considerable *a priori* knowledge to be biologically meaningful. Except for subtle differences in predictive fidelity, all major conclusions

are independent of using normalized ratios instead of absolute flux values (data not shown).

We explicitly considered the fundamental FBA problems of alternate optima and experimental accuracy by size reduction through calculability analysis and by including confidence regions for *in vivo* fluxes in the standardized Euclidean distance, respectively. An important question is whether or not our results are model dependent. To address this point, we verified the key conclusions with two genome-scale models of *E. coli* metabolism (Edwards and Palsson, 2000b; Reed *et al*, 2003). Although specific properties such as flux variability clearly depend on the network structure of the particular stoichiometric network model (see below for details), the above-identified objectives also achieved the best predictions for fluxes in the central carbon metabolism with either genome-scale model (data not shown).

Clearly, alternate optima occurred also in genome-scale network models (see Materials and methods for details). Independent of the model size, however, *in silico* variability can be avoided such that uniquely defined solutions are obtained when low P-to-O ratios are assumed or when all internal proton fluxes are balanced. With the proton-balanced genome-scale model of Reed *et al* (2003), for example, such unique flux solutions can be obtained. The solution space spanning all alternate optima has previously been scanned for biological meaning (Mahadevan and Schilling, 2003; Reed and Palsson, 2004; Wiback *et al*, 2004), and identified interesting correlations with levels of gene expression that point to evolutionary constraints on how tight certain reactions need to be regulated (Bilu *et al*, 2006). A potential complication is that

the solution space itself is not, or not entirely, an inherent network feature, but also a function of the arbitrarily chosen objectives and constraints.

An important distinction must be made between FBA-based prediction of the typically investigated general physiology (i.e., extracellular uptake and production rates and the growth rate) and the here attempted prediction of the underlying intracellular flux distribution, which has several-fold more variables. Hence, there is no immediate contradiction between good prediction of growth physiology obtained by maximizing the growth yield (Price *et al*, 2004) or minimizing the redox production rate (Knorr *et al*, 2007), and their here demonstrated limited capacity to predict the underlying flux state. In some cases, alternate flux optima are responsible for the apparent discrepancy and in others it is primarily the specific combination of chosen constraints that explain why good physiology predictions were achieved with suboptimal objectives. By demonstrating that intracellular fluxes can be approximated with experimentally validated optimality assumptions, we go beyond flux prediction algorithms such as minimization of metabolic adjustment (Segre *et al*, 2002) or regulatory on/off minimization (Shlomi *et al*, 2005) that require a reference flux distribution in the wild type to predict fluxes in mutants. Carefully chosen objectives achieve intrinsically good prediction not only of growth physiology, but also of intracellular fluxes in wild-type *E. coli* without preconditioning the system through additional constraints apart from the experimentally determined growth rate. Since the network model was identical in all cases, the identified optimality functions potentially reflect the evolved regulatory processes that realize the particular flux states under different environmental conditions.

Under nutrient scarcity in chemostat cultures, metabolism normally supports efficient biomass formation with respect to the limiting nutrient (Russell and Cook, 1995). Based on our results, this operational state appears to have evolved under the objective to maximize either the ATP or biomass yield (synonymous to the frequently used maximization of growth rate objective). Under conditions of unlimited growth in aerobic or nitrate-respiring batch cultures, in contrast, energy production is clearly not optimized *per se* because cells secrete large amounts of acetate instead of using the more efficient respiratory chain.

