Transition of an Anaerobic *Escherichia coli* Culture to Aerobiosis: Balancing mRNA and Protein Levels in a Demand-Directed Dynamic Flux Balance Analysis

Joachim von Wulffen*, RecogNice-Team¶, Oliver Sawodny, Ronny Feuer

Institute for System Dynamics, University of Stuttgart, Stuttgart, Germany

¶ Membership list can be found in the Acknowledgments section.
* vonwulffen@isys.uni-stuttgart.de

Abstract

The facultative anaerobic bacterium *Escherichia coli* is frequently forced to adapt to changing environmental conditions. One important determinant for metabolism is the availability of oxygen allowing a more efficient metabolism. Especially in large scale bioreactors, the distribution of oxygen is inhomogeneous and individual cells encounter frequent changes. This might contribute to observed yield losses during process upscaling. Short-term gene expression data exist of an anaerobic *E. coli* batch culture shifting to aerobic conditions. The data reveal temporary upregulation of genes that are less efficient in terms of energy conservation than the genes predicted by conventional flux balance analyses. In this study, we provide evidence for a positive correlation between metabolic fluxes and gene expression. We then hypothesize that the more efficient enzymes are limited by their low expression, restricting flux through their reactions. We define a demand that triggers expression of the demanded enzymes that we explicitly include in our model. With these features we propose a method, demand-directed dynamic flux balance analysis, dddFBA, bringing together elements of several previously published methods. The introduction of additional flux constraints proportional to gene expression provoke a temporary demand for less efficient enzymes, which is in agreement with the transient upregulation of these genes observed in the data. In the proposed approach, the applied objective function of growth rate maximization together with the introduced constraints triggers expression of metabolically less efficient genes. This finding is one possible explanation for the yield losses observed in large scale bacterial cultivations where steady oxygen supply cannot be warranted.

Introduction

The model organism *Escherichia coli* is a facultative anaerobic bacterium, i.e. it is able to grow in both aerobic and anaerobic environments. To do so, cells need to be able to adapt to changes

of the growth conditions. This capability is required in both the natural habitat and in biotechnological applications where, due to inefficient mixing in large scale bioreactors, oxygen supply is unsteady [1–3]. Adaptation takes place on the transcriptional level in multiple ways, e.g. increased expression of genes for aerobic respiration, or decreased expression of less efficient fermentative genes. Due to the high reduction potential of molecular oxygen, cells are able to generate more energy from its substrates, e.g. sugars, in aerobic compared to anaerobic metabolism. Survival is hampered, however, by the toxic effects of oxygen. Oxygen can get reduced to form the superoxide anion (O_2^-) and other ROS which provoke oxidative damage to DNA, RNA, proteins, and lipids (reviewed in [4]).

Adaptation from anaerobic to aerobic metabolism requires extensive adjustments of the enzyme composition of the cell (more than 20% of the genome being differentially expressed [5]). In this article we will focus on the adjustment of metabolic enzyme expression. Cells are constantly competing against each other for resources and so the fastest growing and best adapted cells will prevail. On the other hand, gene expression is a costly and time-consuming process so cells need to evolve a way of minimizing the total adjustments required for adaptation [6]. Transient overexpression of enzymes, especially enzymes of the first steps of a pathway, allows maximal flux through the pathway with concurrent minimal adjustments of expression [7]. Transition data of transcript expression of a shift from anaerobic to aerobic conditions in an *E. coli* culture also indicate transient overexpression of multiple genes [5]. Furthermore, these data suggest expression of metabolically less efficient genes.

Lewis *et al.* [8] used flux balance analysis (FBA) to classify genes with respect to their essentiality and relative efficiency. In FBA, fluxes of the quasi-steady state metabolism of a culture are estimated by an optimization procedure [9]. FBA has been successful in predicting phenotypes of knock-out mutant strains, metabolite exchange rates and growth rates under different environmental conditions and in different organisms [10–13]. The flux distribution is modeled assuming a quasi-steady state of internal metabolites that are interconnected via reactions that are stored in the stoichiometric matrix [11, 12, 14]. The solution space is further constrained thermodynamically by forcing irreversible fluxes to be positive. Concerning the objective function of the optimization there are different approaches according to prerequisites and the addressed problem [13]. A frequently and successfully applied objective function for batch cultivations is maximization of the growth rate with concurrent restriction of substrate uptake (reviewed in [15]). The result of an FBA is then a growth optimal flux distribution of a quasi-steady state culture in the exponential growth phase.

In parsimonious FBA (pFBA) [8], this optimization is followed by a minimization of all fluxes within the solution space of the the growth optimum, to minimize the requirement for enzyme expression. Using this technique, genes can be classified according to whether it is essential, required for optimal growth, metabolically less efficient (MLE, i.e. less energy production, or more consumption when this gene is used), enzymatically less efficient (ELE, i.e. overall flux is higher when this gene is used), or whether the gene is not used in any case.

Transient upregulation of MLE genes might arise from limited capacity of the optimal enzymes. E.g., in the electron transport chain (ETC) the MLE gene *ndh* (catalyzing the reaction NADH5, reaction naming according to iJO1366 [14]) is transiently upregulated whereas expression of the optimal enzyme, encoded by the *nuo* operon (catalyzing the reaction NADH16pp), increases only slightly [5]. NADH5 is MLE, because it translocates no protons across the plasma membrane which could otherwise be used for energy conservation via ATPase (Fig 1). Similarly, expression of the *cyo* operon (catalyzing CYTBO3_4pp) increases after transition, whereas the MLE operon *cyd* (catalyzing CYTBDpp, Fig 1) is transiently overexpressed. Temporary overexpression is useful to minimize the required adjustments of gene expression [7], yet, this was only shown for essential pathways and not for MLE genes. If a
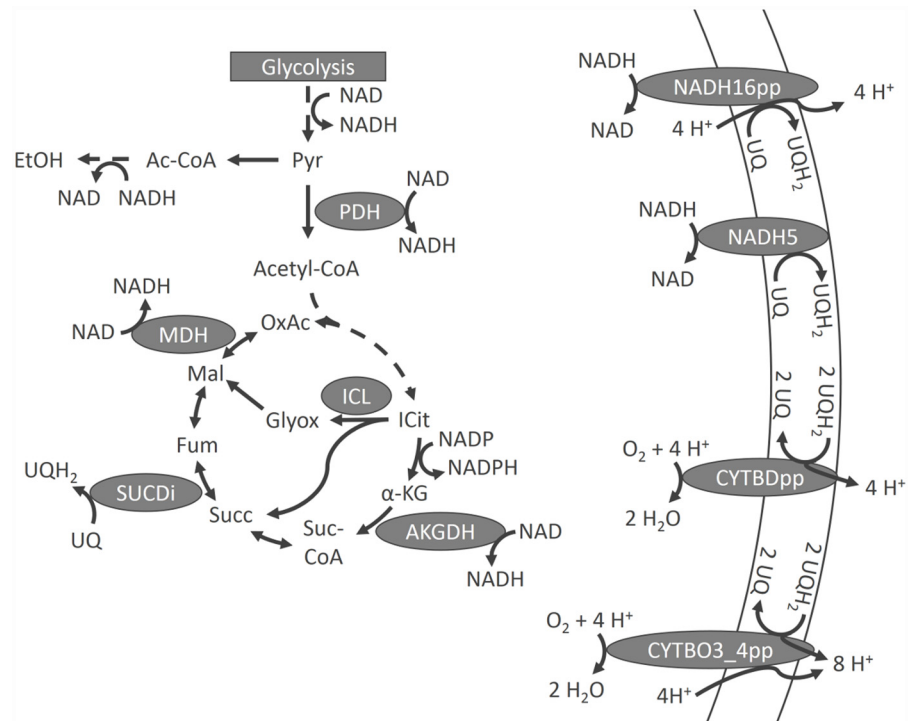
Fig 1. Outline of central aerobic metabolism. Reaction nomenclature according to iJO1366 [14]. Pyr, pyruvate; ICit, Isocitric acid; *α*-KG, *α*-ketoglutaric acid; Suc-CoA, succinyl-coenzyme A; Succ, succinic acid; Mal, malic acid; OxAc, oxaloacetic acid; UQ, ubiquinone; UQH$_2$, ubiquinol; only redox cofactors are considered.

transiently high flux through an optimal enzyme is beneficial for efficient growth, but the enzyme capacity is constrained due to underexpression, MLE genes might help overcome this shortage at the expense of optimal efficiency.

Dynamic FBA (dFBA) is an extension of FBA with the aim of simulating time course experiments [16]. In dFBA with static optimization approach (SOA) the simulation time is divided into small periods which are assumed to be in quasi-steady state. For every time step, an FBA problem is solved and the fluxes are integrated over the time period and extracellular concentrations are calculated, accordingly. dFBA was originally applied to simulations of diauxic growth and has been extended, e.g. with regulatory constraints [17–20].

These studies have in common that they are passive in terms of their regulation. A Boolean regulatory model is superimposed on the reaction network by addition of flux constraints that are deduced from the environmental state of the cell. Boolean models of gene regulation generally have the problem that the interplay between the factors is highly complex, therefore it is hard to identify and model the contribution of individual factors satisfactorily.

Several methods have been published that aim to improve flux predictions based on transcriptomic data. However, in an extensive comparison, none of the methods improved the overall outcome (determined by deviation from fluxes measured using ^{13}C-labeled substrates) compared to pFBA which does not incorporate transcriptomic data [21]. This indicates that for most cases the constraints introduced prevent the FBA from correctly identifying the parsimonious optimum which better approximates the true flux distribution than the other methods. We therefore decided not to use transcription data directly to constrain the metabolic model (as e.g. in [22–25]), but to simulate mRNA and protein expression explicitly and then

deduce constraints depending on these simulated values. Since flux measurement data are only available for steady states, they cannot be used as a benchmark here. We therefore investigate whether the principle behavior of gene expression matches our simulations with the enzymatic capacity constraints and objective functions suggested.

Results and Discussion

Correlation of transcriptome data with FBA fluxes

To investigate whether gene expression can serve as an indicator for respective flux rates, we analyzed their correlation. For this, we needed to define thresholds for differential flux rates and differential gene expression. We applied three thresholds for differential gene expression: a minimal fold change between the conditions of 1.25, a minimal mean logCPM value of 2, and a maximum false discovery rate (FDR) corrected P-value of 0.05. In accordance with Zelezniak *et al.* [26] for enzymes consisting of multiple subunits, we used the fold change value for the subunit with the least fold change and for multiple isoenzymes, catalyzing the same reaction, we used the mean of the isoenzymes' fold changes. Positive correlations between mRNA and protein abundances have been shown in several studies, e.g. [26–28], giving rise to a positive relation of mRNA and maximal enzymatic turnover rate.

To identify reactions with differential flux rates, we first combined fluxes that use the same set of enzymes (same isoenzymes and subunits, according to gene-protein-reaction associations [14, 29]) by applying the sum of absolute flux rates. In this way, reactions that differ only in cofactor usage are lumped together reducing variability of the individual fluxes. We applied two criteria on the flux sums to identify differential fluxes: The threshold for the minimal difference between the respective condition and the anaerobic control was set to $0.25 \frac{mmol}{h \cdot gDCW}$. The second criterion for differential flux rate was non-overlapping intervals in flux variability analysis (FVA).

Fig 2 shows the correlation of flux rate and transcriptome difference with a linear regression. The aerobic steady state expression data correlate well with FBA flux differences with a Spearman rank coefficient of 0.51 and a P-value of 0.003. The linear regression exerts a positive slope, indicating that the relation is also positive. In the following, we want to explore whether this relation appears between a dynamic FBA simulation and transient expression data, as well.

Method development

As pointed out, to our best knowledge, none of the previously published methods is able to map fast dynamic gene expression data to dynamically changing fluxes. Short-term data of the aerobic shift [5] suggest that fluxes through main aerobic pathways, such as the ETC, are initially constrained by enzyme availability, assuming a correlation between gene expression and fluxes. In order to model the dynamically changing fluxes, we conceived a demand-directed dynamic FBA (dddFBA) that integrates a simulation of dynamic gene expression with enzyme kinetic parameters. The method is based on dFBA with SOA [16], where the fluxes are optimized to present conditions only.

In order to capture gene expression dynamics, balance equations for selected mRNA and proteins are set up.

$$\frac{d\,[mRNA_i]}{d\,t} = s_{b,i} + R_i(t) \cdot s_{a,i} - (\mu + \gamma_i) \cdot [mRNA_i](t) \tag{1}$$

This equation incorporates a basal transcription rate, $s_{b,i}$, reflecting constitutive expression, an activated transcription rate, $s_{a,i}$, that is switched *on* or *off* depending on the regulatory signal $R_i(t) \in \{0,1\}$, dilution with the growth rate μ, and a gene specific degradation rate, γ_i, adopted from Bernstein *et al.* [30].
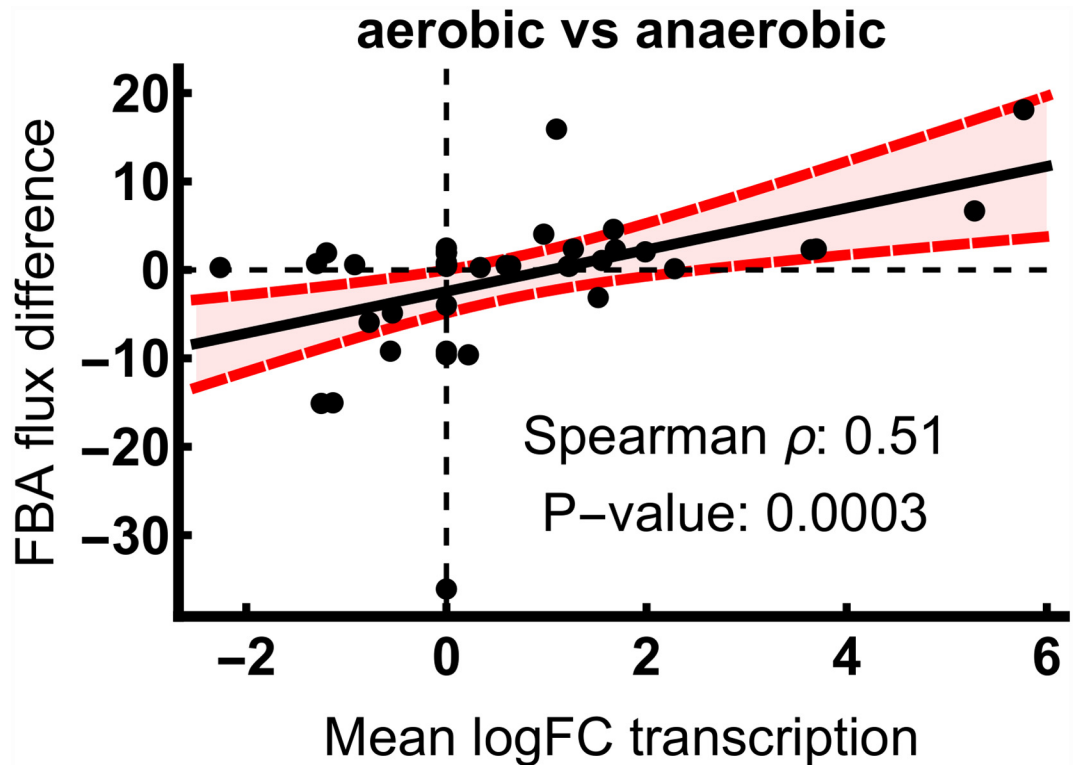
Fig 2. Correlation of transcriptional fold changes with FBA flux rate differences. The difference between the sums of absolute FBA fluxes with identical enzyme composition between anaerobic and aerobic conditions is plotted against the logFC of the associated transcripts. Spearman rank coefficient and P-values are indicated. Linear regression is shown as black line with 95% confidence bands in red.

doi:10.1371/journal.pone.0158711.g002

The value of the regulatory signal determines whether additional transcription is active or not and is directly responsive to the demand for the respective metabolic flux:

$$R_i(t) = \begin{cases} 0 & j_i(t) < \theta \cdot b_i(t) \\ 1 & j_i(t) \geqslant \theta \cdot b_i(t) \end{cases} \tag{2}$$

The regulatory threshold parameter, θ, which decides whether a demand for the flux is present (defined as the proportion of flux $j_i(t)$ to flux bound $b_i(t)$ that triggers regulation) needs to be chosen between 0 and 1. We chose the value 0.6 reflecting that most enzymes are more than half-maximally occupied by their respective substrates (i.e. substrate concentrations are often slightly higher than K_M values, [31]). This choice resulted in good agreement between simulation and measurement of gene expression.