What then is the biological interpretation of the more appropriate maximization of the ATP yield per flux unit? Optimization of this objective is realized by maximizing ATP production (the nominator) and by minimizing the overall intracellular flux (the denominator). Hence, small networks with yet high, albeit suboptimal catabolic ATP formation are identified, which has three potential biological consequences. Firstly, resources are economically allocated because expenditures for enzyme synthesis are, on average, greater for longer pathways. Secondly, suboptimal ATP yields dissipate more energy and thus enable higher catabolic rates because the difference between the free energies of substrates and products must be used for both, energy conservation by synthesizing ATP (increase the yield) and energy dissipation to drive the chemical reaction (increase the rate) (del Valle and Aledo, 2002; Pfeiffer and Schuster, 2005). Thirdly, at a constant catabolic rate, a small network results in shorter residence

times of substrate molecules until they generate ATP and probably other cofactors. The relative contribution of these consequences to the evolution of network regulation is unclear, but simultaneous optimization for ATP yield and catabolic rate under this optimality principle identifies a trade-off between the contradicting objectives of maximum overall ATP yield and maximum rate of ATP formation (Pfeiffer *et al*, 2001). Under nutrient scarcity, in contrast, the metabolic state is closer to an optimal yield of ATP (or biomass) at the cost of the rate of formation.

Materials and methods

Stoichiometric model and constraints

The constructed stoichiometric model of *E. coli* contains all presently known reactions in central carbon metabolism with 98 reactions and 60 metabolites (Supplementary Table I). To apply FBA, the reaction network was automatically translated into a stoichiometric matrix (Schilling and Palsson, 1998) by means of a parser program implemented in Matlab (MATLAB[®], version 7.0.0.1920 (R14), The MathWorks Inc., Natick, MA). Assuming steady-state mass balances, the production and consumption of each of the m intracellular metabolites M_i is balanced to yield

$$Sv = 0 \quad (1)$$

with

$$v_i^{lb} \leq v_i \leq v_i^{ub}$$

S corresponds to the stoichiometric matrix ($m \times n$) and v ($n \times 1$) to the array of n metabolic fluxes with v_i^{lb} as lower and v_i^{ub} as upper bounds, respectively. The above equations represent the conservation law of mass that is fundamental to constraint-based modeling. For all herein presented stoichiometric analyses, maximization of biomass yield is synonymous to the frequently used maximization of growth rate objective (Price *et al*, 2004). This is because stoichiometric models are sets of linear balance equations that are inherently dimensionless, hence maximization of the biomass reaction optimizes the amount of product (i.e., the yield) rather than a time-dependent rate of formation. The P-to-O ratio constraint was implemented by omitting the energy-coupling NADH dehydrogenase I (Nuo), cytochrome oxidase *bo3* (Cyo) and/or cytochrome oxidase *bd* (Cyd) components of the respiratory chain. For a ratio of unity, Cyd and Nuo were set equal to zero. Under anaerobic conditions, electron flow is only possible via the NADH oxidases Nuo or NADH dehydrogenase II (Ndh) to fumarate reductase (Frd), hence coupled to succinate fermentation. For nitrate respiration, the terminal oxidase nitrate reductase (Nar) was used instead of Cyd or Cyo (Unden and Bongaerts, 1997).

For the genome-scale analysis we used two recently reconstructed models of *E. coli* metabolism (Edwards and Palsson, 2000b; Reed *et al*, 2003). *In silico* growth was simulated on glucose minimal medium for all six environmental conditions. ADP remained unbalanced, since otherwise formation of adenosine would be carbon-limited. For the proton-balanced model of Reed *et al* (2003), severe alternate optima occurred in central carbon metabolism given an unlimited proton exchange flux between the cell and the medium and a P-to-O ratio of 2, that is the upper bound of the biologically feasible range of P-to-O ratios (Unden and Bongaerts, 1997). To prevent the unlimited production of ATP equivalents through the ATPS4r reaction under this condition, all external protons involved in the respiratory chain and the transhydrogenase reaction were balanced (specifically, we balanced the external protons around the reactions ATPS4r, TDH2, CYTBD, CYTBO3, NO3R1, NO3R2, NADH6, NADH7, NADH8). A P-to-O ratio of 2 was implemented by assuming both the transport of four protons through CYTBO3 and NADH6 across the membrane and the diffusion of four protons through ATPS4r for the formation of one ATP equivalent.

Objective functions

Linear optimization was used to identify optimal solutions for the objectives maximization of biomass or ATP yield, minimization of glucose consumption, minimization of the redox potential and minimization as well as maximization of ATP producing fluxes. The mathematical definition for all 11 objective functions is given in Table III. While identification of a global optimal value is guaranteed, alternate optima occur frequently. Nonlinear optimization such as the minimization of the overall intracellular flux and maximization of biomass or ATP yield per flux unit do not produce alternate optima. Minimization of the overall intracellular flux always identifies a global optimum because the underlying optimization problem is quadratic and thus convex. Since such convexity cannot be assumed for maximization of biomass or ATP yield per flux unit, we used the general nonlinear solver of the programming package LINDO (Lindo Systems Inc., Chicago, IL) with 100 randomly chosen starting values to find global optima for these two non-convex nonlinear optimization problems. We implemented two independent approaches to validate our results. In a first approach we randomly changed the value of 5% of the variables by 10% iteratively 100 times for all constraint permutations (data not shown). Flux distributions with a higher objective function value were not identified in any of the iterations. In a second approach we reformulated both non-convex nonlinear objective functions

$$\frac{v_{ATP}}{\sum_{i=1}^n v_i^2}$$

and

$$\frac{v_{biomass}}{\sum_{i=1}^n v_i^2}$$

to a nonlinear, but convex form

$$\frac{v_{ATP}}{v_{glucose}} - \varepsilon \sum_{i=1}^n v_i^2 \quad (2a)$$

$$\frac{v_{biomass}}{v_{glucose}} - \varepsilon \sum_{i=1}^n v_i^2 \quad (2b)$$

respectively, corresponding to maximization of (2a) ATP and (2b) biomass yield per flux unit, respectively (Table III contains the mathematical definitions of all objective functions). ε represents a small value that characterizes the unique trade-off between ATP (biomass) maximization and minimization of the flux norm. We assessed the objective function value of ATP (biomass) yield maximization per flux unit for different ε values according to $\varepsilon = \varepsilon_0 (1 \pm 0.5)$, where ε_0 was set equal to the objective function value, which was identified previously with the non-convex nonlinear objective function of max ATP (biomass) yield per flux unit, for the particular environmental condition (Supplementary Figures 7 and 8). Given $\varepsilon = \varepsilon_0$, the previously found flux distribution yielded the optimal solution for every environmental condition (Supplementary Figures 7 and 8). For all other ε values, independent optimizations only lead to suboptimal solutions. Hence, in the present case we have strong indication that global optima are actually identified.

For minimization of the number of reaction steps and for the maximization of the ATP yield per reaction step, we used the mixed-integer solver of the programming package LINDO (Lindo Systems Inc.). All mixed-integer optimizations were formalized as:

$$\min / \max f \quad (3a)$$

$$\text{s.t. } Sv = 0 \quad (3b)$$

$$v_i^{lb} \leq v_i \leq v_i^{ub} \quad (3c)$$

$$v_i - y_i(v_i^{ub} - w_i^{ub}) \leq w_i^{ub} \quad (3d)$$

$$v_i - y_i(v_i^{lb} - w_i^{lb}) \geq w_i^{lb} \quad (3e)$$

$$y_i \in \{0, 1\} \quad (3f)$$

$$w_i^{ub} = w_i + \kappa|w_i| + \varepsilon \quad (3g)$$

$$w_i^{lb} = w_i - \kappa|w_i| - \varepsilon \quad (3h)$$

where for each flux i , $y_i=1$ stands for a non-zero, that is active flux in v_i and $y_i=0$ otherwise, and w_i^{ub} and w_i^{lb} are thresholds for determining non-zero fluxes (equations 3d–f), with κ and ε specifying the relative and absolute ranges of tolerance, respectively (equations 3g and h). The definitions of the objective functions f of linear minimization of reaction steps and nonlinear maximization of ATP yield per reaction step (equation 3a) can be taken from Table III. The constraints of the original linear programming problem with respect to steady state of mass balances and enzyme reversibilities were maintained (equations 3b and c). For κ and ε , we chose minimal values that still resulted in reasonable running times of the mixed-integer solver (specifically, we chose $\kappa=0$, $\varepsilon=0$ and $\kappa=0$, $\varepsilon=1$ for linear minimization of reaction steps and nonlinear maximization of ATP yield per reaction step, respectively). Optimality of the solution obtained by the mixed-integer nonlinear optimization was verified by randomly changing 5% of the integer values 10 times iteratively for the six conditions without additional constraints (data not shown).