The protein concentration $[P_i]$ is modeled in a second ordinary differential equation (ODE, Eq 3). Degradation of proteins in *E. coli* is negligible on this time scale and so only the dilution term appears in the equation [32].

$$\frac{d[P_i]}{dt} = s_{P,i} \cdot [mRNA_i](t) - \mu \cdot [P_i](t) \tag{3}$$

$$s_{P,i} = \frac{\tau}{\Lambda_i} \tag{4}$$

We assume the parameter for the protein synthesis rate, $s_{P,i}$, to be inversely proportional to the length of the respective gene, Λ_i (identified using gene-protein-reaction associations), since ribosomes have to travel along the distance of the gene with limited velocity to produce a polypeptide. The proportionality constant, τ, cannot be derived from literature values, because mRNA levels are given in the normalized unit RPKM. The value of τ may therefore to be chosen to fit the data. In our case it was taken to be equal to $0.6 \frac{bp \cdot mmol}{h \cdot RPKM \cdot gDCW}$ (Eq 4). Finally, the upper bounds $b_i(t)$ for the respective enzyme fluxes result from

$$b_i(t) = k_{cat,i} \cdot [P_i](t) \qquad (5)$$

The required parameters were either estimated from transitional gene expression data ($s_{b,i}$, $s_{a,i}$), or adopted from the literature (γ_i, $k_{cat,i}$, Table 1).

The method contains elements successfully applied to other methodologies. FBA with molecular crowding (FBAwMC [39]) and metabolic modeling with enzyme kinetics (MOMENT [40]) rely on enzyme turnover numbers and restriction arises from limited space of the cytosol available for proteins. These models successfully predict growth rates in multiple media compositions. Similarly, FBA with membrane economics (FBA-ME) limits the available space for membrane-bound proteins and correctly predicts overflow metabolism [41]. E-FLUX limits upper flux bounds according to enzyme capacity that is directly deduced from expression data [24]. This method requires multiple training data that scale the maximal estimated flux to the range of expression data and deduces relative flux bounds from this scaling. Finally, in regulatory FBA (rFBA [17–19]) a Boolean regulatory network determines whether fluxes are allowed or not. This method describes the successive uptake of different carbon sources.

Opposed to these methods we explicitly take gene transcription and translation into account. We do not use measured gene expression values directly to administer bounds to fluxes. Also, we do not impose a regulatory model to determine gene availability but regarded a high flux in relation to its respective bound as a measure for demand for the corresponding enzyme which triggers expression.

Our method takes on the idea from rFBA of switching genes *on* and *off*, but we apply this to transcription instead of enzyme availability, so that continuous flux bounds are obtained. Furthermore, dddFBA does not need a Boolean regulatory model for gene expression, but relies on the flux demand as the regulatory signal. The size of the protein is a critical parameter determining the readiness of availability of a protein. We considered this by normalization of the translation constant τ with the respective gene length. For scaling, enzyme kinetic parameters are applied which limit only the upper bound. In contrast to FBAwMC, MOMENT, or FBA-ME, enzyme production is intrinsically limited by defined synthesis and degradation parameters as well as dilution. Expressed mRNA and protein can be regarded as accumulating compounds. In this way, dynamic effects, such as transient expression, arise.

Simulation

Simulation of mRNA and protein expression was applied to the reactions and parameters indicated in Table 1. Fig 3 illustrates the short-time upregulation of the MLE reactions CYTBDpp and NADH5 simulated by dddFBA with maximization of growth rate as the objective function. Measured gene expression is qualitatively resembled by the simulation. Expression of MLE genes in the simulation represents a temporary demand for the respective reaction flux.

For FVA, the second optimization of pFBA (minimization of total flux) was omitted and each flux was minimized and maximized keeping the optimal growth rate constant. Flux variability is indicated in Fig 3 by shadowing.

Table 1. Parameters for dddFBA modeling.

reaction	gene	gene length	$s_b \left[\frac{RPKM}{min}\right]$	$s_a \left[\frac{RPKM}{min}\right]$	$\gamma \ [min^{-1}]$[30]	$s_P \left[\frac{nmol}{min\ RPKM\ gDCW}\right]$	$k_{cat} \ [s^{-1}]$	Ref.
CYTBD	cydA	1569	116.3	759.1 ± 311.1	0.187	10.16	11.7	[33]
CYTBO	cyoB	1992	2.4	610.9 ± 88.9	0.210	8.07	341	[34]
NADH5	ndh	1305	5.9	504.6 ± 106.8	0.210	12.10	15.8	[35]
NADH16	nuoG	2727	25.1	30.0 ± 6.9	10.68	5.95	100	[36]
AKGDH	sucA	2802	12.1	143.1 ± 25.1	11.76	4.55	49	[37]
PDH	aceE	2664	48.1	2334.0 ± 373.2	7.5	3.79	21.9	[38]

γ and k_{cat} values are adopted from the indicated literature. s_b and s_P are directly calculated from measurement data and gene lengths, respectively. s_a values are estimated from the measurement data, according to Eq 6. 95% confidence intervals are given.

doi:10.1371/journal.pone.0158711.t001

The reaction CYTBDpp (in the following referred to as *cyd*) is MLE compared to CYTBO3_4pp (in the following referred to as *cyo*) as it translocates less protons across the plasma membrane (Fig 1). However, basal anaerobic expression of *cyo* is very low despite its higher aerobic efficiency. The flux through *cyo* reaches its boundary immediately after aeration and expression is turned on. As long as the *cyo* flux is constrained *cyd* is additionally expressed. Furthermore, *cyd* is expressed higher during anaerobic steady state which is probably due to its higher oxygen affinity enabling it to capture even traces of oxygen [42–44]. When *cyo* reaches its final flux value the *cyd* flux drops to zero and *cyd* expression decreases, accordingly. The fluxes of both *cyo* and *cyd* are not variable indicating that the given flux distributions represent the metabolically and enzymatically most efficient solutions.

NADH5 (referred to as *ndh*) is MLE with respect to NADH16pp (referred to as *nuo*) which translocates protons across the plasma membrane (Fig 1). In contrast to the other MLE gene *cyd*, *ndh* is not transcribed anaerobically. The more efficient *nuo* reaction stays constrained throughout the simulation. Despite its very low anaerobic expression, *ndh* is readily available due to its small size. It consists of a single subunit only, whereas *nuo* is a bulky complex consisting of 13 different subunits. Flux through *ndh* is required for an enzymatically efficient solution throughout the simulation causing an overall efficiency loss.

Variability of *ndh* is high and ranges from zero to the upper flux bound indicating that other solutions exist that are metabolically equally efficient. Alternative pathways that reoxidize NADH are listed in Table 2.

The malate dependent way overlaps with the TCA cycle. MDH transfers electrons from malate to NAD$^+$ yielding oxaloacetate and NADH. However, as *nuo* is limited and unable to efficiently reoxidize all NADH generated, the MLE reaction MDH2 assists. This reaction skips NADH and directly transfers the electrons from malate to ubiquinone. However, under *nuo* constrained conditions, MDH2 becomes enzymatically more efficient as it skips the second reaction. This explains the relatively high flux through MDH2 compared to MDH and the high variability of both reactions (Fig 4a and 4b). Expression of both genes, *mdh* catalyzing MDH and *mqo* catalyzing MDH2 increase after oxygenation indicating the demand for both enzymes (Fig 5a and 5b).