Calculability analysis

Instead of comparing all computationally (v_i) and experimentally identified fluxes (v_i^{exp}) in the network, we focused on those that are sufficient to describe the complete systemic degree of freedom because most fluxes are linearly dependent. This minimal subset of fluxes was identified by calculability analysis (Van der Heijden *et al*, 1994; Klamt *et al*, 2002) from the null space of the stoichiometric matrix S and allowed the calculation of all unique reaction rates in the under-determined network. To reduce the considerable difference in magnitude of different fluxes in the network, their rates were expressed as split ratios R of divergent fluxes (Figure 1 and Table I), hence they are scaled to values between zero and unity. Error propagation was used to take the standard deviation of each of the experimentally determined split ratios into account. A default error of 5% was assumed for inactive flux ratios. Secretion of succinate, pyruvate and formate was not considered for calculability analysis, since the corresponding rates are negligible. Non-carbon fluxes such as respiration were also neglected.

Predictive fidelity and alternate optima

Generally, there are two basic principles to quantify the agreement between data series; that is, correlation coefficients that measure linear dependencies and the geometric distance (McShane *et al*, 2002; Segre *et al*, 2002). Since we were interested in the similarity between multiple computational and experimental results rather than their linear dependency, we used the Euclidean distance to quantify the overall agreement. The Euclidean distance belongs to the group of L_2 distance measures, which capture the deviation between two points in absolute terms (Diggle, 1983)

$$\|x\|_2 = \left(\sum_{i=1}^n x_i^2 \right)^{1/2} \quad (4)$$

Our reduction of the overall solution space to a minimal set of 10 linear independent split ratios allowed the direct comparison of complete experimental and computational flux distributions. We defined the term predictive fidelity as the overall agreement between complete experimental and computational flux solutions relative to the specific experimental variance. This explicitly includes putative variability of *in silico* split ratios due to existence of alternate optima. The global optimal value $Z^{(b)}$ of different objective functions f (e.g., maximization of biomass or ATP yield) is determined in a preliminary optimization step, which defines the computational solution space. In case of linear objective functions, with potential alternate optima, the best and the worst possible agreement of the underlying range of flux vectors and

experimental data is subsequently quantified by minimizing and maximizing the standardized Euclidean distance, D^S , respectively:

$$\max / \min D^S \quad (5a)$$

$$\text{s.t. } f^T v = Z^{\text{obj}} \quad (5b)$$

$$Sv = 0 \quad (5c)$$

$$v_j^{\text{lb}} \leq v_j \leq v_j^{\text{ub}} \quad (5d)$$

$$D^S = \varepsilon^T W \varepsilon \quad (5e)$$

$$\varepsilon = R_i^{\text{comp}} - R_i^{\text{exp}} \quad (5f)$$

$$W_{i,i} = \frac{1}{\sigma_i^{\text{exp}}} \left(\sum_i \frac{1}{\sigma_i^{\text{exp}}} \right)^{-1} \quad (5g)$$

$$R_i = f_i(v) \quad (5h)$$

The global optimal value Z^{obj} has to hold (equation 5b) as well as the constraints of the original linear programming problem (equations 1, 5c and d). D^S marks the standardized Euclidean distance (equation 5e), where the deviation ε between *in silico* and *in vivo* ratios, R^{comp} and R^{exp} , respectively (equation 5f), is weighted by the experimental variance σ^{exp} (equation 5g). Finally, the split ratios R are a function of intracellular fluxes v (equation 5h). Note that both computational and experimental split ratios determine the Euclidean distance, such that even small changes in the *in vivo* and *in silico* results can result in a completely different behavior.

Hence, the predictive fidelity explicitly considers both alternate optima and unique solutions. All optimizations for the calculation of the predictive fidelity were performed with linprog and fmincon (MATLAB[®], version 7.0.0.19920 (R14)). Iterative calculations with 100 different, randomly chosen starting points were performed when the nonlinear solver fmincon was used.

Specific agreement for individual split ratios

Predictive fidelity ranks the overall agreement between FBA flux predictions and experimental data. For detailed analysis of the predictive agreement in individual split ratios, experimental confidence intervals and theoretical flexibility due to alternate optima had to be taken into account. The absolute distance between each experimental R^{exp} and the mean computational split ratio $R^{\text{comp,mean}}$ was weighted by the experimental standard deviation δ^{exp} and the possible range ΔR of the computational split ratios (Supplementary Figure 4). The specific agreement ρ was quantified by a standardized variable:

$$\rho_i = |R_i^{\text{exp}} - R_i^{\text{comp,mean}}| \left(\frac{\Delta R_i^{\text{comp}}}{\delta_i^{\text{exp}}} \right) \quad (6)$$

By weighting the absolute distance with the experimental uncertainty, the predictive accuracy is taken into account, that is a large absolute deviation is considered less severe if it is associated with a large experimental uncertainty δ^{exp} . On the other hand, the absolute deviation is considered more severe if it is associated with a large computational uncertainty ΔR . A default value of 0.05 was chosen for δ^{exp} or ΔR , respectively, when the values were zero.

The hierarchical cluster trees were created with the linkage algorithm (MATLAB[®], version 7.0.0.19920 (R14)) using the Euclidean distances among all data points. The cophenetic correlation coefficients (i.e., the correlation coefficients of the distance values) (Legendre and Legendre, 1998) were calculated for all cluster trees to guarantee a faithful representation of the dissimilarities among the 10 ratios for every objective/constraint combination considered. Groups of nodes were assigned where the linkage among the nodes was less than 0.7, when the linkage was normalized to values between 0 and 1.

Supplementary information

Supplementary information is available at the Molecular Systems Biology website (www.nature.com/msb).

Acknowledgements

We are grateful to Annik Nanchen, Eliane Fischer, Tobias Fuhrer, Nicola Zamboni, Matthias Heinemann and Joerg Stelling for fruitful discussions and critical comments on the manuscript. Support for Robert Schuetz through an ETH research grant is acknowledged.