Fluxes through LDH_D and LDH_D2 constitute an ELE alternative pathway for NADH reoxidation (c.f. Table 2). Variation for these pathways is high and the pFBA solution predicts a much lower flux usage because of the increased enzyme demand. Still, a slight overexpression of both corresponding genes could indicate a demand for this functionality (Fig 5c and 5d).

Flux through the succinate dependent pathway is not simulated since both, FRD2 and SUCDi, operate at their respective minima (Fig 4e and 4f). Expression of the characteristic
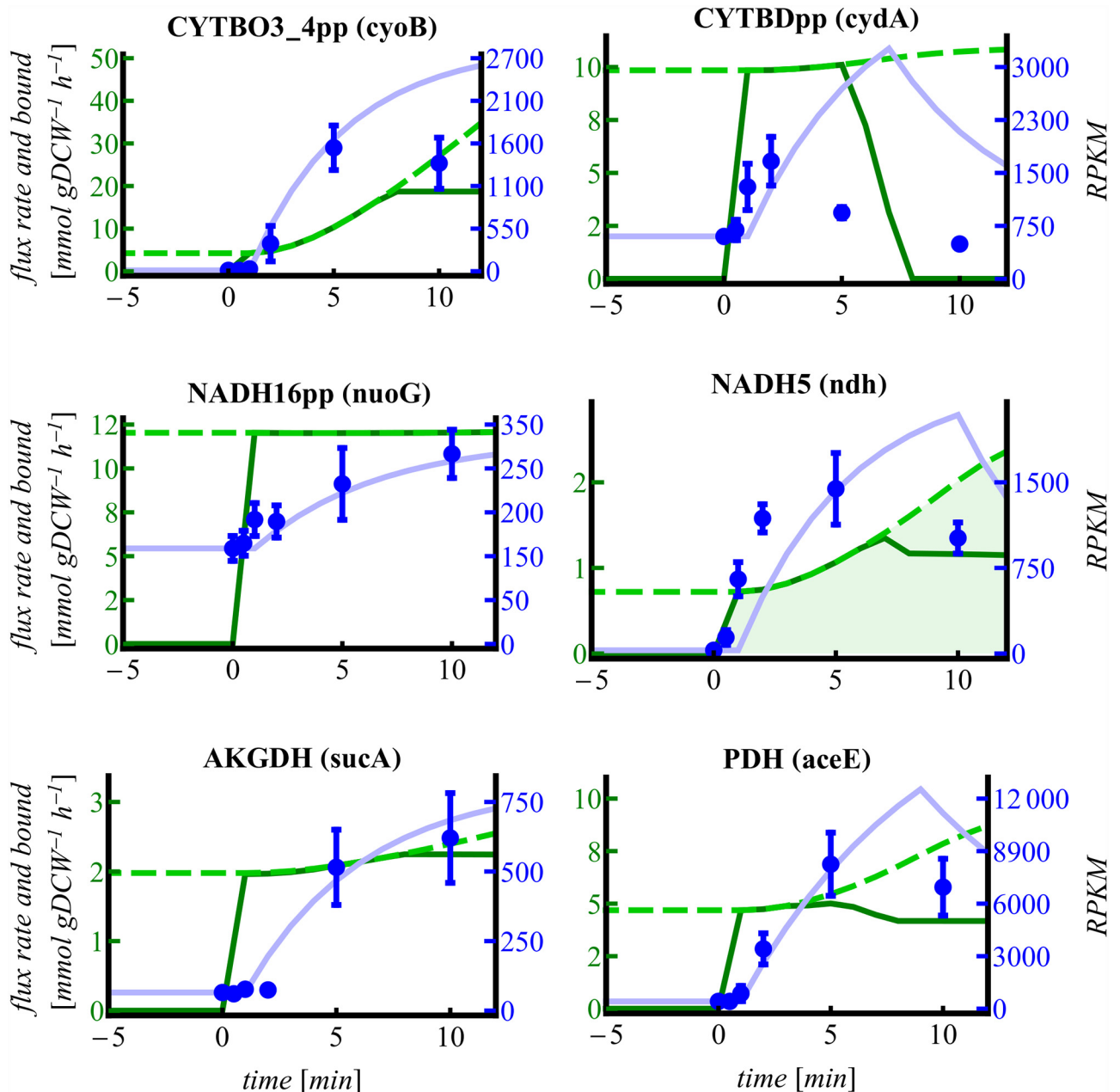
Fig 3. Fluxes of balanced genes in dddFBA. Fluxes are depicted as green lines, upper flux bounds as dashed light green lines, flux variability as shadowed areas and correspond to the left axis; measured mRNA expression as blue dots with standard deviations, and simulated mRNA expression in light blue correspond to the right axis. 0 min denote the onset of aeration.

doi:10.1371/journal.pone.0158711.g003

protein (encoded by the *frd* operon, Fig 5e) decreases, substantiating the estimated absence of this pathway.

Both C-catabolizing reactions AKGDH and PDH are required for aerobic growth (Fig 3). Their flux is initially constrained resulting in activated transcription which is in good qualitative agreement with the measured data. PDH which forms the connector between glycolysis and the TCA cycle is released earlier from the upper bound constraint resulting in elevated flux through the TCA cycle that is still constrained (by AKGDH). This imbalance is resolved by the

Table 2. Alternative pathways for reoxidation of NADH.

reaction name	reaction
malate dependent way	
MDH (reverse)	malate + NAD$^+$ $\rightleftharpoons$ oxaloacetate + NADH + H$^+$
MDH2	malate + UQ $\longrightarrow$ oxaloacetate + UQH$_2$
lactate dependent way	
LDH_D (reverse)	lactate + NAD$^+$ $\rightleftharpoons$ pyruvate + NADH + H$^+$
LDH_D2	lactate + UQ $\longrightarrow$ pyruvate + UQH$_2$
succinate dependent way	
MDH (reverse)	malate + NAD$^+$ $\rightleftharpoons$ oxaloacetate + NADH + H$^+$
MDH3	malate + MQ $\longrightarrow$ oxaloacetate + MQH$_2$
FRD2	fumarate + MQH$_2$ $\longrightarrow$ succinate + MQ
SUCDi	succinate + UQ $\longrightarrow$ fumarate + UQH$_2$

doi:10.1371/journal.pone.0158711.t002

temporary utilization of the glyoxylate shunt reactions ICL and MALS ([Fig 4g and 4h](#)). The glyoxylate shunt omits several energy conserving steps of the TCA and thus can be perceived as MLE. Transient overexpression of both glyoxylate shunt genes, *aceAB* ([Fig 5g and 5h](#)), supports the simulation result.

Parameter sensitivity

For the presented method parameter sensitivity was analyzed qualitatively by parameter variation. Results of variations of the global parameters translation constant τ ([Eq 4](#)) and regulatory threshold θ ([Eq 2](#), [S1 Fig](#)) as well as of the enzyme turnover numbers of CYTBO3_4pp and AKGDH ([S2 Fig](#)) are illustrated.

θ is varied between the columns of [S1 Fig](#) from 0.1 over 0.6 (the value applied in Figs [3](#) and [4](#)) to 0.9. The observed fluxes prove to be insensitive towards θ, only the duration of the expression is influenced θ (e.g. expression of *cyo*, encoding the CYTBO3_4pp enzyme, ceases earlier with increasing θ, [S1a–S1c Fig](#)). τ is varied between the rows of [S1 Fig](#). As this parameter determines the height of the upper bound, it influences both, expressions and fluxes. At the intermediate θ of 0.6 and a low τ of 0.3 $\frac{bp \cdot mmol}{h \cdot RPKM \cdot gDCW}$, flux through the MLE reaction NADH5 is bounded until it reaches its steady-state value after approximately 35 min ([S1b Fig](#)). The initial upper bounds increase linearly with the translation constant and so do the fluxes through the more efficient enzymes. At a τ of 0.6 $\frac{bp \cdot mmol}{h \cdot RPKM \cdot gDCW}$, NADH5 is only bounded until 8 min, whereas no demand for this flux exists any more at a τ of 0.9 $\frac{bp \cdot mmol}{h \cdot RPKM \cdot gDCW}$ ([S1e and S1h Fig](#)) since the more efficient NADH16pp is unrestricted.