References

- Aebersold R (2005) Molecular systems biology: a new journal for a new biology? *Mol Syst Biol* **1**:5
- Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. *Nature* **422**: 198–207
- Almaas E, Kovacs B, Vicsek T, Oltavi ZN, Barabasi AL (2004) Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature* **427**: 839–843
- Bailey JE (2001) Complex biology with no parameters. *Nat Biotechnol* **19**: 503–504
- Beard DA, Liang SD, Qian H (2002) Energy balance for analysis of complex metabolic networks. *Biophys J* **83**: 79–86
- Bilu Y, Shlomi T, Barkari N, Ruppin E (2006) Conservation of expression and sequence of metabolic genes is reflected by activity across metabolic states. *PLoS Comput Biol* **2**: e106
- Blank LM, Kuepfer L, Sauer U (2005a) Large-scale ¹³C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome Biol* **6**: R49
- Blank LM, Lehmbeck F, Sauer U (2005b) Metabolic-flux and network analysis in 14 hemiascomycetous yeasts. *FEMS Yeast Res* **5**: 545–558
- Blank LM, Sauer U (2004) TCA cycle activity in *Saccharomyces cerevisiae* is a function of the environmentally determined specific growth and glucose uptake rate. *Microbiology* **150**: 1085–1093
- Bonarius HPJ, Hatzimanikatis V, Meesters KPH, de Gooijer CD, Schmid G, Tramper J (1996) Metabolic flux analysis of hybridoma cells in different culture media using mass balances. *Biotechnol Bioeng* **50**: 299–318
- Bork P, Serrano L (2005) Towards cellular systems in 4D. *Cell* **121**: 507–509
- Burgard AP, Maranas CD (2003) Optimization-based framework for inferring and testing hypothesized metabolic objective functions. *Biotechnol Bioeng* **82**: 670–677
- Calhoun MW, Oden KL, Gennis RB, de Mattos MJ, Neijssel OM (1993) Energetic efficiency of *Escherichia coli*: effects of mutations in components of the aerobic respiratory chain. *J Bacteriol* **175**: 3020–3025
- Cassman M (2005) Barriers to progress in systems biology. *Nature* **438**: 1079
- Dauner M, Sauer U (2001) Stoichiometric growth model for riboflavin-producing *Bacillus subtilis*. *Biotechnol Bioeng* **76**: 132–143
- del Valle AE, Aledo JC (2002) Glycolysis in wonderland: the importance of energy dissipation in metabolic pathways. *J Chem Educ* **79**: 1336
- Diggle P (1983) *Statistical Analysis of Spatial Point patterns*. Orlando: Academic Press, pp 16–18
- Ebenhoh O, Heinrich R (2001) Evolutionary optimization of metabolic pathways. Theoretical reconstruction of the stoichiometry of ATP and NADH producing systems. *Bull Math Biol* **63**: 21–55
- Edwards JS, Ibarra RU, Palsson BO (2001) *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnol* **19**: 125–130
- Edwards JS, Palsson BO (2000a) Metabolic flux balance analysis and the *in silico* analysis of *Escherichia coli* K-12 gene deletions. *BMC Bioinformatics* **1**: 1