Similarly, the turnover numbers of the enzymes influence the respective fluxes. We exemplify this by bisection and doubling of the turnover numbers of CYTBO3_4pp and AKGDH leaving both τ and θ constant at 0.6 ([S2 Fig](#)). Again, upper bounds increase linearly with the turnover number, which subsequently decreases the duration of bounded flux and expression (c.f. AKGDH in [S2a–S2c Fig](#)). The reduced demand also influences related MLE reactions, such as CYTBDpp whose demand decreases with increasing CYTBO3_4pp turnover ([S2b, S2e and S2h Fig](#)).

Conclusions

Based on the objective function of growth maximization and parsimonious enzyme usage, which are both reasonable assumptions from an evolutionary point of view, and constraints of
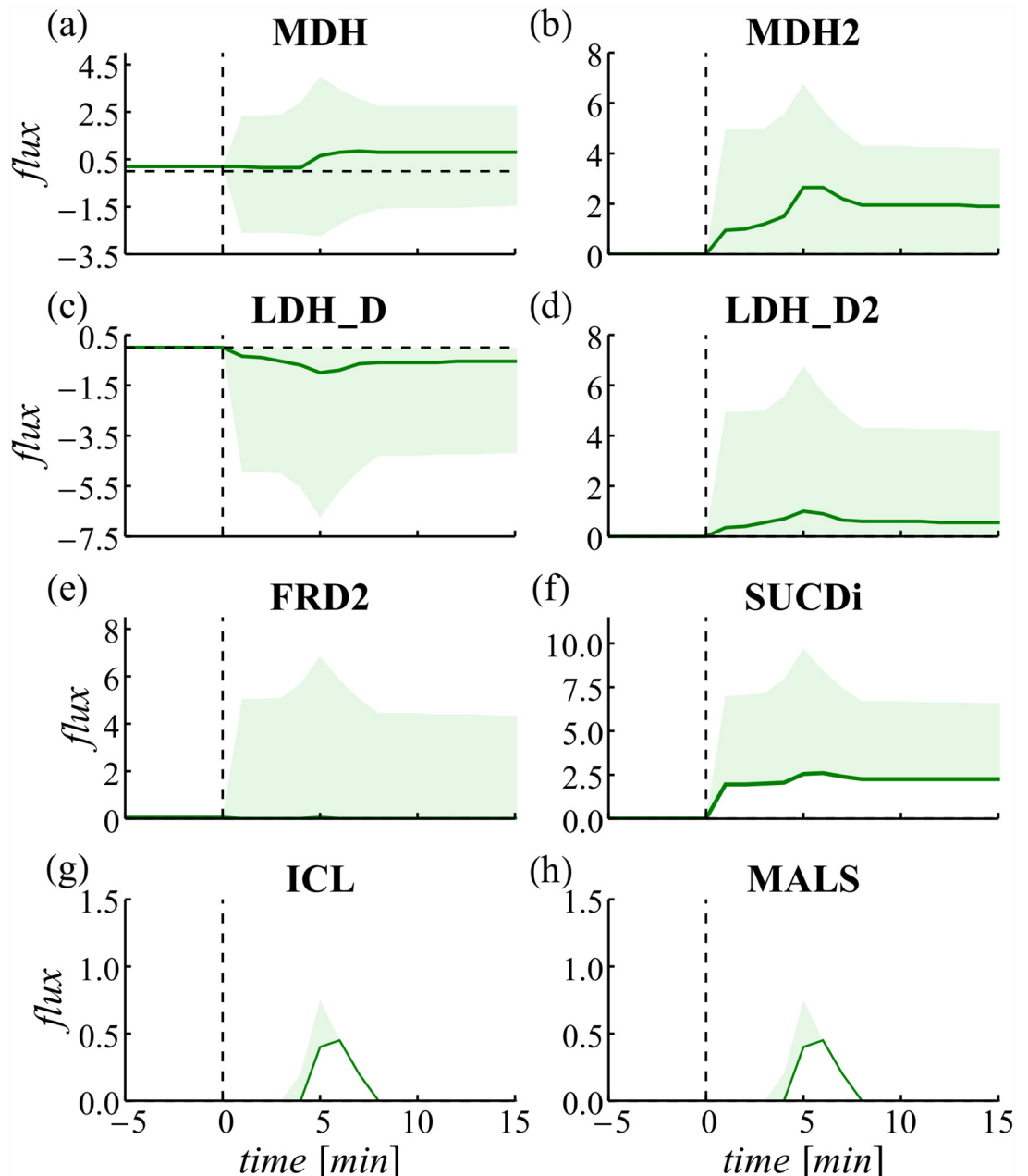
Fig 4. Fluxes of unbalanced reactions in dddFBA. The pFBA solution with minimized squared fluxes is given in black. The shadowed area indicates flux variability.

upper flux bounds reflecting gene expression we simulated mRNA expression using an "on demand" regulatory model to check consistency with measured mRNA. This is in contrast to other methods that incorporate mRNA expression data directly in flux constraints. The presented dddFBA method introduced is able to qualitatively describe the transient overexpression of multiple genes observed during a shift experiment from anaerobic to aerobic conditions. Several overexpressed genes can be classified as MLE or ELE (according to [8]) and thus are not detected by conventional FBA methods. With the introduction of additional constraints arising from gene expression that is directed by the demand for the respective flux we are able to
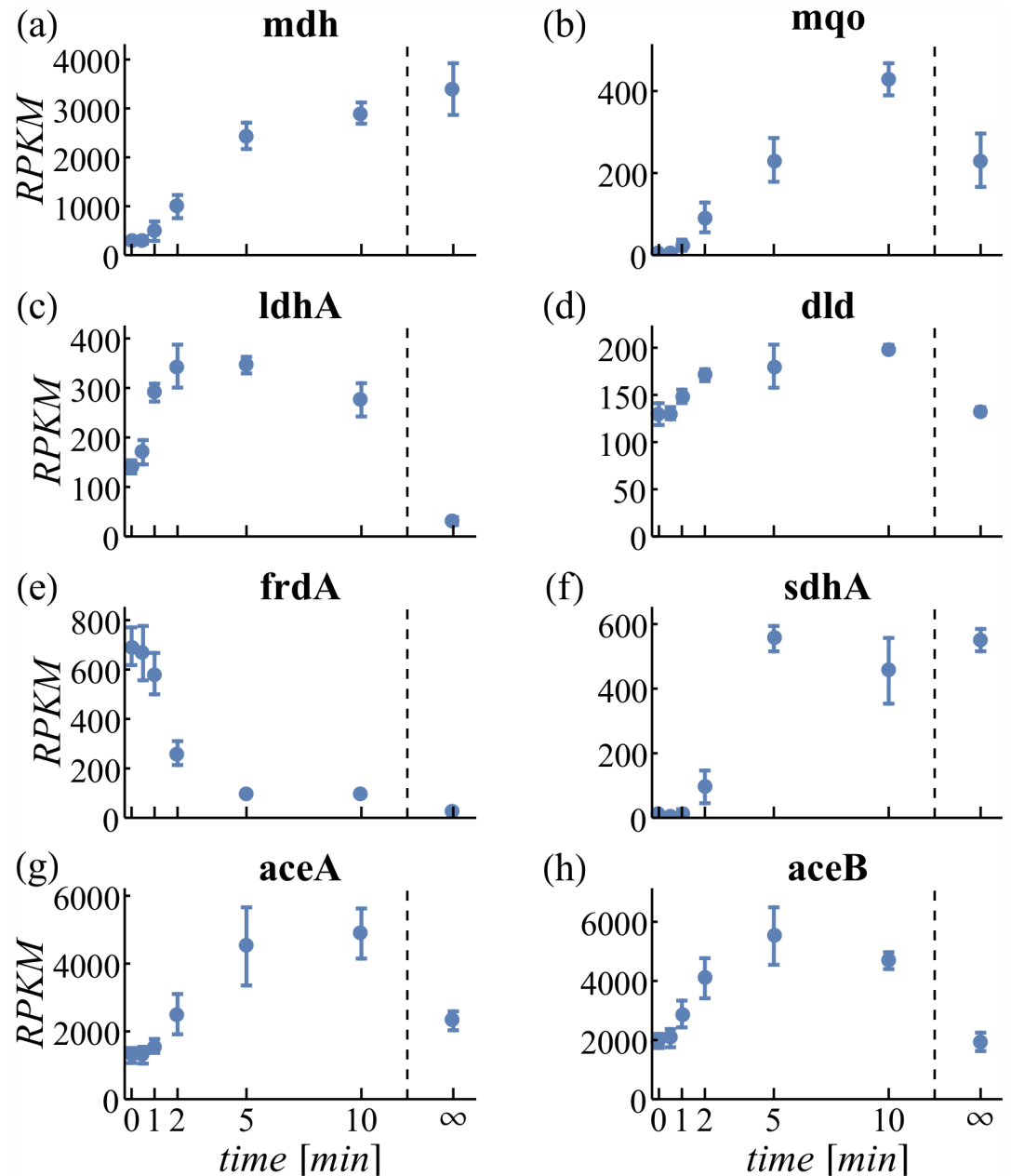
Fig 5. Expression of unbalanced genes. RPKM values of the indicated genes are given with standard deviation. ∞ denotes the aerobic steady-state expression level.