- Edwards JS, Palsson BO (2000b) The *Escherichia coli* MG1655 *in silico* metabolic phenotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci USA* **97**: 5528–5533
- Emmerling M, Dauner M, Ponti A, Fiaux J, Hochuli M, Szyperski T, Wüthrich K, Bailey JE, Sauer U (2002) Metabolic flux responses to pyruvate kinase knockout in *Escherichia coli*. *J Bacteriol* **184**: 152–164
- Fell DA, Small JR (1986) Fat synthesis in adipose tissue. An examination of stoichiometric constraints. *Biochem J* **238**: 781–786
- Fischer E, Sauer U (2003) A novel metabolic cycle catalyzes glucose oxidation and anaplerosis in hungry *Escherichia coli*. *J Biol Chem* **278**: 46446–46451
- Fischer E, Sauer U (2005) Large-scale *in vivo* flux analysis shows rigidity and suboptimal performance of *Bacillus subtilis* metabolism. *Nat Genet* **37**: 636–640
- Fischer E, Zamboni N, Sauer U (2004) High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived ¹³C constraints. *Anal Biochem* **325**: 308–316
- Flores S, Gosset G, Flores N, de Graaf AA, Bolivar F (2002) Analysis of carbon metabolism in *Escherichia coli* strains with an inactive phosphotransferase system by ¹³C labeling and NMR spectroscopy. *Metab Eng* **4**: 124–137
- Forster J, Famili I, Palsson BO, Nielsen J (2003) Large-scale evaluation of *in silico* gene deletions in *Saccharomyces cerevisiae*. *Omic* **7**: 193–202
- Gennis RB, Stewart V (1996) Respiration. In *Escherichia coli and Salmonella. Cellular and Molecular Biology*, Neidhardt FC, Curtiss III R, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umberger HE (eds), pp 217–222. Washington DC: ASM Press
- Heinrich R, Montero F, Klipp E, Waddell TG, Melendez-Hevia E (1997) Theoretical approaches to the evolutionary optimization of glycolysis: thermodynamic and kinetic constraints. *Eur J Biochem* **243**: 191–201
- Hellerstein MK (2003) *In vivo* measurement of fluxes through metabolic pathways: the missing link in functional genomics and pharmaceutical research. *Annu Rev Nutr* **23**: 379–402
- Holzhütter HG (2004) The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *Eur J Biochem* **271**: 2905–2922
- Hoskisson PA, Hobbs G (2005) Continuous culture – making a comeback? *Microbiology* **151**: 3153–3159
- Ibarra RU, Edwards JS, Palsson BO (2002) *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature* **420**: 186–189
- Kacser H, Beeby R (1984) Evolution of catalytic proteins or on the origin of enzyme species by means of natural selection. *J Mol Evol* **20**: 38–51
- Kholodenko BN (2006) Cell-signalling dynamics in time and space. *Nat Rev Mol Cell Biol* **7**: 165–176
- Kholodenko BN, Kiyatkin A, Bruggeman FJ, Sontag E, Westerhoff HV, Hoek JB (2002) Untangling the wires: a strategy to trace functional interactions in signaling and gene networks. *Proc Natl Acad Sci USA* **99**: 12841–12846
- Kitano H (2002) Systems biology: a brief overview. *Science* **295**: 1662–1664
- Klamt S, Schuster S, Gilles ED (2002) Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism in purple nonsulfur bacteria. *Biotechnol Bioeng* **77**: 734–751
- Knorr AL, Jain R, Srivastava R (2007) Bayesian-based selection of metabolic objective functions. *Bioinformatics* **23**: 351–357
- Kuepfer L, Sauer U, Blank LM (2005) Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Res* **15**: 1421–1430
- Lee S, Phalakornkule C, Domach MM, Grossmann IE (2000) Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. *Comput Chem Eng* **24**: 711–716
- Legendre P, Legendre L (1998) *Numerical Ecology*. 2nd edn, Amsterdam: Elsevier Science BV
- Mahadevan R, Schilling CH (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng* **5**: 264–276
- McShane LM, Radmacher MD, Freidlin B, Yu R, Li MC, Simon R (2002) Methods for assessing reproducibility of clustering patterns observed in analyses of microarray data. *Bioinformatics* **18**: 1462–1469
- Melendez-Hevia E, Isidoro A (1985) The game of the pentose phosphate cycle. *J Theor Biol* **117**: 251–263
- Moreira dos Santos M, Gombert AK, Christensen B, Olsson L, Nielsen J (2003) Identification of *in vivo* enzyme activities in the cometabolism of glucose and acetate by *Saccharomyces cerevisiae* by using ¹³C-labeled substrates. *Eukar Cell* **2**: 599–608
- Nanchen A, Schicker A, Sauer U (2006) Nonlinear dependency of intracellular fluxes on growth rate in miniaturized continuous cultures of *Escherichia coli*. *Appl Environ Microbiol* **72**: 1164–1172
- Oliveira AP, Nielsen J, Forster J (2005) Modeling *Lactococcus lactis* using a genome-scale flux model. *BMC Microbiol* **5**: 39
- Papin JA, Price ND, Edwards JS, Palsson BB (2002) The genome-scale metabolic extreme pathway structure in *Haemophilus influenzae* shows significant network redundancy. *J Theor Biol* **215**: 67–82
- Papp B, Pal C, Hurst LD (2004) Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* **429**: 661–664
- Perrenoud A, Sauer U (2005) Impact of global transcriptional regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on glucose catabolism in *Escherichia coli*. *J Bacteriol* **187**: 3171–3179
- Pfeiffer T, Schuster S (2005) Game-theoretical approaches to studying the evolution of biochemical systems. *Trends Biochem Sci* **30**: 20–25
- Pfeiffer T, Schuster S, Bonhoeffer S (2001) Cooperation and competition in the evolution of ATP-producing pathways. *Science* **292**: 504–507
- Price ND, Reed JL, Palsson BO (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Rev Microbiol* **2**: 886–897
- Ramakrishna R, Edwards JS, McCulloch A, Palsson BO (2001) Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints. *Am J Physiol Regul Integr Comp Physiol* **280**: R695–R704
- Reed JL, Famili I, Thiele I, Palsson BO (2006) Towards multi-dimensional genome annotation. *Nat Rev Genet* **7**: 130–141
- Reed JL, Palsson BO (2004) Genome-scale *in silico* models of *E. coli* have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states. *Genome Res* **14**: 1797–1805
- Reed JL, Vo TD, Schilling CH, Palsson BO (2003) An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol* **4**: R54
- Russell JB, Cook GM (1995) Energetics of bacterial growth: balance of anabolic and catabolic reactions. *Microbiol Rev* **59**: 48–62
- Sauer U (2004) High-throughput phenomics: experimental methods for mapping fluxomes. *Curr Opin Biotechnol* **15**: 58–63
- Sauer U (2006) Metabolic networks in motion: ¹³C-based flux analysis. *Mol Syst Biol* **2**: 62
- Sauer U, Canonaco F, Heri S, Perrenoud A, Fischer E (2004) The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of *Escherichia coli*. *J Biol Chem* **279**: 6613–6619
- Schilling CH, Palsson BO (1998) The underlying pathway structure of biochemical reaction networks. *Proc Natl Acad Sci USA* **95**: 4193–4198
- Segre D, Vitkup D, Church GM (2002) Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci USA* **99**: 15112–15117
- Shlomi T, Berkman O, Ruppin E (2005) Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc Natl Acad Sci USA* **102**: 7695–7700
- Sopko R, Huang D, Preston N, Chua G, Papp B, Kafadar K, Snyder M, Oliver SG, Cyert M, Hughes TR, Boone C, Andrews B (2006)