correctly describe transient overexpression. Furthermore, it even enables prediction of temporary fluxes of reactions whose respective enzymes are not explicitly balanced in the model. Qualitatively, the emergence of these temporary fluxes is supported by gene expression data.

Yet, the applicability of this method to other problems is limited. Experimental data are required for parametrization of the model since the estimated transcription rates vary considerably (Table 1).

Still, much can be deduced from this modeling approach: (1), it is reasonable to constrain reaction fluxes with respect to gene expression, (2), limited transcription rates can be conceived

as being causal to the (temporary) assignment of less efficient enzymes, (3), constraints on several key enzymes suffice to predict transient behavior of other reactions, as well, and (4), modeling regulation of gene expression in an "on demand" fashion in a pFBA framework with growth optimization results in a realistic gene expression pattern.

A number of simplifications had to be taken to keep the model feasible. We did not account for posttranscriptional or posttranslational modifications that might restrain enzyme activity. Metabolic regulation can be another source for impaired enzyme functioning. However, by applying only a maximal upper bound on the fluxes solutions with inhibitors are still included. Transcription and translation can be seen as transport processes as polymerases and ribosomes migrate along the polynucleotides. The ODEs in Eqs 1 and 3 do not account for that. Instead, translation increases instantaneously with transcription, which is reasonable for *E. coli* as both processes occur simultaneously at the same RNA molecule.

In reality, gene expression is not the only limitation applying when optimizing cellular enzyme composition. A multitude of constraints are probably effective, including, (1), membrane occupancy by glucose uptake and ETC proteins [41], (2), molecular crowding of cytoplasm that limits total amounts of translated proteins [39], (3), cells need to find a reasonable tradeoff between growth rate and energetic efficiency, also taking the costs of gene expression into account [6, 45, 46], (4), during a shift of the external conditions, minimal and fast dynamic adjustment of gene expression are required [7], (5), in a dynamic setting, competition for limited RNA polymerases, ribosomes, and tRNAs applies [47], and (6), hampered availability of large protein complexes is also conceivable, e.g. by chaperone efficiency.

These constraints are hard to describe through modeling and even harder to parametrize, especially in a dynamic setting. However, depending on the specific problem and the objective of the modeling, not all limitations need to be taken into account, as demonstrated by the good agreement of the dddFBA simulation with gene expression data.

The data show that MLE enzymes seem to be expressed during the transition from anaerobic to aerobic conditions. Here, we present evidence that this might be due to low availability of the more efficient enzymes. Expression of the MLE enzymes helps to overcome this situation and to increase growth rate, however, this is at the cost of additional gene expression. In a large scale bioreactor with frequently-changing oxygen availability, this finding might be part of the explanation for the observed reductions in growth rate (up to 30%, [48]). Furthermore, as oxygen triggers such a large perturbation, with transiently more than 20% genes of the genome differentially expressed [5], these genes are likely to occupy a good portion of the transcription and translation machinery. These polymerases and ribosomes might then be unavailable for expression of recombinant proteins (up to 94% loss reported, [48]). In the long run, cells will need to adapt to the frequently but nonuniformly changing oxygen supply.

Materials and Methods

Data source

We used a dataset of short-time RNA sequencing after an anaerobic to aerobic shift of an *E. coli* batch culture available through NCBI's Gene Expression Omnibus [49] via the GEO Series accession number GSE71562 [5]. The Bioconductor package edgeR [50] was applied for statistical analysis. Read counts were transformed to counts per million (CPM) taking into account the different library sizes. Normalization with the respective gene lengths yields reads per kilobase per million reads (RPKM), which are used for estimating RNA synthesis parameters. A negative binomial model was fitted with the count data and the common and tag-wise dispersions were estimated. Subsequently, a generalized linear model was fitted to the data and p-values were calculated. The resulting p-values were corrected for multiple testing using the FDR

method [51]. Genes with p-values of less than 0.05 and absolute fold change greater than 2 were assumed differentially expressed.

Parameter estimation

From the time course of RPKM values of a gene, synthesis parameters were estimated. This was accomplished by fitting the parameters basal synthesis rate ($s_{b,i}$), activated synthesis rate ($s_{a,i}$, both in $\frac{RPKM}{min}$), and the dead time ($T_{d,i}$, in min) of the ordinary differential equation (ODE) Eq 6. Degradation rates (γ_i) were adopted from the study of Bernstein *et al.* [30]. The squared brackets denote the concentration of the respective entity.

$$\frac{d[mRNA_i]}{dt} = s_{b,i} + \Theta(t - T_{d,i}) \cdot s_{a,i} - (\mu + \gamma_i) \cdot [mRNA_i](t) \tag{6}$$

$$\Theta(t) = \begin{cases} 0 & t < 0 \\ 1 & t \geqslant 0 \end{cases} \tag{7}$$

ODE Eq 6 is analogous to ODE Eq 1, only that the regulatory signal is replaced by the Heaviside function $\Theta(t - T_{d,i})$. The model (Eqs 6 and 7) incorporates a basal synthesis rate that is assumed to be always active. An additionally activated synthesis rate supports gene expression when the Heaviside function (Eq 7) switches from 0 to 1 after a delay of $T_{d,i}$. Transcribed mRNA$_i$ is actively degraded with a certain γ_i and diluted with the growth rate μ.

Assuming only basal synthesis for genes that are known to be inactive in an anaerobic environment the initial condition of Eq 6 simplifies to (with $\Theta(0) = 0$)

$$[mRNA_i](0) = \frac{s_{b,i}}{\mu + \gamma_i}. \tag{8}$$

Inserting RPKM values of a steady state anaerobic culture yields the parameters $s_{b,i}$. The growth rate μ was assumed to be 0.26 h^{-1} for anaerobic growth, as measured in preliminary experiments with the same growth conditions. No reasonable measurements of the growth rate could be performed within the 10 min time frame of the original experiment.

Parameter estimations for the yet missing $s_{a,i}$ were performed in *Mathematica* (Wolfram, Oxfordshire, UK) using the function `NMinimize` and the method `SimulatedAnnealing` [52] using the least squares estimator.

Demand-directed dFBA

dddFBA was implemented as described by the Eqs 1 to 5 and integrated in a dynamic pFBA framework, i.e. in each simulation time step, (1) growth rate is maximized, (2) total fluxes are minimized, (3) external metabolite concentrations are updated, and (4) new bounds for uptake reactions, as well as for the balanced reactions are calculated. We used the genome scale model iJO1366 [14] for flux optimization. We applied this technique to the six enzymes listed in Table 1 employing the indicated parameters.

In Eq 4, we assume the parameter for the protein synthesis rate, $s_{P,i}$, to be inversely proportional to the length of the respective gene i, since ribosomes have to migrate along the distance of the gene with limited velocity to translate the protein. The value of the translation constant was chosen to be equal to 0.6 $\frac{bp \cdot mmol}{h \cdot RPKM \cdot gDCW}$ (Eq 4). If a protein is made up of different subunits, the gene coding the largest subunit will be used, because the longest gene is assumed to require the longest time for transcription and translation and so is supposed to be rate limiting for the expression of the holoenzyme. If the enzyme complex is made up of several (n) copies of the

longest subunit, the gene length for that subunit becomes an apparent gene length with the interribosomal distance (of 72 bp [53]) added n times. For example, the enzyme AKGDH has a subunit stoichiometry of $[(SucA)_{12}][(SucB)_{24}][(Lpd)_2]$. *sucA* is the longest gene of the enzyme complex with a length of 2802 bp. Since 12 copies of that subunit are required and ribosomes are on average 72 bp apart from each other the apparent gene length is $(2802 + 12 \cdot 72)$ $bp = 3666\ bp$.

For FVA, the second optimization step was omitted, instead, each indicated reaction was maximized and minimized within the growth optimal solution space [54].

Parameter variations were employed to analyze sensitivity of the respective parameters to fluxes and simulated gene expression. For this, we performed simulations with combinations of the varying parameters translation constant, θ, k_{cat}(CYTBO3_4pp), and k_{cat}(AKGDH).

Supporting Information

S1 Fig. Variation of global parameters. The translation constant is varied between 0.3 and 0.9 (columns); the threshold for regulation is varied between 0.1 and 0.9 (rows).
(PDF)

S2 Fig. Variation of k_{cat}. The turnover number of CYTBO3_4pp is varied between 170.5 and 682 (columns); the turnover number of AKGDH is varied between 24.5 and 98 (rows).
(PDF)

Acknowledgments

Author Contributions

Conceived and designed the experiments: JW RF OS. Performed the experiments: JW. Analyzed the data: JW RF. Wrote the paper: JW RF OS.

References

1. Enfors SO, Jahic M, Rozkov A, Xu B, Hecker M, Jürgen B, et al. Physiological responses to mixing in large scale bioreactors. J Biotechnol. 2001 feb; 85(2):175–85. doi: 10.1016/S0168-1656(00)00365-5 PMID: 11165362

2. Junker BH. Scale-up methodologies for Escherichia coli and yeast fermentation processes. J Biosci Bioeng. 2004 jan; 97(6):347–64. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16233642. doi: 10.1016/S1389-1723(04)70218-2 PMID: 16233642

3. Takors R. Scale-up of microbial processes: impacts, tools and open questions. J Biotechnol. 2012 jul; 160(1–2):3–9. doi: 10.1016/j.jbiotec.2011.12.010

4. Imlay JA. The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. Nat Rev Microbiol. 2013 may; 11(7):443–454. doi: 10.1038/nrmicro3032 PMID: 23712352

5. von Wulffen J, Jäger G, Sawodny O, Feuer R. RNA sequencing of E. coli transitioning from anaerobic to aerobic conditions reveals temporary expression of several metabolic enzymes. submitted;.

6. Molenaar D, van Berlo R, de Ridder D, Teusink B. Shifts in growth strategies reflect tradeoffs in cellular economics. Mol Syst Biol. 2009; 5(323):1–10. Available from: http://msb.embopress.org/cgi/doi/10.1038/msb.2009.82.

7. Zaslaver A, Mayo AE, Rosenberg R, Bashkin P, Sberro H, Tsalyuk M, et al. Just-in-time transcription program in metabolic pathways. Nat Genet. 2004; 36(5):486–491. Available from: http://www.nature.com/doifinder/10.1038/ng1348. doi: 10.1038/ng1348 PMID: 15107854

8. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, et al. Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. Mol Syst Biol. 2010; 6(390):390. doi: 10.1038/msb.2010.47 PMID: 20664636

9. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nat Biotechnol. 2010 mar; 28(3):245–8. doi: 10.1038/nbt.1614 PMID: 20212490

10. Varma A, Palsson BO. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. Appl Environ Microbiol. 1994; 60(10):3724–3731. PMID: 7986045

11. Edwards JS, Palsson BO. Metabolic flux balance analysis and the in silico analysis of Escherichia coli K-12 gene deletions. BMC Bioinformatics. 2000; 1:1. doi: 10.1186/1471-2105-1-1 PMID: 11001586

12. Edwards JS, Ibarra RU, Palsson BO. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. Nat Biotechnol. 2001; 19(2):125–130. doi: 10.1038/84379 PMID: 11175725

13. Schuetz R, Kuepfer L, Sauer U. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. Mol Syst Biol. 2007 jan; 3(119):119. doi: 10.1038/msb4100162 PMID: 17625511

14. Orth JD, Conrad TM, Na J, Lerman Ja, Nam H, Feist AM, et al. A comprehensive genome-scale reconstruction of Escherichia coli metabolism–2011. Mol Syst Biol. 2011 jan; 7(535):535. doi: 10.1038/msb.2011.65 PMID: 21988831

15. Feist AM, Palsson BO. The biomass objective function. Curr Opin Microbiol. 2010; 13(3):344–349. doi: 10.1016/j.mib.2010.03.003 PMID: 20430689

16. Mahadevan R, Edwards JS, Doyle FJ. Dynamic flux balance analysis of diauxic growth in Escherichia coli. Biophys J. 2002; 83(September):1331–1340. doi: 10.1016/S0006-3495(02)73903-9 PMID: 12202358

17. Covert MW, Schilling CH, Palsson B. Regulation of gene expression in flux balance models of metabolism. J Theor Biol. 2001 dec; 213(1):73–88. doi: 10.1006/jtbi.2001.2405 PMID: 11708855

18. Covert MW, Palsson BO. Constraints-based models: Regulation of Gene Expression Reduces the Steady-state Solution Space. J Theor Biol. 2003 apr; 221(3):309–325. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022519303930712. doi: 10.1006/jtbi.2003.3071 PMID: 12642111

19. Covert MW, Xiao N, Chen TJ, Karr JR. Integrating metabolic, transcriptional regulatory and signal transduction models in Escherichia coli. Bioinformatics. 2008 sep; 24(18):2044–50. doi: 10.1093/bioinformatics/btn352 PMID: 18621757

20. Chandrasekaran S, Price ND. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis. Proc Natl Acad Sci U S A. 2010 oct; 107(41):17845–50. doi: 10.1073/pnas.1005139107 PMID: 20876091

21. Machado D, Herrgård M. Systematic Evaluation of Methods for Integration of Transcriptomic Data into Constraint-Based Models of Metabolism. PLoS Comput Biol. 2014; 10(4):e1003580. Available from: http://dx.plos.org/10.1371/journal.pcbi.1003580. doi: 10.1371/journal.pcbi.1003580 PMID: 24762745

22. Åkesson M, Förster J, Nielsen J. Integration of gene expression data into genome-scale metabolic models. Metab Eng. 2004; 6(4):285–293. doi: 10.1016/j.ymben.2003.12.002 PMID: 15491858

23. Becker SA, Palsson BO. Context-specific metabolic networks are consistent with experiments. PLoS Comput Biol. 2008; 4(5):e1000082. doi: 10.1371/journal.pcbi.1000082 PMID: 18483554

24. Colijn C, Brandes A, Zucker J, Lun DS, Weiner B, Farhat MR, et al. Interpreting expression data with metabolic flux models: predicting Mycobacterium tuberculosis mycolic acid production. PLoS Comput Biol. 2009; 5(8):e1000489. doi: 10.1371/journal.pcbi.1000489 PMID: 19714220

25. van Berlo RJP, de Ridder D, Daran JM, Daran-Lapujade PaS, Teusink B, Reinders MJT. Predicting metabolic fluxes using gene expression differences as constraints. IEEE/ACM Trans Comput Biol Bioinform. 2011; 8(1):206–16. doi: 10.1109/TCBB.2009.55 PMID: 21071808

26. Zelezniak A, Sheridan S, Patil KR. Contribution of network connectivity in determining the relationship between gene expression and metabolite concentration changes. PLoS Comput Biol. 2014 apr; 10(4):e1003572. doi: 10.1371/journal.pcbi.1003572 PMID: 24762675

27. O'Brien EJ, Lerman Ja, Chang RL, Hyduke DR, Palsson BØ. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. Mol Syst Biol. 2013 jan; 9(693):693.

Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3817402&tool=pmcentrez&rendertype=abstract. doi: 10.1038/msb.2013.52 PMID: 24084808

28. Kosuri S, Goodman DB, Cambray G, Mutalik VK, Gao Y, Arkin AP, et al. Composability of regulatory sequences controlling transcription and translation in Escherichia coli. Proc Natl Acad Sci U S A. 2013; 110(34):14024–9. Available from: http://www.pnas.org/content/110/34/14024.full. doi: 10.1073/pnas.1301301110 PMID: 23924614

29. Thiele I, Palsson BØ. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc. 2010; 5(1):93–121. Available from: http://www.nature.com/doifinder/10.1038/nprot.2009.203. doi: 10.1038/nprot.2009.203 PMID: 20057383

30. Bernstein JA, Khodursky AB, Lin PH, Lin-Chao S, Cohen SN. Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. Proc Natl Acad Sci. 2002 jul; 99(15):9697–9702. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.112318199. doi: 10.1073/pnas.112318199 PMID: 12119387

31. Bennett BD, Kimball EH, Gao M, Osterhout R, Van Dien SJ, Rabinowitz JD. Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. Nat Chem Biol. 2009 aug; 5(8):593–599. Available from: http://www.nature.com/doifinder/10.1038/nchembio.186. doi: 10.1038/nchembio.186 PMID: 19561621

32. Maurizi MR. Proteases and protein degradation in Escherichia coli. Experientia. 1992 feb; 48(2):178–201. Available from: http://link.springer.com/10.1007/BF01923511. doi: 10.1007/BF01923511 PMID: 1740190

33. Rice CW, Hempfling WP. Oxygen-limited continuous culture and respiratory energy conservation in Escherichia coli. J Bacteriol. 1978 apr; 134(1):115–24. PMID: 25879

34. Bolgiano B, Salmon I, Poole RK. Reactions of the membrane-bound cytochrome bo terminal oxidase of Escherichia coli with carbon monoxide and oxygen. Biochim Biophys Acta. 1993; 1141:95–104. doi: 10.1016/0005-2728(93)90194-K PMID: 8382081

35. Rapisarda VA, Montelongo LR, Farias RN, Massa EM. Characterization of an NADH-Linked Cupric Reductase Activity from the Escherichia coli Respiratory Chain. Arch Biochem Biophys. 1999 oct; 370(2):143–150. doi: 10.1006/abbi.1999.1398 PMID: 10510271

36. Leif H, Sled VD, Ohnishi T, Weiss H, Friedrich T. Isolation and characterization of the proton-translocating NADH: ubiquinone oxidoreductase from Escherichia coli. Eur J Biochem. 1995; 230:538–548. PMID: 7607227

37. Waskiewicz DE, Hammes GG. Elementary steps in the reaction mechanism of the alpha-ketoglutarate dehydrogenase multienzyme complex from Escherichia coli: kinetics of succinylation and desuccinylation. Biochemistry. 1984; 23(1982):3136–3143. doi: 10.1021/bi00309a005 PMID: 6380583

38. Schulze E, Westphal AH, Obmolova G, Mattevi A, Hol WG, de Kok A. The catalytic domain of the dihydrolipoyl transacetylase component of the pyruvate dehydrogenase complex from Azotobacter vinelandii and Escherichia coli. Expression, purification, properties and preliminary X-ray analysis. Eur J Biochem. 1991 nov; 201(3):561–8. doi: 10.1111/j.1432-1033.1991.tb16315.x PMID: 1935951

39. Beg QK, Vazquez A, Ernst J, de Menezes MA, Bar-Joseph Z, Barabási AL, et al. Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity. Proc Natl Acad Sci U S A. 2007; 104(31):12663–12668. doi: 10.1073/pnas.0609845104 PMID: 17652176

40. Adadi R, Volkmer B, Milo R, Heinemann M, Shlomi T. Prediction of microbial growth rate versus biomass yield by a metabolic network with kinetic parameters. PLoS Comput Biol. 2012; 8(7):e1002575. Available from: http://dx.plos.org/10.1371/journal.pcbi.1002575. doi: 10.1371/journal.pcbi.1002575 PMID: 22792053

41. Zhuang K, Vemuri GN, Mahadevan R. Economics of membrane occupancy and respiro-fermentation. Mol Syst Biol. 2011 apr; 7(1):500. Available from: http://dx.doi.org/10.1038/msb.2011.34. doi: 10.1038/msb.2011.34 PMID: 21694717

42. Iuchi S, Chepuri V, Fu HA, Gennis RB, Lin EC. Requirement for terminal cytochromes in generation of the aerobic signal for the arc regulatory system in Escherichia coli: study utilizing deletions and lac fusions of cyo and cyd. J Bacteriol. 1990; 172(10):6020–6025. PMID: 2170337

43. Iuchi S, Weiner L. Cellular and molecular physiology of Escherichia coli in the adaptation to aerobic environments. J Biochem. 1996 dec; 120(6):1055–1063. doi: 10.1093/oxfordjournals.jbchem.a021519 PMID: 9010748

44. Lindqvist A, Membrillo-Hernández J, Poole RK, Cook GM. Roles of respiratory oxidases in protecting Escherichia coli K12 from oxidative stress. Antonie Van Leeuwenhoek. 2000 jul; 78(1):23–31. doi: 10.1023/A:1002779201379 PMID: 11016692

45. Oyarzún DA, Ingalls BP, Middleton RH, Kalamatianos D. Sequential activation of metabolic pathways: A dynamic optimization approach. Bull Math Biol. 2009; 71(8):1851–1872. doi: 10.1007/s11538-009-9427-5 PMID: 19412635

46. Waldherr S, Oyarzún DA, Bockmayr A. Dynamic optimization of metabolic networks coupled with gene expression. J Theor Biol. 2015 jan; 365:469–485. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022519314006250. doi: 10.1016/j.jtbi.2014.10.035 PMID: 25451533

47. De Vos D, Bruggeman FJ, Westerhoff HV, Bakker BM. How molecular competition influences fluxes in gene expression networks. PLoS One. 2011; 6(12):e28494. Available from: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0028494. doi: 10.1371/journal.pone.0028494 PMID: 22163025

48. Lara AR, Galindo E, Ramírez OT, Palomares LA. Living With Heterogeneities in Bioreactors: Understanding the Effects of Environmental Gradients on Cells. Mol Biotechnol. 2006; 34(3):355–382. Available from: http://link.springer.com/10.1385/MB:34:3:355. doi: 10.1385/MB:34:3:355 PMID: 17284782

49. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002; 30(1):207–210. doi: 10.1093/nar/30.1.207 PMID: 11752295

50. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26(1):139–140. doi: 10.1093/bioinformatics/btp616 PMID: 19910308

51. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc. 1995; 57(1):289–300.

52. Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by Simulated Annealing. Science (80-). 1983; 220 (4598):671–680. doi: 10.1126/science.220.4598.671

53. Brandt F, Etchells SA, Ortiz JO, Elcock AH, Hartl FU, Baumeister W. The Native 3D Organization of Bacterial Polysomes. Cell. 2009; 136(2):261–271. doi: 10.1016/j.cell.2008.11.016 PMID: 19167328

54. Mahadevan R, Schilling CH. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab Eng. 2003; 5(4):264–276. doi: 10.1016/j.ymben.2003.09.002 PMID: 14642354