- Mapping pathways and phenotypes by systematic gene overexpression. *Mol Cell* **21**: 319–330
- Stelling J (2004) Mathematical models in microbial systems biology. *Curr Opin Microbiol* **7**: 513–518
- Thiele I, Price ND, Vo TD, Palsson BO (2005) Candidate metabolic network states in human mitochondria. *J Biol Chem* **280**: 11683–11695
- Uden G, Bongaerts J (1997) Alternative respiratory pathways of *Escherichia coli*: energetics and transcriptional regulation in response to electron acceptors. *Biochim Biophys Acta* **1320**: 217–234
- Van der Heijden RTJM, Heijnen JJ, Hellinga C, Romein B, Luyben KChAM (1994) Linear constraint relations in biochemical reaction systems: I. Classification of the calculability and the balanceability of conversion rates. *Biotechnol Bioeng* **43**: 3–10
- van Gulik WM, Heijnen JJ (1995) A metabolic network stoichiometry analysis of microbial growth and product formation. *Biotechnol Bioeng* **48**: 681–698
- Varma A, Boesch BW, Palsson BO (1993) Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. *Appl Environ Microbiol* **59**: 2465–2473
- Varma A, Palsson BO (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl Environ Microbiol* **60**: 3724–3731
- Warner GJ, Adeleye YA, Ideker T (2006) Interactome networks: the state of the science. *Genome Biol* **7**: 3001
- Wiback SJ, Mahadevan R, Palsson BO (2004) Using metabolic flux data to further constrain the metabolic solution space and predict internal flux pattern: the *Escherichia coli* spectrum. *Biotechnol Bioeng* **86**: 317–331
- Wiechert W, Nöh K (2005) From stationary to instationary metabolic flux analysis. *Adv Biochem Eng Biotechnol* **92**: 145–172
- Xu B, Jahic M, Enfors S-O (1999) Modeling of overflow metabolism in batch and fed-batch cultures of *Escherichia coli*. *Biotechnol Prog* **15**: 81–90
- Zhao J, Shimizu K (2003) Metabolic flux analysis of *Escherichia coli* K12 grown on ^{13}C -labeled acetate and glucose using GC-MS and powerful flux calculation method. *J Biotechnol* **101**: 101–117



Molecular Systems Biology is an open-access journal published by *European Molecular Biology Organization* and *Nature Publishing Group*.

This article is licensed under a Creative Commons Attribution License.